

A photograph of a brick building with a large glass window and a balcony, surrounded by greenery and a fountain. The building is the main focus, with a balcony featuring a metal railing and a large glass window to the right. In the foreground, there is a well-manicured lawn with a fountain and a path. The sky is clear and blue.

# Global Wireless Summit -2016

Technology for the Betterment of Human Communication

## Proceedings

Fourth Global Wireless Summit (GWS-2016),  
November 27-30, 2017  
Aarhus University, Aarhus, Denmark

## **About Global Wireless Summit 2016**

Global wireless summit is the world's leading conference on wireless technologies covering the latest advances in wireless communication technologies. The conference brings together academia, industry and standardization bodies to explore activities, trends and future challenges towards ICT globalization including energy and security issues in existing and future wireless technologies such as cellular, short-range, sensors, radio access, vehicular communication and embedded ones. The conferences also focus on issues in cognitive and self-organizing networks, Internet of Things, nano sensors, positioning and localization, networks without borders, recent advances in information theory and its application, multimedia applications and services. A key emphasis of GWS-2016 was interdisciplinarity - bridging across industries, public and private companies, universities, faculties and other knowledge societies in exploration of the newest wireless technologies pursuing new ways of communicating, new business models as well as innovative and sustainable solutions for the future.

The fourth conference in 2016 took place in Aarhus, Denmark under the theme of “**Technology for the Betterment of Human Communication.**”

# About Aarhus University

The university was founded in 1928 and today it has several world-class research fields.

Aarhus University (AU) is a top ten university among universities founded within the past 100 years. It has a long tradition of partnerships with some of the world's best research institutions and university networks.

AU has a strong commitment to the development of society that is realised through its collaboration with government agencies and institutions and the business community.

The university's goal is to contribute towards solving the complex global challenges facing the world. The university therefore strives to combine the high level of academic standards of its researchers with collaboration across disciplinary boundaries to combine research in new ways.

This takes place in close contact with the world around us and creates the basis for the university to be internationally competitive within the areas of research, education, talent development and knowledge exchange.

## Facts about AU

- Number of students 2015: 42,500
- Number of employees 2015: 11,550
- Number of Bachelor's degree graduates 2015: 4,489
- Number of Master's degree graduates 2015: 4,520
- Number of PhD graduates 2015: 471

## Aarhus University Faculty



**JOHNNY LAURSEN,**  
Dean, AU Arts  
Nordre Ringgade 1, building 1431, 223,  
8000 Aarhus C, Denmark



**THOMAS PALLESEN,**  
Dean, Aarhus BSS,  
Bartholins Allé 14,  
building 1327, 330,  
8000 Aarhus C, Denmark

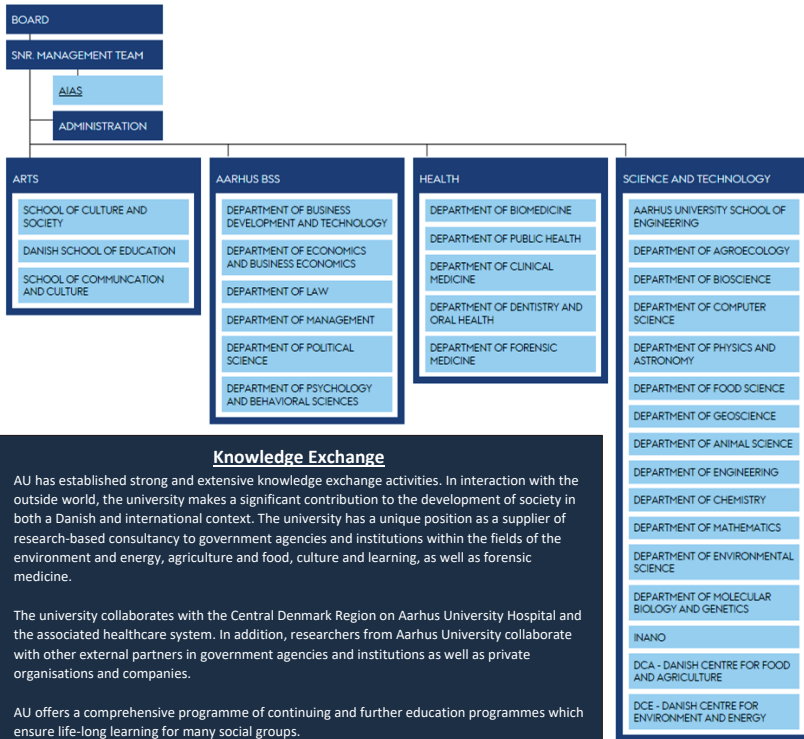


**NIELS CHRISTIAN NIELSEN,**  
Dean, AU Science and Technology  
Ny Munkegade 120, building 1521, 216,  
8000 Aarhus C, Denmark



**OLE STEEN NIELSEN,**  
Dean, AU Health Sciences,  
Nordre Ringgade 1,  
building 1431, 113,  
8000 Aarhus C, Denmark

# Arhus University Organization



## Knowledge Exchange

AU has established strong and extensive knowledge exchange activities. In interaction with the outside world, the university makes a significant contribution to the development of society in both a Danish and international context. The university has a unique position as a supplier of research-based consultancy to government agencies and institutions within the fields of the environment and energy, agriculture and food, culture and learning, as well as forensic medicine.

The university collaborates with the Central Denmark Region on Aarhus University Hospital and the associated healthcare system. In addition, researchers from Aarhus University collaborate with other external partners in government agencies and institutions as well as private organisations and companies.

AU offers a comprehensive programme of continuing and further education programmes which ensure life-long learning for many social groups.

## Research

Research at Aarhus University is partly organised in departments distributed among four main academic areas, and partly in basic research centres. Aarhus University also comprises a number of interdisciplinary centres.

Two researchers from the university have received the Nobel Prize, firstly in Chemistry (1997), and subsequently in economics (2010) respectively. Several highly coveted research awards and grants have also gone to a large number of researchers at Aarhus University.

There are more than 6,000 researchers (4,300 full-time equivalents) at the university, including approximately 2,000 PhD students.

## Education

Research at Aarhus University is The study programmes at AU are based on and integrated with the research activities. They are developed in a close collaboration between employers, academic staff and students. Aarhus University's most important contribution to society are its graduates.

AU has 44,500 students (2012) – of which 5,000 are foreign students. The university has an annual intake of just under 7,000 students, while 4,000 students graduate each year.

AU was the first university in Denmark to recruit students to the PhD programme before they were awarded their Master's degree.

## Education

Research at Aarhus University is The study programmes at AU are based on and integrated with the research activities. They are developed in a close collaboration between employers, academic staff and students. Aarhus University's most important contribution to society are its graduates.

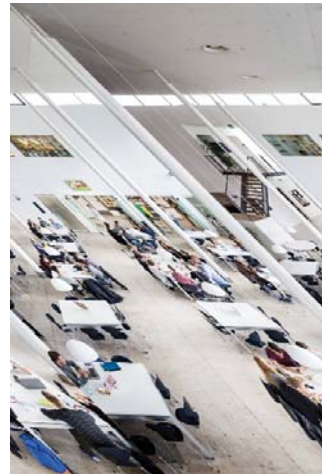
AU has 44,500 students (2012) – of which 5,000 are foreign students. The university has an annual intake of just under 7,000 students, while 4,000 students graduate each year.

AU was the first university in Denmark to recruit students to the PhD programme before they were awarded their Master's degree.

## **The Department of Business Development and Technology (BTECH)**

The department's focus is on research and educational excellence and offers four undergraduate and two graduate study programmes as well as a number of part-time studies all within business engineering, business economics and management. Furthermore, Aarhus School of Engineering offers the Electronic Engineer degree programme and the pre-admission course for engineering programmes at AU Campus Herning. The study programmes offered by BTECH are based on interaction across professional boundaries as well as close interaction with students and the surrounding business community. All study programmes have an innovative and international perspective.

The department was originally founded in 1995 as a result of a merger between the Engineering College of Western Jutland and the Central Jutland Business School. The merger with Aarhus University in 2006 resulted in an even higher academic level and an increased focus on research and development.



## **The MBIT Group**



The MBIT (Multi-Business Model Innovation and Technology) research group conducts research within the area of engineering and business development by applying the 'MBIT mindset'. The research group is headed by Professor Peter Lindgren and Associate Professor Annabeth Aagaard, and currently exists of a series of interdisciplinary competences. MBIT has three pillars: research, business development and education. We use these to support the MSc in Engineering (Technology Based Business Development) study programme at BTECH.

The core of MBIT is business model innovation, which is a generic, highly efficient and fresh approach to developing any business in any environment. Solid businesses all around the world with excellent offerings in terms of products, services and processes have suffered in the last decades. Not because they could not innovate their offerings, but because someone else changed the game of the ecosystem by radically changing the business model, and thereby disrupting or dislocating the ecosystem and shortening the business model lifecycles. Traditional product- or process-oriented innovation, such as adding features and/or performance to an existing product, is not sufficient anymore; it is simply too easy for competitors and externals to follow or disrupt.

Business model innovative moves are often seen during a crisis and when a business is forced to do something radically and bold. However, we believe that applying business model innovation as an ongoing activity can help any business stay competitive and ensure renewal in a continuous, incremental way.

When conducting business model innovation, you get an overview of the different business models that actually operate in your business. This overview will help you ensure and create synergies between the business models in your portfolio. In addition, it can help you determine which areas would be most beneficial to focus on in order to strengthen your business.

## Research Objectives and Interests

### Education

MBIT contributes to the engineering study programmes at BTECH/AU, and in particular, to the research base underpinning the MSc in Engineering study programme (cand.polyt.)

### Business

MBIT aims to build bridges to businesses and public institutions by means of business development, business development technologies, entre-, intra- and interpreneurship as well as strategic business model innovation.

### Research

MBIT pursues on and publishes research in the area of engineering and engineering management as consisting of network-based and integrated businesses and their business engineering design methodologies, which include:

- Business engineering operations frameworks
- Business technological artefacts
- Multi-business model and innovation systems
- Business model information systems
- Business model information technology
- Multi-business model modelling
- Entre-, intra and interpreneurship
- Strategic multi-business model innovation
- Sustainable business models
- Data-driven business models

To pursue these themes, we execute activities such as:

Combining business models and related big data from experiments with (applied) science to understand the processes in scientific and engineering business model challenges and problems.



# **GWS 2016 Committee**

## Committee Members



**RAMJEE PRASAD,**  
Steering Committee Chair, GWS



**PETER LINDGREN,**  
General Chair, Global Wireless Summit 2016  
Professor, Multi Business Model Innovation and  
Technology  
Aarhus University, AU Herning , Birk Centerpark 15,  
7400 Herning, Denmark



**ANNABETH AAGAARD,**  
TPC Co-Chair, Global Wireless Summit 2016  
Department of Business Development and Technology  
Birk Centerpark 15, building 8001, Innovatorium  
7400 Herning, Denmark,



**DINA SIMUNIC,**  
TPC Co-Chair, Global Wireless Summit 2016  
University of Zagreb, Croatia



**NEELI R. PRASAD,**  
TPC Co-Chair, Global Wireless Summit 2016  
SPA Solutions-LLC, Mountain View, USA



**VLADIMIR POULKOV,**  
TPC Co-Chair, Global Wireless Summit 2016 Technical  
University of Sofia , Bulgaria



**MAHASWETA SARKAR,**  
Workshop Co-Chair, Global Wireless Summit 2016  
San Diego State University, USA



**PREETAM KUMAR,**  
Workshop Co-Chair, Global Wireless Summit 2016  
IIT Patna, India



**MORTEN KARNØE,**  
Workshop Co-Chair, Global Wireless Summit 2016  
Aarhus University, Denmark



**JACOB STEENDAHL NIELSEN,**  
Tutorial Co-Chair, Global Wireless Summit 2016  
Aarhus University, Denmark



**RAMA RAO,**  
Tutorial Co-Chair, Global Wireless Summit 2016  
SRM University, India



**PERIKLIS CHATZIMISIOS,**  
Tutorial Co-Chair, Global Wireless Summit 2016  
Alexander Technological Educational Institute of  
Thessaloniki, Greece



**HENRIK KNUDSEN,**  
Panel Co-Chair, Global Wireless Summit 2016  
Aarhus University, Denmark



**DILIP KRISHNASWAMY,**  
Panel Co-Chair, Global Wireless Summit 2016  
IBM Research, India



**PETER MÜLLER,**  
Panel Co-Chair, Global Wireless Summit 2016  
IBM Research, Switzerland



**BORIVOJ MODLIC,**  
Patronage Chair, Global Wireless Summit 2016  
University of Zagreb, Croatia



**ERNESTINA CIANCA,**  
Publication Chair, Global Wireless Summit 2016  
University of Rome Tor Vergata, Italy



**MAJA ILIC-DELIBASIC,**  
Publicity Chair, Global Wireless Summit 2016  
University of Montenegro, Montenegro



**AMBUJ KUMAR,**  
Web Chair, Global Wireless Summit 2016  
Aarhus University, Denmark

# **Welcome Speeches**

On behalf of the steering committee of the global wireless summit (GWS) 2016, it is my great pleasure to bid you a warm welcome to this highly renowned annual event and the beautiful historic city of Aarhus, Denmark.

After successful editions of GWS in Atlantic City, USA in 2013, followed by GWS-2014 in Aalborg, Denmark, and GWS-2015 in Hyderabad, India, GWS-2016 brought forward the theme of “technology for the betterment of human communications”.

Wireless communications have evolved over the years to open great potential for monetizing on multimedia data and diversified user-generated content, making them a ubiquitous means of connecting people and businesses. The current communications landscape is characterized with billions and billions of devices interworking through a myriad of technologies for the delivery of smart personalized services and applications. At the center of these is the human user who drives his/her own interconnected cluster. The vision pushed forward by GWS-2016 was a technological and communications landscape of large-scale and dense human centric connectivity, the complexity of which can be used to extract new business value propositions.

GWS-2016 brought together interdisciplinary themes and the respective stakeholders, researchers, industrial and standardization and policy experts to focus on both the technological and human aspect arising from that vision. The setting of GWS-2016 was in the heart of jutland, Denmark, in the city of Aarhus, the second largest city in Denmark, known for its rapid businesses and market growth, which is also supported by the young talents graduating from Aarhus university. In the heart of the city is the historical open-air museum where one may learn about the traditional danish lifestyles.

The GWS-2016 program featured an unusually distinguished group of keynote and invited speakers, as well as a strong technical program with novel research contributions from the relevant thematic areas from all over the world. In addition, an extensive list of social events were scheduled to allow for a break from the exciting technical sessions. The participants took the advantage of the many opportunities to network and expand their technical and cultural horizons.

I would like to extend my appreciation to the GWS2016 general chair and organizing committee for their dedication and effort to put together a strong event.



**Ramjee Prasad**, Steering Committee Chair, Global Wireless Summit

It is a great honour for me to welcome you all to the GWS 2016 Summit at Aarhus University, Denmark. Aarhus is placed in the very heart of the Central Denmark Region and is called “The City of Smiles”. Please go and take a look at our city and enjoy your stay here in Aarhus. My team and I have been looking forward to hosting you all and to holding this IEEE conference. We hope this will be a valuable and inspiring conference and visit stay for you.

A key theme in my research field - Multi Business Model Innovation and Technology - is to explain the strategic value of technology for business, its users, customers, employees and networks – and find out how we turn technology into business. Despite many advances in our understanding of the way in which technology contributes to business model value creation and to the betterment of humans, there are still fundamental issues, challenges and opportunities to be addressed in order for us to deal with the challenges of the 21st century landscape. The purpose of GWS 2016 IEEE Summit was to explore new technologies and interdisciplinary businesses models related to future dynamic, agile and real-time value creation. GWS 2016 IEEE Summit addressed key topics of technologies and business models in the 21st century. Security, human bond communication, persuasive technologies, energy and climate saving together with co-creation represent some of the topics in a world of increasing complexity and turbulence.

The GWS 2016 Global Wireless (IEEE) Summit at Aarhus University brought researchers and industry together from different fields and countries to present, debate and reflect on creative approaches to Global Wireless technologies and business models. Throughout the summit the betterment of man will be looked at from an interdisciplinary perspective and will involve the viewpoints of Business and Social Sciences, Science and Technology, Arts and Health. GWS 2016 summits conferences keynote speeches, paper presentations, special sessions, “living business model innovation labs on sustainable energy”, tutorials and many more value propositions gave an unique opportunity to create, capture, deliver, receive and consume knowledge and not least - meet with old and new network partners to discuss existing and new projects.

On behalf of Aarhus University – Business Development and Technology and of my research group MBIT I welcome all of you to Aarhus and to Aarhus University.



**Peter Lindgren**, Professor, Multi Business Model Innovation and Technology



It is a privilege to host the GWS 2016 Summit at Aarhus University, Denmark, and I personally look very much forward to welcome you all to my hometown, Aarhus. The venue of the conference, Aarhus, was ranked second in Lonely Planet's 'Best City in Europe' ranking for 2016, behind the Peloponnese in Greece and ahead of Venice (Italy), so you are in for a treat!

If time, remember to visit the 'The Old Town' a 3-star attraction in Guide Michelin that will take you back in time and which is placed only 500 meters from Aarhus Business School, or take a stroll down the main shopping street, Stroegget, to visit the cathedral of Aarhus, Sct. Clement, which goes back to year 1100 and contains some of the most beautiful and well kept frescoes. So this old city will host the new frontier of technology bridging from the old to the new world. The main theme of this year's Global Wireless Summit GWS 2016 was one of my core research fields - Multi Business Model Innovation and Technology. The main objective of the conference was to bridge wireless technologies and business and to explain the strategic value creation of technology for business and for all stakeholders. Recent research has emphasized the contributions of technology in business model development and in the betterment of humans. However, the frontier of wireless technologies is continuously pushing the boundaries of what we can do for the betterment of humans and in designing new and multiple types of businesses. This conference addressed the opportunities and challenges of wireless technologies and business development in dealing with the key challenges of the 21st century. The GWS 2016 Global Wireless (IEEE) Summit at Aarhus University was a global event that will brought together researchers and industry across from different countries, contexts and fields of expertise in exploration of the future potentials of Global Wireless technologies and business models.

Our team and I are looking forward to hosting this conference and hope that GWS2016 will be a valuable and inspiring conference for you. On behalf of Aarhus University – Business Development and Technology and of our research group MBIT I welcome all of you to Aarhus and to Aarhus University.



**Annabeth Aagaard**, Professor, Multi Business Model Innovation And Technology

# **Keynote Speakers**

## Keynote Speakers



**Anand R. Prasad,**

NEC, Japan

Dr. Anand R. Prasad, Delft University of Technology, The Netherlands, Fellow IETE, is Chief Advanced Technologist, Executive Specialist, at NEC Corporation, Japan, where he leads the mobile communications security activity. Anand is the chairman of 3GPP SA3 and GISFI Security & Privacy group. He has 20+ years of experience in all aspects mobile networking industry in companies around the globe.

Anand has published 6 books, authored 50+ peer reviewed papers and is a passionate speaker on security (Globecom, MWC, RSA etc.). He is editor-in-chief of the “Journal of ICT Standardization” published by River Publishers.

He is recipient of the 2014 ITU-AJ “Encouragement Award: ICT Accomplishment Field” and the 2012 (ISC)<sup>2</sup> “Asia Pacific Information Security Leadership Achievements” (ISLA) Award as a Senior Information Security Professional

### **Keynote Title: 5G Internet of Things Security**



**Seshadri Mohan,**

Professor, Systems Engineering Department, University of Arkansas at Little Rock, USA

Seshadri Mohan is currently a professor in Systems Engineering Department at University of Arkansas at Little Rock, where, from August 2004 to June 2013, he served as the Chair of the Department of Systems Engineering. Prior to the current position he served as the Chief Technology Officer (CTO) and Acting CEO of IP SerVoniX, where he consulted for several telecommunication firms and venture firms, and served as the CTO of Telsima (formerly known as Kinera). At Telsima, he carried out extensive business development with telecommunications and wireless carriers both in the US and India. Prior to launching IP SerVoniX, he was with Comverse, Wakefield, Massachusetts, as Chief Technology Officer. At Comverse, he was responsible for starting and guiding an advanced development center, developing architectures and solutions to facilitate the development of wireless and next generation messaging products, enhancing the patent portfolio, and evaluating technology strategies to pursue and providing

recommendations to the executive management team. Besides these positions, his industry experience spans New Jersey-based Telcordia (formerly Bellcore) and Bell Laboratories. At Telcordia, as a Senior Research Scientist, he led many projects in the area of wireless and next generation technologies, including advancing a multimedia-capable softswitch, in which areas he holds several patents. Prior to joining Telcordia, he was an associate professor at Clarkson and Wayne State Universities, where he developed the communications curriculum and conducted research in computer networking and source coding algorithms.

**Keynote Title: Internet in the Sky**



**Markus Eisenhaur,**  
Fraunhofer FIT, Germany

Markus Eisenhaur studied Psychology and Informatics at the University of Trier in Germany and received a PhD in Cognitive Psychology from the same University. Since 2001 he works at the Fraunhofer Institut for applied Information Technology (FIT) and is leading the Research Department on User Centered Computing, focusing on Ubiquitous Computing, Internet of Things, embedded and Cyber-Physical Systems, as well on M2M and HM-Interaction, Usability and Web-Compliance. Dr. Eisenhaur is the coordinator of several big European Projects and has participated in a multitude of large-scale national funded and European projects (German Industrie 4.0 and European IoT). He works as a Program Chair of international Conferences, Editor of scientific Journals and as Expert for the European Commission.

**Keynote Title: IoT Services for the Betterment of Smart Cities**

Dr. Walter Weigel graduated from the Technical University in Munich, Germany, with the Master Degree in electrical engineering in 1984 and with the Ph. D. degree in 1990. From 1984 to 1991 he was assistant professor at the Institute of Data Processing at the Technical University in Munich.

Dr. Weigel is since 1st April 2015 VP and CSO of the European Research Institute of Huawei, based in Leuven Belgium and Munich Germany.



**Walter Weigel ;**

VP European Research Institute  
of Huawei Technologies  
Duesseldorf GmbH

He used to be from September 2006 to July 2011 the Director General of the European Telecommunication Standards Institute ETSI. Between February 1991 and February 2015 he held several positions within Siemens AG, including VP of External Cooperations and Head of Standardization in Corporate Technology, VP of the Research & Concept-department of the Mobile Networks business unit as well as Head of the business segment Video Processing for the semiconductor business unit (today Infineon).

He is a lecturer at the Technische Universitaet Muenchen, member of the IEEE SA Board of Governors, of the Key Enabling Technologies working group of the European Commission, the Innovationsdialog of the German Government and of the Aalborg University Industry Advisory Council

**Keynote Title: Future Wireless Services Anticipating on Next Generation Wireless Technologies and Systems**



**Hermann Brand,**

Director Innovation, ETSI,  
France

Hermann Brand received his PhD from the Technical University of Vienna. He started his professional carrier as a SW developer and system designer in telecommunications. He worked as a researcher in the semiconductor industry and managed several international R&D teams in mobile communications including researchers, system engineers and delegates to different standards developing organizations. While his employer evolved into a diversified IT service provider, Dr. Brand also worked as technology manager, innovation manager and business developer.

Hermann Brand joined ETSI in 2008. Until 2012 he was responsible for various institutional services of ETSI, including new initiatives, partnership management, membership care, and meeting support. Since 2012, as Director Innovation, he has worked closely with members and other

stakeholders to setup new standardization committees/groups, in particular ISGs.

**Keynote Title: ETSI 5G Standards**



**Leo Ligthart,**  
CONASENSE, Netherlands

**Education**

Doctor of Technology, Delft University of Technology; Thesis: "Antenna Design and Characterization based on the Elementary Antenna Concept", December 1985.  
Engineers degree Cum Laude (M.Sc.), Delft University of Technology; Thesis: "Electromagnetic Reflections from a transparent diffraction grating", June 1969.

**Career/Employment**

1968-2011 : Faculty of Electrical engineering, Delft University ;1976-1977 : Chalmers University Gothenburg, Sweden  
1987-1997 : Advisor-Getronics Telecom and Industrial Electronics; 1988-2011 : Professor Delft University  
1994-2011 : Advisor-Radio communications Rotterdam; 1994-2011: Director IRCTR, Delft University of Technology  
2009-2012 : Guest professor ITB University of Technology Bandung, Indonesia;2011-present : Emeritus Prof. Delft University of Technology  
2011-present : Scientific advisor IRCTR-Indonesia; 2012-present : Chairman CONASENSE  
2012-present : Adjunct- professor Universitas Indonesia, Indonesia;  
2013-present : Adjunct-professor ITS Surabaya, Indonesia  
2013-present : Board of Governors IEEE-AESS; 2014-present : Adjunct-professor Beijing Institute of Technology, China

**Keynote Title: Future Wireless Services Anticipating on Next Generation Wireless Technologies and Systems**



**Steen M. Lynenskjold,**  
Executive Vice President &  
Chief Commercial

Steen Lynenskjold has been employed by Terma since 1989. Prior to his current role, Steen has had several assignments within Terma – latest as Senior Vice President, Defense & Security, which covered the global supply of airborne and naval self-protection systems, SCANTER radar systems, and tactical command and control systems for mission-critical defense and security applications. Steen Lynenskjold has had a leading role in making Terma a more global company. Terma A/S is headquartered at Aarhus Denmark and maintains Danish facilities in Copenhagen and Grenaa.

Internationally, Terma has subsidiaries and operations in The Netherlands, Germany, the U.S., Singapore, United Arab Emirates, United Kingdom, and India.

In 1999, Steen relocated as project manager for 18 months to Seattle, U.S., to support the Boeing Joint Strike Fighter (JSF) aircraft competition within the area of autonomic logistics. In the period 1990-1996, Steen worked in Darmstadt, Germany, to develop command and control systems for satellites used by the European Space Agency. Steen holds an MSc (EE) degree from the Technical University of Denmark in 1987. In addition, he graduated as executive MBA from the Technical University of Denmark in 2002.

Steen is chairman of the Board of Directors of the Alexandra Institute, which is a research-based limited company that bridges the gap between the IT corporate sector, research, and education. He is a member of the IEEE and ACM.

Steen works at the corporate headquarter at Lystrup. He and his wife, Eva, have a daughter, Anna Katrine, and a son, Christoffer.

**Keynote Title: New applications of SCANTER radar technology**



# **Invited Speakers**

## Invited Speakers



**Rajeev Shorey;**

Principal  
Scientist/Researcher, TCS  
Innovation Labs

- 18 years experience in Industry & academia with strong business acumen, technical, analytical, research skills
- Proven ability to combine business, engineering, research skills to develop innovative ideas and solutions
- Strong focus on products, services, applications in industry
- Instrumental in hiring, managing and mentoring individuals and teams of highly qualified professionals in industry and academia in the above areas
- Strong ability to collaborate effectively across institutional and cultural boundaries and ability to lead teams across the globe
- Specialties: Research & Development (Communications/Networking & Computer Science, Data Analytics, Big Data)

**Areas of Interest:**

- o Wireless, Wireline Networks, Internet
- o Cyber Physical Systems
- o Telecommunications & Telematics
- o Communication Security
- o Emerging Applications and Services in Telecom/Telematics Industry
- o Data Analytics and Big Data

**Talk Title: ICT, Capacity Building and Globalisation**



**Marja Anneli  
Matinmikko,**

University Researcher and  
Project Manager, Centre for  
Wireless Communications  
(CWC), University of Oulu,  
Finland

University of Oulu, Centre for  
Wireless Communications (CWC),  
Oulu, Finland,

University Researcher and  
Project Manager since 08.01.2016  
VTT Technical Research Centre of  
Finland, Oulu, Finland

Senior Scientist and  
Project Manager 01.12.2010 –  
07.01.2016

Research Scientist  
01.01.2002 – 30.11.2010

Research Trainee  
12.02.2001 – 31.12.2001  
University of Oulu, Finland

Teaching assistant,  
Industrial Engineering and  
Management 01.02.2000 –  
18.06.2000

(part time 50 %)  
01.09.2000 – 31.12.2000  
Nokia, Oulu, Finland

Logistics coordinator at  
Nokia base station plant  
01.06.2000 – 31.08.2000

Testing in base station production  
17.05.1999 – 20.08.1999

**Talk Title: Sharing economy in 5G to  
enable local micro operators**



**Geir M. Køien,**

University of Agder, Norway

Geir M. Køien is a professor at the University of Agder, Norway, within the field of ICT security.

Currently he is with the Norwegian Defense Research Institute (FFI), where he holds a position as principal scientist. Køien holds a PhD from University of Aalborg, Denmark.

He has previously worked for Ericsson Norway and Telenor R&D amongst others, and has been a delegate to the 3GPP SA3 (Security) standardization group.

He has worked extensively with access security in cellular systems, with security protocol designs, cyber security, privacy aspects and with security designs for system architectures.

He is senior member of IEEE and a senior member of ACM.

**Talk Title: Security by Default: Time to Deliver on a Promise?**



**Silvano Pupolin**

, University of Padova, Italy

Silvano Pupolin (S67, M71, SM83), received the Laurea degree in Electronic Engineering from the University of Padova, Italy, in 1970. Since then he joined the Department of Information Engineering, University of Padua, where currently is Full Professor of Electrical Communications. He was Chairman of the Faculty of Electronic Engineering (1990-1994), Chairman of the PhD Course in Electronics and Telecommunications Engineering (1991-1997), (2003-2004) and Director of the PhD School in Information Engineering (2004-2007). Chairman of the board of PhD School Directors of the University of Padua (2005-2007), Member of the programming and development committee of the University of Padua (1997-2002), Member of Scientific Committee of the University of Padua (1996-2001), Member of the budget Committee of the Faculty of Engineering of the University of Padua (2003-2009),

Member of the Board of Governor of CNIT “Italian National Interuniversity Consortium for Telecommunications” (1996-1999), (2004-2007), Director of CNIT (2008-2010), Chairman of the budget committee of the Department of Information Engineering (2014-present), General Chair of the 9-th, 10-th and 18-th Tyrrhenian International Workshop on Digital Communications devoted to “Broadband Wireless Communications”, “Multimedia Communications” and “Wireless Communications”, respectively, General Chair of the 7th International Symposium on Wireless Personal Multimedia Communications (WPMC'04). He is actively engaged in research on Digital Communication Systems and Brain Communication Interface for motor rehabilitation.



**Peter Farkaš**

Institute of Telecommunications,  
Slovak University of  
Technology in Bratislava  
(STU)

Peter Farkaš is with Institute of Telecommunications, Slovak University of Technology in Bratislava (STU) and also with Institute of Applied Informatics, Faculty of Informatics, Pan European University in Bratislava as a Professor. From 2002 till 2007 he was Visiting Professor at Kingston University, UK and senior researcher at SIEMENS PSE. In 2003 SIEMENS named him VIP for his innovations and patents. In 2004 he was awarded with the Werner von Siemens Excellence Award for research results on two-dimensional Complete Complementary Codes. From 2008 to 2009 he worked also as Consultant for SANDBRIDGE Tech. (USA). He was responsible leader of a team from STU in projects funded by the European Community under the 5FP and 6FP “Information Society Technologies” Programs: NEXWAY IST -2001-37944 (Network of Excellence in Wireless Applications and technology) and CRUISE

(Creating Ubiquitous Intelligent Sensing Environments) FP6 IST-2005- 4- 027738, (2006-2007). His research interests include Error Control Coding, Communications Theory and sequences for CDMA. He published 3 chapters in books 45 papers in reviewed scientific journals and about 98 papers in international conferences. He is author or coauthor of 7 patents. He is and was serving in TPC of about 60 international conferences and presented 12 invited lectures.

As an IEEE volunteer, Prof. Farkaš was serving in IEEE Czechoslovakia Section Executive Committee in different positions from 1992 to 2014 and from 2005 to 2006 he served as a chair of Conference Coordinator Subcommittee in IEEE Region 8. He organized IEEE R8 Conference EUROCON 2001 and was chairman of SympoTIC'03, SympoTIC'04, SympoTIC'06 and co-organizer of Winter school on Coding and Information Theory 2005.

**Talk Title: ARQ Strategy Inspired by Slepian - Wolf Correlated Source Coding**



**Radosveta Sokullu,**  
Turkey

Radosveta Sokullu received her PhD in Telecommunications in 1992 from the Technical University Sofia and in 2000 she joined Ege University as an Assistant Professor. She became Head of Telecommunications Division in 2004. Currently an Associate Professor at the same department she is and also Head of the Wireless Communications Lab. Her research covers different wireless communication networks and protocols (IEEE 802.15.4, Bluetooth, Wireless Sensor Networks, Cellular Networks) with a focus on MAC layer and PHY layer protocol design, resource sharing, resource allocation, and M2M communications. She participated in FP 7 CRUISE project as a leader of the Turkish partners, and has lead a number of projects

sponsored by the National Research Council in Turkey (TUBITAK), by the Ministry of Industry (SANTEZ) in cooperation with large electronic companies in Turkey as well as numerous projects sponsored by the University Research Fund. She is an active member of the International Federation of University Women (IAUW), an IEEE Member and a consultant of the Student IEEE Branch at the Department of EEE.



**Lars Møller Mikkelsen**

Networks and Distributed Systems, AAU

Lars Møller Mikkelsen has a Master of Science (MSc) in Networks and Distributed Systems (2013) from AAU. He is currently finishing his PhD project titled 'Enhancing IoT systems by exploiting crowd sourcing collected information from communication networks', at AAU, Department of Electronic Systems, Section for Wireless Communication Networks. His research interests include IoT connectivity, distributed systems and networks and communication network performance. He has extensive experience in working in EU projects and collaborating with industrial partners from ITS domain. In the ITS project MOBiNET he works with platform development, particularly automated M2M service discovery, and development of an intelligent parking assistant. In the H2020 IoT project BIG IoT he works with use case development, particularly related to public transport optimization based on people density information.

**Talk Title: On the benefits and challenges of crowd-sourced network performance measurements for IoT scenarios**

**Session Tracks  
&  
Workshops**



## **Tracks**

### Technology

- Machine-to-Machine Communications
- Technology Applications
- Future Wireless Networks and Architectures
- Investigation of Detection, Estimation, Precoding, and Delay Tolerance of networks
- Antenna Propagation and Channel Modelling
- ICT Assistance for Medical Systems
- Communication and Coding Theory
- Physical Layer
- Wireless Sensor Network
- Communication, Navigation, Sensing and Services
- Single-Carrier, Multi-Carrier and Broadband Systems
- Reliability, Survivability and Cooperative Communications
- Resource Allocation and Interference Management
- Entrepreneurship and Competence
- Communications
- Security

### Business

- Supporting, Persuasive and Sustainable Business Models & invited Talk
- Unlock Your Personalization towards Quality of life
- Multi business model innovation in a world of human bond communication
- Imperative Business Models for the Technology for Society: Intelligent Systems, Sustainable Energy, eHealth, and Big Data

## **Workshops**

- Workshop on Creative Services, Users and Usable Privacy
- Workshop on eWALL Visions 2017 and beyond
- Biogas 2020 Session-1 &2
- Pearl Project Demonstration

# Program

## Program: Sunday, November 27

### TUTORIAL

Time Slot	Day1: Sunday, November 27   Tutorials	
09:00-18:00	Registration	
<b>Morning Tutorial Session</b>		
09:00-10:30	<b>Tutorial on “Advanced Business Modelling in a world of disruptive technologies ”</b> <b>Peter Lindgren;</b> <i>Multi Business Model Innovation and Technology, Aarhus University, AU Herning, Denmark</i>	<b>HALL-1</b>
10:30-11:00	<b>Coffee Break</b>	
11:00-13:00	<b>Tutorial on “Advanced Business Modelling in a world of disruptive technologies ”(Contd.)</b> <b>Peter Lindgren;</b> <i>Multi Business Model Innovation and Technology, Aarhus University, AU Herning, Denmark</i>	<b>HALL-1</b>
13:00-14:00	<b>LUNCH</b>	
<b>Afternoon Tutorial Session</b>		
14:00-15:30	<b>Tutorial on “Smart UniverCity: A Pathway to World with Thailand as A Case Study”</b> <b>Chayaporn Wattanasiri,</b> Mae Fah Luang University, <i>Thailand</i> <b>Thongchai Yooyativong,</b> Ph.D., Senior Lecturer, School of Information Technology, Mae Fah Luang University, <i>Thailand</i>	<b>HALL-1</b>
14:00-15:30	<b>Tutorial on “Localization/sensing with LTE (in the framework of the CONASENSE initiative) ”</b> <b>Ernestina Cianca;</b> <i>University of Rome Tor Vergata, Italy</i>	<b>HALL-2</b>
15:30-16:00	<b>Coffee Break</b>	
16:00--18:00	<b>Tutorial on “Smart UniverCity: A Pathway to World with Thailand as A Case Study”(Contd.)</b> <b>Chayaporn Wattanasiri,</b> Mae Fah Luang University, <i>Thailand</i> <b>Thongchai Yooyativong,</b> Ph.D., Senior Lecturer, School of Information Technology, Mae Fah Luang University, <i>Thailand</i>	<b>HALL-1</b>
16:00--18:00	<b>Tutorial on “Localization/sensing with LTE (in the framework of the CONASENSE initiative) ”(Contd.)</b> <b>Ernestina Cianca;</b> <i>University of Rome Tor Vergata, Italy</i>	<b>HALL-2</b>
18:00-19:00	<b>Welcome</b>	

## Program: Monday, November 28

### GRAND OPENING

Time Slot	Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions   Part-1	
09:00-18:00	<b>Registration</b>	
09:00-10:15	Grand Opening	
09:00-09:10	<b>Welcome Speech</b>	<b>Ramjee Prasad</b> , Steering Board Chair GWS ,Founder Chair, GISFI
09:10-09:20	<b>Inaugural Speech</b>	<b>Peter Lindgren</b> , General Chair, GWS 2016
09:20-09:35	<b>Opening Speech</b>	<b>Brian Bech Nielsen</b> , Rector (President), Aarhus University
09:35-10:05	<b>Keynote Speech -1</b>	<b>Anand R. Prasad</b> , NEC, Japan; 5G Internet of Things Security
10:05-10:15	<b>Vote of Thanks</b>	<b>Annabeth Aagaard</b> , TPC Co-Chair, Global Wireless Summit 2016
<b>10:15-18:00</b>	<b>END of DAY-1 EVENTS</b>	

## Program: Monday, November 28

Time Slot	Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions	
09:00-10:15	Grand Opening	
	HALL 1	
	KEYNOTE SPEECH	
10:15-11:15	Keynote Speech -2	Leo Ligthart, Conasense, Netherlands; Future Wireless Services Anticipating on Next Generation Wireless Technologies and Systems
	Keynote Speech -3	Seshadari Mohan, University of Arkansas at Little Rock, USA; Internet in the Sky
	HALL-1	
11:15-11:45	Coffee Break	
11:45-13:15	Technical and Business Sessions	
	See the details of Technical and Business Sessions of Day 2 below this page	
	ALL ROOMS/ HALLS	
13:15-14:15	Lunch	
14:15-15:45	Session A-1: Keynote and Invited Talks-1	
	Keynote Speech-4	Walter Weigel , VP European Research Institute of Huawei Technologies Duesseldorf GmbH; State of the art in 5G research
	Invited Talks-1	Rajeev Shorey; Principal Scientist/Researcher, TCS Innovation Labs.
	Invited Talks-2	Marja Anneli Matinmikko , Centre for Wireless Communications (CWC), University of Oulu, Finland; <i>Sharing economy in 5G to enable local micro operators</i>
	HALL-1	
15:45-16:15	Coffee Break	
16:15-17:15	Session A-2: Invited Talks-2	
	Invited Talks-3	Geir M. Køien , Professor , University of Agder, Norway; <i>Security by Default: Time to Deliver on a Promise?</i>
	Invited Talks-4	Silvano Pupolin, DEI - University of Padua, Padova, Italy; <i>Signal Processing in Neuroscience: a Telecommunications view</i>
	HALL-1	
17:30-18:30	Welcome Reception at Town Hall Aarhus	
19:00-19:15	Banquet Speech at Music House in Aarhus; Indian Ambassador to Denmark, Mr. Rajeev Shahare	
20:00-22:30	GALA DINNER at Music House in Aarhus	
	End of Day 2	

<b>Time Slot</b>	<b>Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions   (Continued)</b>	
09:00-18:00	Registration	
11:45-13:15	<b>Session A-3: Machine to Machine Communications-1 (HALL-1)</b>	
	1570314888	<b>Resource Allocation Strategy for Ultra-Reliable Communication in a Factory Environment;</b> Bikramjit Singh and Olav Tirkkonen (Aalto University, Finland); Zexian Li and Mikko Uusitalo (Nokia Bell Labs, Finland)
	1570315717	<b>Linear Wireless Sensor Networks in M2M Communications: MAC Layer Protocol Comparison;</b> Radosveta Sokullu, Eren Demir, Abdullah Balci and Burhan Şeker (Ege University, Turkey)
	1570316804	<b>Decentralized Mobile Cloud Computing using 5G Networks;</b> George Suci (Politehnica University of Bucharest & BEIA Consult International SRL, Romania); Codrin Burla (University Politehnica of Bucharest, Romania); Ioana Marcu (University POLITEHNICA of Bucharest, Romania); Simona Halunga and Carmen Voicu (University Politehnica of Bucharest, Romania)
1570319530	<b>Data Recovery For Workgroup Distribution File System (WDFS);</b> Ravi Upa and Mayoong Yaibuates (Mae Fah Luang University, Thailand); Rungsan Chairicharoen (Mae Fah Luang University & School of Information Technology, Thailand)	
13:15-14:15	Lunch	

	<b>Session A-4: Technology Applications (HALL-2)</b>		<b>Page</b>
11:45-13:15	1570325812	<b>The Influence of Network Neutrality on CONASENSE Innovation;</b> Yapeng Wang (Aalborg University, Denmark)	72
	1570321320	<b>Improving Consideration Performance of Student Loan Fund using Data Clustering Approach;</b> Klangwaree Chaiwut (Mae Fah Luang, Thailand)	85
	1570277648	<b>Overview and challenges with HF RFID systems;</b> Kevin D'hoë, Nobby Stevens, Jean-Pierre Goemaere, Bart Nauwelaers and Lieven De Strycker (KU Leuven, Belgium)	91
	1570321737	<b>Realization of Global Precision Agriculture Monitoring System using WiFi-Based Sensor Nodes and Internet of Things Open Data Platform: A Case Study in Thailand;</b> Thongchai Yooyativong (Mae Fah Luang University, Thailand)	96
	1570324232	<b>Role of Mobile Technology Evolution in Accelerating Digital Transformation of Traditional Businesses and Deriving a Score Card Model that can Estimate the Growth Factor</b> Jeevarathinam Ravikumar (Stanley Black and Decker, United Kingdom); Harshini Arumugam (ACS Egham International School, United Kingdom); Ramjee Prasad (Aarhus University, Denmark)	101
13:15-14:15	Lunch		

Time Slot	Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions   (Continued)		
09:00-18:00	Registration		
11:45-13:15	<b>Session A-5: Future Wireless Networks and Architectures (HALL-3)</b>		<b>Page</b>
	1570325293	<b>Distributed Dynamic Backhauling in Self-Itinerant Intelligent Aerial Radio Architecture;</b> Purnima Lala Mehta (Aalborg University, India); Troels B. Sørensen (Aalborg University, Denmark); Ramjee Prasad (Aarhus University, Denmark)	<b>107</b>
	1570313903	<b>Hierarchical Network Design and Optimal Deployment of LTE Enabled RSUs in VANET;</b> Feng Jiayong, Wu Muqing and Min Zhao (Beijing University of Posts and Telecommunications, P.R. China)	<b>112</b>
	1570287885	<b>Scalable Software-Defined Networking Based on Heap Structure;</b> Yang Liu (Beijing University of Posts and Telecommunications, P.R. China); Zhuqing Jiang (Beijing University of Posts and Telecommunications, P.R. China); Yi Zhao (Beijing Institute of Spacecraft System Engineering, P.R. China); He Shanbao (China Academy of Space Technology, P.R. China); Chan Wang, Aidong Men and Beihang Zhang (Beijing University of Posts and Telecommunications, P.R. China)	<b>117</b>
	1570321566	<b>Towards a configurable state model for the 5G radio access networks;</b> Sofonias Hailu (Aalto University, Finland); Mikko Säily (Nokia Bell Labs, Finland); Olav Tirkkonen (Aalto University, Finland)	<b>123</b>
	1570287854	<b>Enhancing Peer-to-Peer Communications for Future Wireless Systems;</b> Ugljesa Urosevic and Zoran Veljovic (University of Montenegro, Montenegro); Ramjee Prasad (Aalborg University, Denmark)	<b>128</b>
13:15-14:15	Lunch		

11:45-13:15	<b>Session A-6: : Investigation of Detection, Estimation, Precoding, and Delay Tolerance of networks (HALL-4)</b>		<b>Page</b>
	1570321697	<b>A Survey on Live Virtual Machine Migration;</b> Arsch Sharma Vimal, Ashu Saxena and Karthick Nanmaran (SRM University, India)	<b>133</b>
	1570316721	<b>Enhanced Energy Saving Mechanism for Multi-Radio Multi-Channel Wireless Mesh Networks;</b> Samreen Umer, Kenneth N Brown and Cormac J. Sreenan (University College Cork, Ireland)	<b>139</b>
	1570286941	<b>A Topological Map Merging Method Based on Localization of Multi-Robot SLAM;</b> Rui Qu and Zhuqing Jiang (Beijing University of Posts and Telecommunications, P.R. China); Guanghua Zhang (Shenzhen Skycaster Micronics Inc., P.R. China); Beihang Zhang, Chan Wang and Aidong Men (Beijing University of Posts and Telecommunications, P.R. China); Chonghua Liu (Beijing Institute of Spacecraft System Engineering, P.R. China)	<b>145</b>
	1570292461	<b>DDoS Attack Detection in SDN-based VANET Architectures;</b> Anton Katov (Aalborg University & CTIF, Denmark); Martina Todorova and Stamelina Todorova (Aalborg University, Bulgaria); Albena Mihovska (Aalborg Universitet, Denmark); Ramjee Prasad (Aalborg University, Denmark); Vladimir K. Poulkov (Technical University of Sofia, Bulgaria)	<b>151</b>
	1570315724	<b>Coordinated Hybrid Precoding for 3D Millimeter Wave Massive MIMO in 5G H-CRAN;</b> Jialin Cai and Sun Songlin (Beijing University of Posts and Telecommunications, P.R. China); Bo Rong (Communications Research Center Canada, Canada)	<b>157</b>
	1570315928	<b>Precoding Design for Massive MIMO Time Division</b> Gao Wei, Wu Muqing, Zhao Min, Cheng Yanqing Beijing Laboratory of Advanced Information Network Beijing University of Posts and Telecommunications Beijing 100876, P. R. China n Duplex System;	<b>163</b>
13:15-14:15	Lunch		

Time Slot	Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions   (Continued)		
09:00-18:00	Registration		
11:45-13:15	<b>Session A-7: Antenna Propagation and Channel Modelling (HALL-5)</b>		<b>Page</b>
	1570287528	<b>Performance Improvement of Relay System Using Polarization Diversity;</b> Maja Delibasic (University of Montenegro & Research Center for ICT, Montenegro); Milica Pejanovic-Djuricic (University of Montenegro, Montenegro)	<b>168</b>
	1570295045	<b>Estimation of Radio Signal Spatial Local Mean;</b> Pamela Njemcevic (University of Sarajevo, Bosnia and Herzegovina); Vlatko Lipovac (University of Dubrovnik, Croatia)	<b>173</b>
	1570322064	<b>Microstrip Antenna with Inserted Slot for Dual-Polarized;</b> Teeravitit Laohapensaeng (Mae Fah Luang University, Thailand)	<b>178</b>
	1570322285	<b>A Contemporary Survey on Accuracy Demands of Mobile Navigation Systems;</b> Karthick Nanmaran and Amutha Balakrishnan (SRM University, India); Gokul Ashok N (Larsen & Toubro, India)	<b>183</b>
13:15-14:15	Lunch		

Time Slot	<b>Session A-8: Machine to Machine &amp; Green Communications-2 (HALL-6)</b>		<b>Page</b>
11:45-13:15	1570278672	<b>Modeling Service Interaction in M2M Connectivity Management;</b> Evelina Pencheva, Ivaylo Atanasov, Anastas Nikolov and Kamelia Nikolova (Technical University of Sofia, Bulgaria)	<b>189</b>
	1570291407	<b>Research and Experiment of the Indoor Position System Based on Smartphones;</b> Yingxin Lou and Zhimiao Wang (Beijing University of Posts and Telecommunications, P.R. China)	<b>194</b>
	1570308577	<b>Relay Assisted Device to Device Communication: Approaches and Issues;</b> Uyoata Uyoata (University of Cape Town); Mqhele E. Dlodlo (University of Cape Town, South Africa)	<b>199</b>
	1570313406	<b>Energy-efficient Management System for Smart Air Conditioners;</b> Sarang Gosavi (Cyber Advanced Technology, USA); Neeli Rashmi Prasad (Center for TeleInfrastructure (CTIF), USA)	<b>205</b>
	1570316359	<b>Energy-Harvesting Aware Clustering Scheme for Wireless Sensor Networks with Non-uniform Energy-Harvesting rate;</b> Daiki MAEMOTO, Kazuo MORI, Kosuke SANADA, and Hideo KOBAYASHI Graduate School of Engineering, Mie University 1577 Kurimamatiya-cho, Tsu-shi, Mie, 514-8507, Japan	<b>210</b>
13:15-14:15	Lunch		



Time Slot	Day2: Monday, November 28   Grand Opening, Workshops, Special Sessions, Business and Technical Sessions   (Continued)		
09:00-18:00	Registration		
11:45-13:15	<b>Session A-9: ICT Assistance for Medical Systems (HALL-7)</b>		
		<b>Page</b>	
	1570286094	<b>Denosing Images by Variational Mode Decomposition and Spectral Subtraction;</b> Bogdan Gavriloiu, Radu Vizireanu and Octavian Fratu (University Politehnica of Bucharest, Romania); Cristian Barca (University of Pitesti, Romania); Constantin Mara (University of Pitesti & Dacia S.A., Romania); Gheorghe Gavriloiu (University of Pitesti, Romania)	<b>215</b>
	1570291544	<b>Trend Predictive Model of Cardiovascular Complication for Type 2 Diabetes Mellitus with Hypertension Patients;</b> Napa Rachata and Punnarumol Temdee (Mae Fah Luang University, Thailand)	<b>219</b>
	1570294317	<b>Classification of Mental Tasks from EEG Signals using Spectral Analysis, PCA and SVM;</b> Nikolay Neshov, Agata Manolova and Ivo Draganov, (Technical University of Sofia, Bulgaria)	<b>224</b>
	1570315237	<b>Increased Personalization and Intelligent Behavior of Future Workplaces Based on Applied CBR-Based Decisions Support Service;</b> Anton Katov (Aalborg University & CTIF, Denmark); Sofoklis Kyriazakos (Aalborg Universitet, Denmark); Ramjee Prasad (Aalborg University, Denmark)	<b>228</b>
	1570321456	<b>An overview of children activity monitoring framework for attention deficit hyperactive disorder analysis;</b> Tew Hongthong (Mae Fah Luang University & School of Information Technology, Thailand)	<b>233</b>
1570321559	<b>Empowerment of Autistic Children in Learning Experience of Life Skills and Society Adaptation through Digital Interactive Online Media;</b> Banphot Nobaew (Mae Fah Luang University & Mae Fah Luang University, Thailand); Weeranut Intagan (Silpakorn University, Thailand)	<b>237</b>	
13:15-14:15	<b>Lunch</b>		

11:45-13:15	<b>Session A-10: Communication and Coding Theory-1 (HALL-8)</b>		
		<b>Page</b>	
	1570279192	<b>Mobile Surveillance Application for Hybrid WiMAX-Wi-Fi Systems;</b> Smart Lubobya, Mqhele E. Dlodlo and De Jager Gerhard (University of Cape Town, South Africa); Ackim Zulu (University of Zambia, Zambia)	<b>242</b>
	1570287834	<b>Maximizing LTE Network Performance by using Differential Scheduling and Carrier Aggregation;</b> Madalina Oproiu (University Politehnica of Bucharest & Orange Romania, Romania); Vlad Boldan (Network Operational Skill Center-Orange, Romania); Alexandru Vulpe and Ion Marghescu (University Politehnica of Bucharest, Romania)	<b>247</b>
	1570314189	<b>Robust C-RAN Precoder Design for Wireless Fronthaul with Imperfect Channel State Information;</b> Dong Wang, Ying Wang, Ruijin Sun and Xiangyang Zhang (Beijing University of Posts and Telecommunications, P.R. China)	<b>252</b>
	1570316856	<b>Multiple-access performance and interleaver collision in IDMA systems;</b> Olanrewaju B Wojuola (University of KwaZulu-Natal, South Africa); Stanley Mneney (University of St. Joseph in Tanzania, Tanzania); Viranjay M Srivastava (University of KwaZulu-Natal, Durban & South Africa, South Africa)	<b>258</b>
1570321594	<b>A Combined method of LDPC-Polar code for image transmission</b> Tanaporn Payommai (KMUTT, Thailand)	<b>263</b>	
13:15-14:15	<b>Lunch</b>		

## Program: Tuesday, November 29

Time Slot	Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-1		
09:00-18:00	Registration		
	<b>KEYNOTE SPEECH</b>		
09:00-10:45	<b>Keynote Speech -5</b>	<b>Markus Eisenhaur, Fraunhofer FIT, Germany;</b>	
	<b>Keynote Speech -6</b>	<b>Hermann Brand, ETSI, France;</b> ETSI 5G Standards	
	<b>Keynote Speech -7</b>	<b>Steen M. Lynenskjold, Executive Vice President &amp; Chief Commercial Officer (CCO) Terma;</b>	
10:45-11:15	<b>Coffee Break</b>		
11:15-13:15	<b>Session B-1: Invited Talks-3</b>		
	<b>Invited Talks-5</b>	<b>Peter Farkaš , Institute of Telecommunications, Slovak University of Technology, Bratislava (STU) &amp; Frank Schindler Faculty of Informatics, Pan European University Bratislava, Slovakia;</b> <i>ARQ Strategy Inspired by Slepian - Wolf Correlated Source Coding</i>	
	<b>Invited Talk -6</b>	<b>Radosveta SOKULLU, Abdullah BALCI;</b> M2M Communications – promises and challenges	
	<b>Invited Talk -7</b>	<b>Lars Møller Mikkelsen, AAU;</b> <i>The benefits and challenges of crowd-sourced network performance measurements for IoT scenarios</i>	
	<b>Invited Talk-8</b>	<b>Shyam Chakraborty (Trinnect Ltd , Finland);</b> <i>mHealth: taking healthcare to rural India;</i>	
13:15-14:15	<b>Lunch</b>		
14:15-15:45	<b>Technical and Business Sessions Day 3 (Afternoon -1)</b>		
	See the details of Technical and Business Sessions (afternoon) of Day 3 below this page		<b>ALL HALLS/ ROOMS</b>
15:45-16:15	<b>Coffee Break</b>		

<b>Time Slot</b>	<b>Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-2</b>		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-2: Resource Allocation and Interference Management (HALL-1)</b>		<b>Page</b>
	1570319520	<b>ICMP Based IP address Recovery Method for DHCP;</b> Mayoon Yaibuates (Mae Fah Luang University, Thailand); Roungsan Chaisricharoen (Mae Fah Luang University & School of Information Technology, Thailand); Ravi Uptra (Mae Fah Luang University, Thailand)	267
	1570285073	<b>Inter-cell interference avoidance by active coordination of scheduling for downlink on LTE like systems;</b> Ján Kulík (Softca, s. r. o., Slovakia); Peter Farkaš (Slovak University of Technology, Slovakia)	272
	1570320530	<b>Challenges in Wireless Technologies Coexistence in Smart Home Environments;</b> Antonija Marinčić (FER, Croatia); Antun Kerner (Faculty of Electrical Engineering and Computing (FER), Croatia); Dina Simunic (University of Zagreb, Croatia)	277
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-3 Communication and Coding Theory-2 (HALL-1)</b>		<b>Page</b>
	1570316862	<b>Cross-correlation index and multiple-access performance of Gold codes;</b> Olanrewaju B Wojuola (University of KwaZulu-Natal, South Africa); Stanley Mneney (University of St. Joseph in Tanzania, Tanzania); Viranjay M Srivastava (University of KwaZulu-Natal, Durban & South Africa, South Africa)	282
	1570321835	<b>Leveraging TV White Spaces as a Tool for Improved Rural Broadband Connectivity in Developing Countries: An Operational Perspective;</b> Alexander Osei-Owusu (Aalborg University, Denmark); Kwesi Biney-Assan (Ghana Technology University College, Denmark)	286
	1570322217	<b>Performance of Non-binary Tail-biting LDPC Convolutional Code Encoder Design for Image Transmission;</b> Tanaporn Payommai (KMUTT, Thailand)	293
	1570321585	<b>Teaching Chinese as a Foreign Language Teachers' Epistemic Beliefs, Pedagogical Beliefs and Their Beliefs about the Use of ICT in Northern Thailand;</b> Fang Yuan (School of Sinology, Mae Fah Luang University, Thailand); Nikorn Rongbuttsri (School of Information Technology & Mae Fah Luang University, Thailand); Thanon Phunsapphaisan (School of Sinology, Mae Fah Luang University, Thailand)	302
18:00-	End of Day 3		

Time Slot	Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part 3		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-4: Physical Layer (HALL-2)</b>		<b>Page</b>
	1570286761	<b>Extend SDN Network by Multi-controller Based on Fibonacci Heap;</b> Chan Wang and Zhuqing Jiang (Beijing University of Posts and Telecommunications, P.R. China); Yi Zhao (Beijing Institute of Spacecraft System Engineering, P.R. China); Yang Liu (Beijing University of Posts and Telecommunications, P.R. China); Rui Qu and Aidong Men (Beijing University of Posts and Telecommunications, P.R. China); Chonghua Liu (Beijing Institute of Spacecraft System Engineering, P.R. China); Li Zhendong (China Academy of Space Technology, P.R. China)	<b>308</b>
	1570287348	<b>In-Service Assessment of Mobile Services QoE from Network Parameters;</b> Sibila Isak-Zatega (BH Telecom, Bosnia and Herzegovina); Vlatko Lipovac (University of Dubrovnik, Croatia)	<b>314</b>
	1570291893	<b>OFDM Error Floor Based EVM Estimation;</b> Adriana Lipovac (University of Dubrovnik, Croatia); Borivoj Modlic (University of Zagreb, Croatia); Mislav Grgic (University of Zagreb & Faculty of Electrical Engineering and Computing, Croatia)	<b>320</b>
	1570294214	<b>Integrated Microstrip Resonator Network for Multi-band Microwave Filter;</b> Bogdan Gavriloua, Simona Halunga, Octavian Fratu and Radu Vizireanu (University Politehnica of Bucharest, Romania); Gheorghe Gavriloua (University of Pitesti, Romania)	<b>324</b>
	1570316800	<b>Sustainable Routing Protocol To Improve Life Time Of MANET;</b> Manohar S. Chaudhari and Soumya Samal (Technical University of Sofia, Bulgaria); Pavlina Koleva and Vladimir K. Poulkov (Technical University of Sofia, Bulgaria)	<b>329</b>
	1570321570	<b>The New Methodology for Vehicular Network;</b> Krittika Kantawong (Mae Fah Luang University, Thailand)	<b>333</b>
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-5: IoT and WSN (HALL-2)</b>		<b>Page</b>
	1570291858	<b>Primary Evaluation of a Software-defined Security Architecture for an IoT Environment;</b> Stefan-Ciprian Arseni, Alexandru Stancu and Alexandru Vulpe (Faculty of Electronics, Telecommunication and Information Technology; University Politehnica of Bucharest, Romania); Madalina Oproiu (University Politehnica of Bucharest & Orange Romania, Romania); Simona Halunga and Octavian Fratu (Faculty of Electronics, Telecommunication and Information Technology; University Politehnica of Bucharest, Romania)	<b>339</b>
	1570309750	<b>DCDT: Degree-based Multi-hop Clustering Scheme with Directional Packets Transmission for EH-WSNs;</b> Dan Liu (Beijing University of Posts and Telecommunications, P.R. China); Yan Sun (Queen Mary University of London, United Kingdom); Changchuan Yin (Beijing University of Posts and Telecommunications, P.R. China)	<b>344</b>
	1570316846	<b>Effect of GPS errors on Emission model;</b> Anders Lehmann (Aarhus University, Denmark); Allan Gross (University of Aarhus, Denmark)	<b>349</b>
	1570321323	<b>Design and Performance Analysis of RTS/NCTS Based MAC Protocol in IEEE 802.11 Systems;</b> Md. Emdadul Haque (University of Rajshahi, Bangladesh); Faisal Tariq (Queen Mary University of London, United Kingdom); Ghazanfar A. Safdar (University of Bedfordshire, United Kingdom); Yue Chen (Queen Mary University of London, United Kingdom)	<b>355</b>
18:00-	End of Day 3		

Time Slot	Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-4		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-6: Workshop on Creative Services, Users and Usable Privacy -1 (HALL-3)</b>		Page
		<b>Organiser:</b> Knud Erik Skouby, Professor, Director: Professor and founding director of center for Communication, Media and Information technologies (CMI), Aalborg University/ Copenhagen	
	1570321653	<b>Cybersecurity Challenges for SMEs;</b> Samant Khajuria , Knud Erik Skouby, and Lene T. Sørensen; Center for Communication Media and Information Technologies (CMI), Electronics Systems Aalborg University, Copenhagen, Denmark	<b>633</b>
	1570321656	<b>Cybersecurity Challenges for Public Authorities</b> Samant Khajuria , Knud Erik Skouby, and Lene T. Sørensen; Center for Communication Media and Information Technologies (CMI), Electronics Systems Aalborg University, Copenhagen, Denmark	<b>638</b>
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-7: Workshop on Creative Services, Users and Usable Privacy -2 (HALL-3)</b>		
	1570281939	<b>Organiser:</b> Knud Erik Skouby, Professor, Director: Professor and founding director of center for Communication, Media and Information technologies (CMI), Aalborg University/ Copenhagen	<b>HALL-3</b>
18:00-	End of Day 3		

Time Slot	Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-5		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-8: Workshop on eWALL Visions 2017 and beyond-1</b>		<b>HALL-4</b>
	1570285359	Dina Simunic (University of Zagreb, Croatia); Ramjee Prasad (Aalborg University, Denmark); Sofoklis Kyriazakos (Aalborg Universitet, Denmark); Antun Kerner (University of Zagreb, Cote d'Ivoire)	
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-9: Workshop on eWALL Visions 2017 and beyond-2</b>		<b>HALL-4</b>
	1570285359	Dina Simunic (University of Zagreb, Croatia); Ramjee Prasad (Aalborg University, Denmark); Sofoklis Kyriazakos (Aalborg Universitet, Denmark); Antun Kerner (University of Zagreb, Cote d'Ivoire)	
18:00-	End of Day 3		

Time Slot	Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-6		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-10: Communication, Navigation, Sensing and Services (HALL-5)</b>		<b>Page</b>
	1570317388	<b>Detection of offensive lateral movements using finite-state-machine-based patterns;</b> Alessandro Greco and Giuseppe Bianchi (University of Rome "Tor Vergata", Italy)	<b>360</b>
	1570319771	<b>Experimental Low Complexity CubeSat-based Network for Alert Messages Broadcasting;</b> Alessio Fanfani (University of Florence - CNIT, Italy); Sara Jayousi (CNIT University of Florence, Italy); Simone Morosi (University of Florence - CNIT, Italy); Luca Simone Ronga (CNIT, Italy); Enrico Del Re (University of Florence & CNIT, Italy)	<b>366</b>
	1570321660	<b>A coherence study on EEG and EMG signals;</b> Giulia Cisotto (University of Padua, Italy); Umberto Michieli (University of Padua, Italy); Leonardo Badia (Università degli Studi di Padova, Italy)	<b>372</b>
	1570328185	<b>EGNSS High Accuracy System Improving Photovoltaic Plant Maintenance using RPAS integrated with Low-cost RTK Receiver;</b> Marco Nisi, Fabio Menichetti Sistema, tica S.p.a, Via Bramante, 43, 05100 Terni TR, Italy; Bilal Muhammad , Ramjee Prasad CTIF Section, Department of Electronic Systems, Aalborg University, Denmark; Ernestina Cianca CTIF Section, Department of Electronics, University of Rome Tor Vergata, Italy; Alberto Mennella, Graziano Gagliarde TopView srl, Via Santa Croce, 9, San Nicola La Strada, 81020 (CE), Italy Davide Marenchino Entec S.p.a, Via Montegrappa, 2, 12038 Savigliano (CN), Italy	<b>376</b>
	1570321685	<b>Challenges in self-coordinating and adaptive CPS with human in the loop;</b> Vladimir K. Poulkov, Agata Manolova, Krasimir Tonchev and Ognian Boumarbarov (Technical University of Sofia, Bulgaria)	<b>383</b>
	1570320554	<b>Advanced Multi-Constellation EGNSS Augmentation and Monitoring Network (AUDITOR);</b> Esther Lopez, Jacobo Dominguez and Beatriz Quijano (ACORDE Technologies S.A., Spain); Carles Fernández-Prades, Javier Arribas (Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Spain); Manuel Hernández Pajares, Alberto Garcia-Rigo (gAGE/UPC, Spain); Michael Schmidt and Andreas Goss (Technische Universität München, Germany); Emiliano Spaltro Juan Grosso (Alpha Consultants, Italy); Machi Symeonidou, Spiros Fountas and Zisis Tsiropoulos (Draxis Environmental, Greece); Frits Van Evert and Pieter Blok (Wageningen Plant Research, The Netherlands); David Roma (University of Barcelona, Spain)	<b>389</b>
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-11: Single-Carrier, Multi-Carrier and Broadband Systems (HALL-5)</b>		<b>Page</b>
	1570316383	<b>Performance Analysis of Filter Bank Based Multicarrier Waveform for Future Wireless Systems;</b> Deepa. T and T Rama Rao (SRM University, India);	<b>395</b>
	1570316444	<b>Optical Channel Analysis of Turbo coded MIMO-OFDM system for Visible Light Communication;</b> Sabitha Gauni (SRM University, India); Manimegalai Ct (INDIA & Srm University, India); T Rama Rao (SRM University, India)	<b>399</b>
	1570286737	<b>A Hybrid Pedestrian Navigation Based on Activity Sequence Recognition, Map Matching and HMM;</b> Beihang Zhang and Zhuqing Jiang (Beijing University of Posts and Telecommunications, P.R. China); Guanghua Zhang (Shenzhen Skycaster Micronics Inc., P.R. China); Rui Qu (Beijing University of Posts and Telecommunications, P.R. China); Yang Liu (Beijing University of Posts and Telecommunications, P.R. China); Aidong Men (Beijing University of Posts and Telecommunications, P.R. China); Yi Zhao (Beijing Institute of Spacecraft System Engineering, P.R. China); Shanbao He (China Academy of Space Technology, P.R. China)	<b>405</b>
	1570292332	<b>Personalized Learning in a Virtual Learning Environment Using Classification of Objective Distance;</b> Sataworn Chaichumpa (Mae Fah Luang University, Thailand)	<b>411</b>
	1570316596	<b>Maritime wireless multimedia communication system based on Software Defined Radio;</b> Chen Zhenjia (University of Hainan, P.R. China); Yonghui Zhang (Hainan University, P.R. China)	<b>415</b>
18:00-	End of Day 3		

<b>Time Slot</b>	<b>Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-7</b>		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	<b>Session B-12: Reliability, Survivability and Cooperative Communications (HALL-6)</b>		<b>Page</b>
	1570278491	Characterizing the Service Usage of Online Video Sharing System: Uploading v.s. Playback; Chenyu Li, Jun Liu, Dr. and Shuxin Ouyang (Beijing University of Posts and Telecommunications, P.R. China)	420
	1570285305	<b>Characterizing the Content Popularity of Online Video Service;</b> Chenyu Li, Jun Liu, Dr. and Shuxin Ouyang (Beijing University of Posts and Telecommunications, P.R. China)	425
	1570320524	<b>Evaluating QoS in Large-Scale Wireless Sensor Networks in Context of Smart Homes;</b> Ana Koren (FER, Croatia); Dina Simunic (University of Zagreb, Croatia)	430
	1570322062	<b>Load-balanced Clustering Approach for Cooperative D2D Communications in Cellular Networks;</b> Enis Kocan (University of Montenegro, Montenegro); Milica Pejanovic-Djurisic (University of Montenegro & Centre for Telecommunications, Montenegro); Sofoklis Kyriazakos (Aalborg Universitet, Denmark)	435
	1570286718	<b>Energy-efficient two-hop OFDM DF relay system;</b> Enis Kocan (University of Montenegro, Montenegro); Milica Pejanovic-Djurisic (University of Montenegro & Centre for Telecommunications, Montenegro); Sofoklis Kyriazakos (Aalborg Universitet, Denmark)	440
15:45-16:15	Coffee Break		
16:15-17:45	<b>Session B-13: Resource Allocation and Interference Management (HALL-6)</b>		<b>Page</b>
	1570315021	<b>Identifying and Analyzing Hotspots in Urban Cities from Cellular Network Data;</b> Sudong Zhang, Yuanyuan Qiao and Jie Yang (Beijing University of Posts and Telecommunications, P.R. China)	445
	1570316421	<b>Resource Scheduling Approach for LTE-A based Network Incorporating a Moving Relay Node System Equipped Train;</b> Ayotunde Laiyemo (Centre for Wireless Communication, Finland); Harri Pennanen and Pekka Pirinen (University of Oulu, Finland); Matti Latva-aho (UoOulu, Finland)	450
	1570321674	<b>Motion Estimation in Video Sequences by Using Region Based and Optical Flow Algorithm;</b> Narasak Boonthep and Kosin Channongthai Department of Electronic and Telecommunication Engineering King Mongkut's University of Technology Thonburi Bangkok, Thailand	455
18:00-	End of Day 3		

<b>Time Slot</b>	<b>Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-8</b>		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	Session B-14: Pearl Project Demonstration		HALL-7
	Pearl Project Demonstration Organizers: Sofoklis Kariazakos; Anton Nikolaev Katov		
15:45-16:15	Coffee Break		
16:15-17:45	Session B-15: Unlock Your Personalization towards Quality of life (HALL-7)		Page
	1570316568	A Novel Naturalistic Concept Proposal "ATMAN" towards Human Evolution for Personal Bliss, Harmony and Global Peace; Shobha Lala , Ramjee Prasad; Aarhus University, Denmark	460
	1570324236	Towards quality of life - A case study from the Netherlands and Denmark; Ramjee Prasad; Aarhus University, Denmark	465
	1570321630	The Practice and Impact of ATMAN on Human Mind and Body Shobha Lala , Ramjee Prasad; Aarhus University, Denmark	498
18:00-	End of Day 3		

<b>Time Slot</b>	<b>Tuesday, November 29   Workshops, Special Sessions, Business and Technical Sessions   Part-7</b>		
09:00-18:00	Registration		
13:15-14:15	Lunch		
14:15-15:45	Session B-16: Biogas 2020 Session-1		HALL-8
15:45-16:15	Coffee Break		
16:15-17:45	Session B-17: Entrepreneurship and Competence		HALL-8
18:00-	End of Day 3		



## Program: Wednesday, November 30

Time Slot	Wednesday, November 30   Plenary Session, Business and Technical Sessions   Part-1		
09:00-18:00	Registration		
09:00-11:00	<b>Session C-1: Supporting, Persuasive and Sustainable Business Models &amp; invited Talk (HALL-1)</b>		<b>Page</b>
09:30-11:00	1570319761	<b>Supporting sustainable use of persuasive technology in the creation of sustainable business models;</b> Annabeth Aagaard (Aarhus University, Denmark)	<b>503</b>
	1570321728	<b>Human Bond in a world persuasive business models embedded with Disruptive Technologies;</b> Peter Lindgren (Aarhus University Birk, Italy); Annabeth Aagaard (Aarhus University, Denmark)	<b>507</b>
	1570322056	<b>BUSINESS MODEL INNOVATION COMPETENCES "What interdisciplinary competences can really value and do Business Model Innovation?"</b> Peter Lindgren (Aarhus University)	<b>513</b>
	1570316089	<b>Biogas in Denmark and the Wireless Business Model - Discussions on Potential Value Upgrading;</b> Morten Soendergaard (Aarhus University, Denmark)	<b>521</b>
	1570320380	<b>Business strategy on service and data driven multi business model platform in a world of persuasive technologies;</b> Troels Andersen (Aarhus University BSS, Denmark)	<b>526</b>
<b>11:00-11:30</b>	<b>Coffee Break</b>		
11:30-13:00	<b>Technical and Business Sessions Day 4 (Morning)</b>		<b>HALL 1</b>
<b>13:00-14:00</b>	<b>Lunch</b>		
14:00-15:30	<b>Technical and Business Sessions Day 4 (Afternoon)</b>		<b>HALL 1</b>
<b>15:30</b>	<b>END OF GWS 2016</b>		

Time Slot	Wednesday, November 30   Plenary Session, Business and Technical Sessions   Part-2		
09:00-18:00	Registration		
11:00-11:30	Coffee Break		
11:30-13:00	<b>Session C-2 Multi business model innovation in a world of human bond communication (HALL-1)</b>		<b>Page</b>
	1570316833	<b>Applying Persuasive Design for Reducing Urban Unrest;</b> Kristian Løbner (Aarhus University & COWI A/S, Denmark)	531
	1570322041	<b>Technological driven business model innovation: a pragmatic take on human bond communication;</b> John Bang Mathiasen and Jacob Nielsen (University of Aarhus, Denmark)	538
	1570322058	<b>Global Standards, IPR and Patents Business model Ecosystems "An Outlook to future Global SIP BMES Business Models";</b> Peter Lindgren (Aarhus University, Denmark); Krishna Sirohi (I2tb Research Foundation, India); Jesper Bandsholm (Aarhus University, Denmark)	543
	1570322196	<b>Integrating Public Engagement in the Business Model of a Research project: citizen science activities in a robotic project through Action Research lens;</b> Nello Colella (University of Cassino and Southern Lazio, Italy); Francesco Bolici (University of Cassino and Southern Lazio & OrgLab, Italy)	549
13:00-14:00	Lunch		
14:00-15:30	<b>Session C-3: Imperative Business Models for the Technology for Society: Intelligent Systems, Sustainable Energy, eHealth, and Big Data (HALL-1)</b>		<b>Page</b>
	1570316813	<b>Sustainable Business Models for the Adoption of Energy Management Cloud Platforms within Enterprises;</b> George Suci (Politehnica University of Bucharest & BEIA Consult International SRL, Romania); Octavian Fratu (University Politehnica of Bucharest, Romania); Lucian Necula (BEIA Consult International, Romania); Laura Ghenciu (Politehnica University of Bucharest & BEIA Consult International SRL, Romania); Victor Suci (BEIA / UPB, Romania); Carmen Voicu (University Politehnica of Bucharest, Romania)	553
	1570321201	<b>Data Driven Business Models definitions and barriers in a European Electronics Manufacturing Company. - A Preliminary Case study;</b> Torben Bjerrum (Aarhus University, Denmark)	559
	1570316009	<b>5G, an approach towards future telemedicine</b> Sadia Anwar (Aalborg University, Denmark); Tahir Riaz (Radio Analyzer, Denmark); Ambuj Kumar (Aalborg University, India); Ranjee Prasad (Aalborg University, Denmark)	565
15:30	END OF GWS 2016		

<b>Time Slot</b>	<b>Wednesday, November 30   Plenary Session, Business and Technical Sessions   Part-2</b>		
09:00-18:00	Registration		
11:00-11:30	Coffee Break		
11:30-13:00	<b>Session C-4 : Communications (HALL-2)</b>		<b>Page</b>
	1570291946	<b>Sparse Modeling Methods for Misbehavior Detection in LSA Networks;</b> George Papageorgiou, Konstantinos Voulgaris and Constantinos B. Papadias (Athens Information Technology, Greece)	<b>570</b>
	1570316092	<b>DOA Estimation Based Bayesian Compressive Spectrum Sensing Using Dual Polarization Antenna Receiving;</b> Kun Su, Xuekang Sun, Caili Guo and Shiyu Ren (Beijing University of Posts and Telecommunications, P.R. China)	<b>575</b>
	1570316667	<b>Bayesian Cognitive Radio Spectrum Sensing in Multipath Rayleigh Fading Channels</b> Usama Y. Mohamad and Dirk Dahlhaus (University of Kassel, Germany)	<b>580</b>
	1570315926	<b>Multi-current doubling used in multi-band wireless energy harvesting;</b> Chen Xinying, Wu Muqing, Zhao Min Beijing Key Laboratory of Network System Architecture and Convergence Beijing University of Post and Telecommunications Beijing 100876, P.R.China	<b>586</b>
13:00-14:00	Lunch		
14:00-15:30	<b>Session C-5: Security (HALL-2)</b>		<b>Page</b>
	1570287825	<b>Cyber security algorithms for distributed denial of service attack in botnet;</b> Karuna Bhosale (Technical University of Sofia); Maria Nenova and Georgi Iliev (Technical University of Sofia, Bulgaria)	<b>592</b>
	1570287841	<b>Comparative Analysis of Current Security Techniques Employed in Various Attacks of Bluetooth Communication;</b> Varsha Khatod (University of Mumbai, India); Agata Manolova (Technical University of Sofia, Bulgaria)	<b>597</b>
	1570321565	<b>State of the Art Protection Techniques against Denial of Service Attacks In E-commerce Network;</b> Javed Shaikh (Savitribia phule Pune University & SKN SITS, India); Vandana Rohokale (CTIF, Aalborg University & SKN Sinhgad Institute of Technology and Sciences, India); Georgi Iliev (Technical University of Sofia, Bulgaria)	<b>602</b>
	1570316840	<b>User privacy in big data analytics for eHealth;</b> Nidhi N (Vishwaniketan iMeet, India); Albena Mihovska (Aalborg Universitet, Denmark); Ramjee Prasad (Aalborg University, Denmark)	<b>606</b>
15:30	END OF GWS 2016		

<b>Time Slot</b>	<b>Wednesday, November 30   Plenary Session, Business and Technical Sessions   Part-2</b>	
09:00-18:00	Registration	
11:00-11:30	Coffee Break	
11:30-13:00	<b>Session C-6: Special Session on Internet of Things Infrastructure (HALL-6)</b>	
		<b>Organisers:</b> Birger Andersen, Center for Wireless Systems and Applications, Aalborg University, Copenhagen; Knud Erik Skouby & Samat Khajuria, Department of Electronic Systems, Aalborg University, Copenhagen
	1570321228	<b>An Energy-Efficient Link with Adaptive Transmit Power Control for Long Range Networks;</b> Tomasz Blaszczyk (Technical University of Denmark, Denmark); Per Lynggaard (Aalborg University, Denmark)
	1570321654	<b>CCM-R: Secure Counter Synchronization for IoT Wireless Link;</b> Jiachen Yin (Technical University of Denmark); Upal Roy and Birger Andersen (Technical University of Denmark, Denmark)
	1570321684	<b>Comparative study of Internet of Things Infrastructure and security;</b> Bipjeet Kaur (DTU Diplom, Denmark); Bhupjit Singh (DTU, Denmark)
1570287701	<b>Analysis of Deployment Options to Enhance Horizontal Information Sharing and Networking in Internet of Things;</b> A Andrej Mihailovic (King's College London, United Kingdom); Marko Simeunovic (University of Montenegro, Montenegro); Milica Pejanovic-Djurisic (University of Montenegro & Centre for Telecommunications, Montenegro)	
13:00-14:00	Lunch	
14:00-15:30	C-7: for Biogas 2020 Session-2	
15:30	END OF GWS 2016	

**Technical and  
Workshop Papers**

# Resource Allocation Strategy for Ultra-Reliable Communication in a Factory Environment

Bikramjit Singh, Olav Tirkkonen

Department of Communications and Networking,  
Aalto University, Espoo, Finland

{bikramjit.singh,olav.tirkkonen}@aalto.fi

Zexian Li, Mikko A. Uusitalo

Nokia Bell Labs,  
Espoo, Finland

{zexian.li,mikko.a.uusitalo}@nokia-bell-labs.com

**Abstract**—The focus on mission-critical communications, such as industrial automation and public safety, presses the demand for 5G *ultra-reliable communication*. In this paper, we investigate resource allocation to improve the *rate available* to a user in a factory network setting over time and locations. High availability, e.g., 99.999 % can be strived for to reduce outage. This results in a cellular network-based solution for industrial automation. Machines deployed in a multi-user multi-cell network may enjoy different levels of Quality-of-Service due to the environment variation. As a result, these machines have different outage capacities, e.g., at 99.999 % availability target. The network optimizes resource allocation to improve the outage capacities of the disadvantaged machines, and thereby improves the availability in the ultra-reliable communication region.

**Keywords**—5G; mission-critical communication; mission-type communications (MTC); ultra-reliable communication (URC); factory; availability; scheduling

## I. INTRODUCTION

5G will be introduced in the early 2020s, enabling expansion of International Mobile Telecommunications (IMT) that go beyond IMT-2000 and IMT-Advanced services. New services and use cases are also envisioned [1]. These new services are not only for human communication, the significance of Machine-Type Communications (MTC) is predicted to grow. MTC facilitates a broad range of use cases, which can be classified in two main categories. (1) *Massive MTC* provides low-rate connectivity for enormous amounts of small and low-power devices, such as sensors, meters, actuators, etc. and enables new services like building automation, assisted living, logistics tracking, Internet of Things (IoT), etc. [1]. (2) *Ultra-Reliable Communication (URC)* enables real-time control and automation of dynamic processes in, e.g., factory automation, manufacturing and control, smart grid protection, traffic management and safety, or “Tactile Internet” [2]. URC services require very high reliability and often short latencies down to the millisecond level [1], [3], [4].

There are three aspects to reliable use of communication; latency, reliability and availability [5]. It is important to distinguish between reliability and availability. Reliability is related to short-term quality of communication, and is affected by multipath fading and rapidly varying interference. Availability characterizes the service enjoyed by an end-user on longer term. It is affected by path loss, shadowing, and other slowly varying characteristics of the channel. Intuitively, availability of a service is a geographic characteristic, which varies from place to place, and potentially on long time scales, e.g., related to hardware failures. Reliability of a service is temporal characteristic on short time scale, related to varying service quality

and delay. In the literature on mission-critical MTC, the concepts of availability and reliability are often not distinguished. They are referred to with the wholesale term “reliability”.

We explore a 5G design that ensures URC for mission-critical MTC on a unified 5G air interface and network architecture. The most stringent reliability requirement on URC currently being studied is 99.999 % under the radio latency bound of 1 ms [3]. The maximum packet error rate must not be higher than  $10^{-5}$ , where maximum allowable End-to-End (E2E) latency, including jitter/re-transmissions is 1 ms. The error performance depends on path loss, shadowing, fading, interference, etc. Inter-cell interference and fading may be countered using Multiple-Input Multiple-Output (MIMO) technologies, e.g., a diversity of order 4 results in at least 14 dB fading margin [4]. On the other hand, shadow fading poses stringent problems to *availability*, e.g., 99.999 % availability requires a 28 dB margin [4].

Long-Term Evolution (LTE) offers Guaranteed Bit Rate (GBR) that can support packet error down to  $10^{-6}$ . For this the Radio Link Control (RLC) Automatic Repeat Request (ARQ) protocol, combined with physical/Medium-Access Control (MAC)-layer Hybrid Automatic Repeat Request (HARQ) is used. The delay budget, however, goes up to 300 ms [6]. This delay includes radio, transport and core network latencies. In the Radio Access Network (RAN), each ARQ/HARQ round-trip takes at least 8 ms. Thus, LTE systems may guarantee 99.999 % availability, but the RAN delay budget by far exceeds 10 ms [1], [3]. These solutions are accordingly unfit for URC.

In this paper we concentrate on *URC rate availability* in a multi-user, multi-cell network. The availability metric can be viewed as the network’s ability to provide the targeted throughput for 99.999 % of the time to the connected User Equipments (UEs). The factory environment encounters excessive shadow events, e.g., moving cranes or big blocks, and it is very challenging to guarantee the targeted availability. In [7] scheduling policies for ultra-reliable MTC are described, but the performance degrades with high or low node densities. Solutions to combat both shadow and fast fading could be routing diversity resulting from having multiple alternate communication routes to choose from, using multi-connectivity, multi-hop communication, etc. [8], [9]. However, the need of additional infrastructure [8], [9] may limit their usage. We propose a resource allocation mechanism where the latency component is taken into account by considering short 5G frame structures, such as the one envisioned in [10], [11], shown in Fig. 1, and by allowing no re-transmissions. Each UE is characterized with a spectral efficiency distribution, which is constructed over the time. We calculate outage spectral

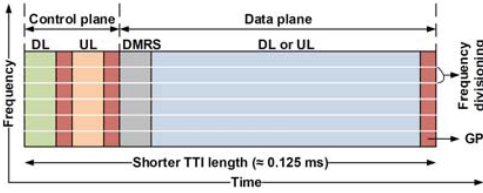


Fig. 1. Envisioned 5G sub-frame design.

efficiency, e.g., for 99.999 % availability for each of the UEs. Based on this information, resources are allocated to achieve better outage performance, and thereby improving availability in the URC region.

## II. SYSTEM MODEL

We assume a network topology consisting of Access Points (APs) and UEs in industrial environments such as factories, shown in Fig. 2. We consider Downlink (DL) communication with low mobility, which works independently of Uplink (UL) communication, e.g., Time Division Duplex (TDD) can be employed. Each UE uses all the bandwidth available to the network and UEs are scheduled based on Time Division Multiplexing (TDM), i.e., each AP provides one DL channel to one selected UE at a time. Time is considered to be slotted to sub-frames. Cell selection is based on Reference Signal Received Power (RSRP).

We do not consider temporal diversity, e.g., in the form of re-transmissions. Accordingly, the delays are related to processing at the AP or server. For DL transmission, we assume Channel State Information (CSI) updated regularly in UL, and CSI feedback latency is not counted towards the URC delay budget. Reserving one sub-frame for processing at the AP, and one sub-frame for the DL transmission, 2 sub-frame transmission delay is incurred in DL. Thus, with a 0.125 ms sub-frame duration shown in Fig. 1, the E2E round-trip delay would be 1 ms or under.

We concentrate on a network scenario with set of APs  $B$  and network load of  $N$  UEs. For AP  $i$  serves a group of UEs  $\mathcal{N}_i$ , its instantaneous load value is  $N_i$  UEs, and  $\sum_{i \in B} N_i = N$ . Each node is equipped with multiple antennas

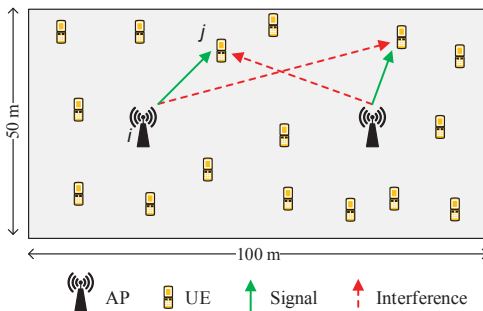


Fig. 2. A factory URC scenario.

for spatial multiplexing MIMO communications. Considering MIMO channels with  $N_a$  antennas at each node, the achievable data rate on a link between AP  $i$  and UE  $j$  is given by<sup>1</sup>

$$r_{ij} = W t_{ij} \log_2 \det (\mathbb{1} + \mathbf{W}_{ij} \mathbf{H}_{ij} \mathbf{H}_{ij}^H \mathbf{W}_{ij}^H \mathbf{X}_j^{-1} p_i) \quad (1)$$

where  $W$  is the available bandwidth,  $\mathbf{H}_{ij}$  is the channel matrix,  $\mathbf{W}_{ij}$  is the precoding matrix,  $\mathbb{1}$  is the  $N_a \times N_a$  identity matrix,  $t_{ij}$  is the time scheduling weight of the data flow and  $p_i$  is the transmission power of AP  $i$  which constrained by a maximum transmission power  $p^{\max}$ , i.e.,  $p_i \leq p^{\max}$ . Finally, the noise-plus-interference covariance matrix at UE  $j$  is

$$\mathbf{X}_j = \sigma_N^2 \mathbb{1} + \mathbf{I}_j$$

where  $\sigma_N^2$  is the noise variance and  $\mathbf{I}_j$  is the covariance matrix representing aggregate inter-cell interference.

The channel matrix  $\mathbf{H}$  and precoding matrix  $\mathbf{W}$  of a link can be represented as

$$\mathbf{H}_{ij} = \begin{bmatrix} h_{ij}^{11} & \dots & h_{ij}^{1N_a} \\ \vdots & \ddots & \vdots \\ h_{ij}^{N_a 1} & \dots & h_{ij}^{N_a N_a} \end{bmatrix}, \mathbf{W}_{ij} = \begin{bmatrix} w_{ij}^{11} & \dots & w_{ij}^{1N_a} \\ \vdots & \ddots & \vdots \\ w_{ij}^{N_a 1} & \dots & w_{ij}^{N_a N_a} \end{bmatrix}$$

where  $h_{ij}^{uv}$  and  $w_{ij}^{uv}$  are the channel coefficient and the applied precoding weight from antenna  $u$  of node  $i$  to antenna  $v$  of node  $j$ . The channel coefficient is  $h_{ij}^{uv} = \sqrt{\Gamma_{ij}} \tilde{h}_{ij}^{uv}$  where  $\tilde{h}_{ij}^{uv}$  represents the fast fading coefficient which is a zero-mean, unit variance circular complex Gaussian random variable. The link gain due to the large scale effects is  $\Gamma_{ij} = \alpha d_{ij}^{-\eta} 10^{-\delta_{ij}/10}$  where  $\alpha d_{ij}^{-\eta}$  represents the distance dependent path loss effects with  $d_{ij}$  the distance between the transmitting and the receiving node,  $\alpha = 10^{-\frac{PL(d_0)}{10}} d_0^\eta$  is the path loss constant and  $\eta$  the path loss exponent. Shadow fading is represented by zero mean normally distributed random variables  $\delta_{ij}$  with standard deviation  $\sigma_S$ .

## III. URC RESOURCE ALLOCATION PROTOCOL

In a factory setting, some machines may face constant degradation in their Quality-of-Service (QoS). The reason could be due to the obstruction from big blocks, nearby machines, robots, etc. In order to avoid starvation of resources to machines with *weak* channels, the network can adjust the scheduling decisions to prioritize their resource allocation. However, while distributing the resources, the communication reliability is kept in mind. The network diverts resources from the better UEs to the weaker ones only if communication reliability of the sacrificing UEs is maintained.

In [8], we strived to improve the availability around the URC region of 99.999 % in a factory setting by means of multi-hop relaying. We identified the outage capacity subject to the URC point, i.e., the rate available at 99.999 % of the network locations. Here, we use the same URC target to define the availability over time across the network coverage area. For example, one machine, availability of more than 1 Mbps over 99.999 % of time, whereas any other may have less than 1 Mbps. The AP considers a scheduling decision reducing the resources allocated to the better UE compared to the weaker

<sup>1</sup>Notations: the transpose, complex conjugate, complex conjugate transpose (or Hermitian), trace and determinant operations for a matrix are denoted by  $(\cdot)^T$ ,  $(\cdot)^*$ ,  $(\cdot)^H$ ,  $\text{Tr}(\cdot)$  and  $\det(\cdot)$  respectively.

ones, and thus improves the rate available in the network at the 99.999 % URC point.

The URC resource allocation protocol works as follows. For each UE  $j \in \mathcal{N}_i$ , we estimate a spectral efficiency Probability Distribution Function (PDF). The PDF is constructed over time whenever a UE is served, based on the spectral efficiency information and collected at the serving AP. The instantaneous spectral efficiency  $\rho_{ij}$  of UE  $j$  connected to AP  $i$  is defined as the achievable rate over unit time scheduling weight,

$$\rho_{ij} = \frac{r_{ij}}{t_{ij}}. \quad (2)$$

Given the availability target  $\mathbb{P}^{\text{rel}}$ , there exists an outage spectral efficiency  $\rho_{ij}^{\text{out}}$  for UE  $j$ , which is

$$\mathcal{F}_{ij}(\rho_{ij}^{\text{out}}) = 1 - \mathbb{P}^{\text{rel}} \quad (3)$$

where  $1 - \mathbb{P}^{\text{rel}}$  is the outage probability and  $\mathcal{F}_{ij}$  the spectral efficiency Cumulative Distribution Function (CDF) of UE  $j$  in cell  $i$ . Due to the varying radio communication environment, UE locations  $\{d_j\}$  have different spectral efficiency CDFs, and therefore different outage spectral efficiencies.

The network is interested in improving URC rate availability, and therefore optimizes rate allocation amongst its UEs. One method is to use max-min scheduling [12]. It targets the worst performing UEs, and alleviates the minimal available rate in the cell. For URC, we modify max-min scheduling as follows. At a given time instant, AP allocates the fractional time resources using max-min protocol but contingent on outage spectral efficiency rather than instantaneous spectral efficiency. Thereupon AP  $i$  serving a UE  $j$  allocates time scheduling weight  $t_{ij}$  according to

$$\{t_{ij} \rho_{ij}^{\text{out}}\}_{j \in \mathcal{N}_i} = C_i \quad (4)$$

where  $C_i$  is a constraint specific to AP  $i$  and  $\mathcal{N}_i$  represents AP  $i$ 's UE set. The sum of scheduling weights fulfills

$$\sum_{j \in \mathcal{N}_i} t_{ij} = 1, \quad (5)$$

so that

$$C_i = \left( \sum_{j \in \mathcal{N}_i} \frac{1}{\rho_{ij}^{\text{out}}} \right)^{-1}. \quad (6)$$

Substituting  $C_i$  to (4), we obtain the scheduling weight for UE  $j$  as

$$t_{ij} = \left( \rho_{ij}^{\text{out}} \sum_{j \in \mathcal{N}_i} \frac{1}{\rho_{ij}^{\text{out}}} \right)^{-1}. \quad (7)$$

As a result, UEs with poor availability in the URC region can enjoy increased allocation of resources. Situations may occur where a UE with rather poor instantaneous spectral efficiency gets less resources, but this only happens if this UE has a better outage spectral efficiency at the URC point than another UE being scheduled. Thus this UE typically enjoys better rates, and can afford to momentarily have a lower rate, so that URC rate availability of the whole network is optimized.

Consider a special case, where the UEs in a cell would have approximately identical spectral efficiency CDFs. From (3), the

outage spectral efficiencies for different UEs would also be approximately the same,

$$\rho_{ij}^{\text{out}} \approx \rho_{ik}^{\text{out}} \quad (8)$$

where  $j, k \in \mathcal{N}_i$ . From (7), we obtain time fraction weight of a UE  $j$  in case of a uniform distributed scenario as

$$t_{ij} \approx \frac{1}{N_i}$$

which is a round-robin scheduling weight.

The benefit over the traditional max-min method is that their the resource allocation is based on instantaneous Signal-to-Noise Ratio (SNR). With the instantaneous condition in consideration when a UE's instantaneous SNR is below URC threshold, network has to give so much resources to this UE that the rates of some other UEs may fall below their URC rates, which makes overall URC performance worse. If we consider max-min application based on the some SNR point within the sample distribution, the diversion of resources may not be as drastic, and the URC performance remains adequate. The reason is that in the proposed scheme the sample extrema is never picked for the max-min application. However, in case of the traditional max-min, the sample extrema could be the instantaneous SNR at some period of time, and during that extent the URC performance gets worse.

#### IV. NUMERICAL RESULTS

To assess the performance of the discussed resource allocation protocol for increasing availability of URC services, we

TABLE I.  
NUMERICAL PARAMETERS

Factory dimension	50 m x 100 m x 5 m
Number of APs/cells ( $ B $ )	2
Number of UEs ( $N$ )	100
Cell selection by UE	RSRP
Carrier frequency	2.4 GHz
Bandwidth ( $W$ )	10 MHz
Bandwidth efficiency	90 %
Number of carriers	1
Power budget ( $p^{\text{max}}$ )	24 dBm
Thermal noise power	-174 dBm/Hz
Noise figure	9 dB
Implementation margin	2 dB
Number of antennas ( $N_a$ )	2
Antenna Pattern	Omnidirectional, uncorrelated, single-polarized
MIMO configuration	2 x 2
MIMO gain	Multiplexing
Precoding type	Singular value decomposition with water-filling
Traffic model	Full-buffer

TABLE II.  
FACTORY USE CASE PARAMETERS [13]

Factory topography	Path loss $PL(d_o)$ [dB] at $d_o = 15$ m	Path loss exponent ( $\eta$ )	Shadowing standard deviation ( $\sigma_S$ ) [dB]
Heavy clutter	80.48	1.69	6.62



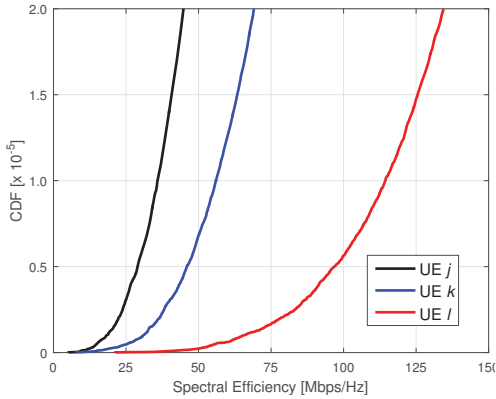


Fig. 3. Distribution of spectral efficiency of some of the UEs at the end of the initialization phase.

consider a factory deployment scenario with two cells. The parameters used for numerical simulations are summarized in Table I. For channel modeling, we apply the heavy clutter model of [13], which is characterized by moderate to heavy losses and is appropriate for transmissions between the AP in the roof and UE on the floor. The factory channel modeling parameters at 2.4 GHz center frequency are listed in Table II. UE locations are uniformly distributed inside the factory hall, and in each instance, UEs are provided with limited mobility around the locations. Path loss, shadowing and fading multi-antenna channels are properly modeled both for own-cell transmissions and inter-cell interference.

First, we consider an initialization phase of 10 000 000 simulation time units. At each time unit, APs keep track of spectral efficiencies for their UEs. Fig. 3 depicts the spectral efficiency statistics of some of the UEs at the end of the initialization phase. We see that the UEs have different outage spectral efficiencies at the URC point, reflecting different average channel conditions in the factory setting. If the resources are distributed evenly amongst the UEs, the result may be poor availability across the network at the URC point.

Next, we evaluate the performance of the proposed scheme in terms of the UE rate distribution over a finite time horizon of 10 000 000 time units following the initialization phase. We strive to maximize the availability at 99.999 % URC point. The gains are depicted in Fig. 4. By applying the allocation scheme, we obtain outage capacity improvement of 900 % at the URC point with respect to the default round-robin allocation. The UEs with the poor outage spectral efficiency at the URC point are allocated more resources at the expense of better performing UEs.

## V. CONCLUSION

A downlink resource allocation scheme is proposed in a factory setting in the context of 5G ultra-reliable machine-type communications. For this, network nodes collect and construct spectral efficiency statistics for each of the machines. Due to excessive shadow events in the factory caused by, e.g.,

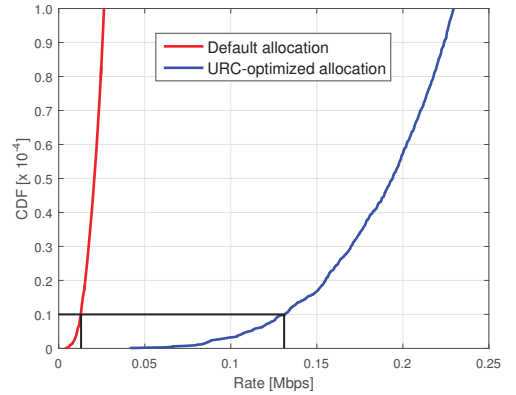


Fig. 4. UE rate distribution in URC region around 99.999 % availability. Default resource allocation vs. URC-optimized allocation.

robots, cranes or big block movements, machines at different locations have different outage spectral efficiencies, which are thus extracted from the statistics with 99.999 % availability target. The scheme tends to exploit this variation while allocating time scheduling weights where machines with poor availability boost their rates by securing the resources from better performing machines, and overall improves the network availability in the ultra-reliable communication region.

## REFERENCES

- [1] ITU-R, "IMT vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Rec. M.2083-0, Sep. 2015.
- [2] G. Fettweis *et al.*, "The tactile internet," ITU-T, Tech. Watch Rep., Aug. 2014.
- [3] 3GPP, "Study on scenarios and requirements for next generation access technologies," TR 38.913 V0.3.0, Mar. 2016.
- [4] B. Singh *et al.*, "Ultra-reliable communication in a factory environment for 5G wireless networks: link level and deployment study," in *Proc. IEEE PIMRC Workshop*, pp. 1-5, Sep. 2016.
- [5] D. Xu *et al.*, "Elastic service availability: utility framework and optimal provisioning," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 55-65, Aug. 2008.
- [6] ETSI, "Speech and multimedia Transmission Quality (STQ); QoS of connections from current technologies to LTE for delay sensitive applications," TR 103 122 V1.1.1, Nov. 2011.
- [7] B. Farayev and S. C. Ergen, "Towards ultra-reliable M2M communication: scheduling policies in fading channels," in *Proc. ICT*, pp. 1-6, Jun. 2016.
- [8] B. Singh *et al.*, "Selective multi-hop relaying for ultra-reliable communication in a factory environment," in *Proc. IEEE PIMRC*, pp. 1-6, Sep. 2016.
- [9] M. Simsek *et al.*, "5G-enabled tactile internet," *IEEE J. Sel. Areas Commun.*, Early Access Article, 2016.
- [10] P. Mogensen *et al.*, "Centimeter-wave concept for 5G ultra-dense small cells," in *Proc. IEEE VTC Spring*, pp. 1-6, May 2014.
- [11] Nokia and Alcatel-Lucent Shanghai Bell, "Flexible numerology for 5G," 3GPP TSG-RAN WG1 Meeting #85, May 2016.
- [12] D. P. Bertsekas and R. G. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, N.J., 1992.
- [13] E. Tanghe *et al.*, "The industrial indoor channel: large-scale and temporal fading at 900, 2400, and 5200 MHz," *IEEE Trans. Commun.*, vol. 7, no. 7, pp. 2740-2751, Jul. 2008.

# Linear Wireless Sensor Networks in M2M Communications: MAC Layer Protocol Comparison

Radosveta Sokullu, Eren Demir, Abdullah Balci, Burhan Şeker  
 Department of Electrical and Electronics Engineering  
 Ege University  
 Izmir, Turkey  
[abdullah.balci@ege.edu.tr](mailto:abdullah.balci@ege.edu.tr)

**Abstract** — A large number of WSN applications like monitoring of pipelines, roads and bridges, as well as recently evolving industrial M2M applications, specifically require linear topologies, which have very different characteristics from traditional WSNs and pose new challenges in terms of overall performance, delay and energy efficiency. Even though there are numerous and also very successful MAC protocols for WSN in general, very few of them perform well with linear topologies. Furthermore, M2M applications have more stringent delay and energy consumption constraints which have to be balanced with simplicity, robustness and compatibility. In this paper we briefly summarize the specifics of LWSN and provide a comparative study of delay and energy efficiency performance of the recently suggested LINE-MAC and the widely used S-MAC protocol. Our simulations of some very typical implementation scenarios show that while the two protocols exhibit similar energy consumption performance, LINE-MAC achieves significant reduction of end-to-end delay and increase in throughput as compared to S-MAC. These results are important because similar underlying scenarios will be quite common for newly emerging M2M applications, where there are much more stringent requirements for delay and energy efficiency than in any applications known so far.

**Keywords**—M2M communications; energy efficiency, delay, linear wireless sensor networks

## I. INTRODUCTION

Machine-to-Machine (M2M) communications is defined in [1] and refers to the technology that enables the communication between different devices (e.g. smartphones, personal health devices and other electronically controlled devices) and allows them to perform a variety of actions without or with only limited human intervention. Because of the extremely large number of devices involved as well as the different communication patterns, the implementation of M2M poses many new challenges related to data transmission, protocol design, system integration, power and spectrum efficiency which make it a very hot research topic. [2 – 7] M2M can be viewed as a myriad of different components (sensors, actuators, middleware, software and applications) that help improve the efficiency and quality of our lives, environment and society. In many environmental and industrial applications, the wireless sensors, which lie in the

bottom of this myriad, are physically deployed in a linear topology. Examples include monitoring of highways, gas, oil and water pipelines; the “Smart Lighting” system, which provides different services from luminosity control for street lightening, lamp monitoring to tracking of elderly people [8]; the PipeNet system which collects pressure and vibration data at high sampling rates to provide near-real-time data pipe monitoring [9], the “Stream-and Waterflood Tracking System” which allows to detect, identify and locate important abnormalities in pipelines without human intervention [10]; railway, bridge and tunnel monitoring [11]. In this paper we refer to such types of networks as Linear Wireless Sensor Networks (LWSNs). LWSNs differ from other WSN in terms of the way communication is carried out. The long chained, single-neighbor transmission pattern, increased network delays and the well-known “relay burden problem” are additional challenges posed by linear topology. To solve these issues, not only routing, but MAC layer protocols should be carefully considered, because they provide effective possibilities to regulate important network performance parameters like throughput, energy efficiency and latency. Since delay and energy constraints in M2M are at least an order more stringent than in WSN known so far, in this paper we focus on examining the advantages of a MAC protocol specifically designed for LWSN (LINE-MAC) as compared to a very efficient but general WSN MAC protocol (S-MAC).

From here on the paper is organized as follows: in section II we briefly introduce the specifics of LWSN and some related work. In section III the two protocols under consideration, namely LINE-MAC and S-MAC, are introduced. Section IV details the examined scenarios followed by the simulation results and conclusions.

## II. DEFINITIONS AND RELATED WORK

### A. Definition of LWSN

The topology of a WSN is determined by the positions of the sensor devices that belong to the network, i.e a LWSN consists of  $n$  sensors and one base station (sink) arranged along a line. The authors of [12] give two different definitions as follows: a) a simple linear network is a connected non-

cyclic graph where each node has exactly one or two distinct neighbors. The two nodes with one neighbor are referred to as the network endpoints; b) a  $(N, d)$ -linear network is a linear network containing  $N$  nodes, each able to communicate with nodes up to  $d$  hops away, where the nodes are denoted by  $n_1, n_2, \dots, n_N$ , with  $n_1$  and  $n_N$  being the endpoints and node  $n_i$  is located between node  $n_{i-1}$  and  $n_{i+1}$  for  $1 < i < N$ . LWSN in themselves can be categorized into several groups: single line (one-dimensional), chain or chain-tree (multiline) networks.

In LWSNs, directional transmission creates significant latency when the end-to-end distance is long. To solve this problem, hierarchical architectures of LWSN have been proposed where nodes have different roles: basic sensor nodes (BSN), data relay nodes (DRN) and data dissemination nodes (DDN). BSN nodes are responsible for sensing and transmitting data to the nearest DRN nodes, which in turn perform data processing and communicate the collected information through a multi-hop process to the DDN nodes. The latter, located at different intervals in the network communicate the information to the network control center. An example of an innovative hierarchical multi-tier architecture for LWSN is presented in Fig. 1 (from [13]).

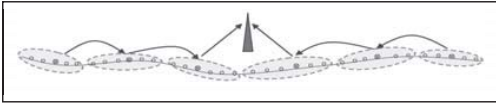


Fig. 1. A three-tier Linear Wireless Sensor Network.

### B. Specifics and Research Challenges for LWSN

In LWSNs data packets are relayed hop-by-hop in a chain-to-one pattern. This creates one of the main problems in LWSNs, referred to as the “relay burden problem”. It results in very unequal energy consumption among the nodes and leaves the ones close to the sink with much less energy, exacerbated by the reduced number of neighbors as compared to general WSN. Therefore the risk of prematurely terminating the network’s operation is greatly increased. On the other hand, long sleep times, i.e short duty cycle solutions are not acceptable because these nodes have extended relaying functions. The “relay burden problem” creates higher congestion for nodes closer to the sink, making channel access more difficult and increasing buffer overflows and packet drops. Assigning bigger buffers to the nodes closer to the sink is a simple solution but not very practical for off-the-shelf components with standard configurations. Heterogeneity is an alternative, which however leads to increased implementation costs. Another important issue is that data delivery is more exposed to failure as compared to general WSNs since packet delivery relies on a more limited number of relay nodes. A single node failure can totally disturb the communication process in the network which is a strong weakness of LWSNs. For M2M related industrial and control applications these constraints are even more important and more stringent.

Despite these challenges LWSNs offer some potential benefits. Since the nodes know their neighbors they can schedule packet transmission ahead and regulate their duty cycle accordingly. In many cases nodes are deployed at equal distances which create advantages in positioning and

synchronization [14]; random node deployment along a line or chain provides advantages in forming clusters. As the topology is already known, additional control overhead for network discovery is minimal. Thus, high energy consuming techniques like flooding are not required in LWSN.

### C. MAC Layer Protocols for LWSN

With the increased number of applications requiring linear topologies, there have been quite a few MAC protocols designed for LWSN and in a previous work we have provided a survey of the most interesting ones. [15] Among them we can mention DiS-MAC [16] which guarantees collision-free communication between synchronized sensor nodes; LC-MAC [17] is a duty cycle medium access control protocol that exploits a mechanism for relay nodes booking in advance and transmitting in a burst manner in order to reduce the end-to-end delivery delay; CMAC-T [18], a chain-type medium access control protocol based on tokens, designed and implemented for monitoring crop’s growing environment; MFT-MAC (Multi frame transmission MAC) [19], a contention-based duty cycled MAC protocol using a synchronized approach to data transmission based on declaring the numbers of data frames to be sent; WiWi [20], a contention-free MAC protocol based on fixed size packets, following a time staggered transmission schedule; and LINE-MAC [21], a LPL (low-power listening) protocol with adaptable duty cycle and “forward wake-up” mechanism. Detailed comparison and evaluation of these protocols can be found in [15].

Until recently wireless sensor networks have been considered and studied as a stand-alone technology, serving specific applications, connecting through a sink or gateway to another data network or actuator network. Both cases differ from the role WSN are envisaged to play in the context of M2M communications, where they will be more of an “invisible”, “underlying” enabler for more complex systems, excluding human intervention. On the one hand this new role places much more tight requirements on delay and energy consumption and on the other hand it also poses the need for greater simplicity and uniformity in the WSN. That is why, well-established, well-tested and simple protocols, will be in demand again. Some of the protocols discussed in [15] really provide quite optimal solutions for LWSNs but are too specific and elaborate to be considered in the context of M2M communications. That is why in this work we provide a comparative performance evaluation of two protocols representative of the two categories mentioned above: a very generic WSN protocol, S-MAC and a simple but quite efficient LINE-MAC designed for LWSNs. Of major interest are the energy consumption and end-to-end-delay. The comparison is carried out based on communication scenarios representative of a large group of applications. In the following section we provide a brief overview of the two MAC protocols and then in section IV continue with the detailed description of the scenarios and the simulation results.

## III. LINE-MAC AND S-MAC OVERVIEW

### A. LINE-MAC

LINE-MAC, previously described in [21], is a duty cycle based protocol using low power listening (LPL). Nodes follow

a dynamically adaptive sleep-wake-up duty cycle; nodes send series of very short preambles and wait for a preACK from downstream neighbors before sending the data itself. Upon awake a node listens for preambles and if none is detected goes to short sleep. If it receives a preamble it immediately sends a short preACK reply and stays awake to receive the number of packets specified in the preamble. When a next hop neighbor overhears a preACK it can deduct how long the data transmission is and sets its timer accordingly to receive the data during the next hop (long sleep). This mechanism of “forward wake-up” provides both distributed synchronization and allows minimizing the hop-to-hop delay. The fact that preambles are very short allows each node to conserve energy while still communicating with its neighbors. The suggested mechanism also has provisions for varying the number of data packets to be sent. It results in gracefully minimizing (or totally avoiding) the “relay burden problem” since every next node will know how many data packets are to be transmitted and will stay awake for the required transmission interval. Previous work has shown that LINE-MAC is simple to implement while at the same time alleviates the 3 major consequences of the “relay burden problem”: reduces the end-to-end delay, minimizes the consumed energy, and promotes network lifetime by ensuring equal distribution of the consumed energy.

### B. S-MAC

S-MAC, is a well-established MAC protocols for WSNs. [22]. It uses a coarse-grained sleep/wakeup cycle, allowing nodes to sleep most of their time. Communication is organized in frames, where each frame begins with a listen period (coordination among nodes that have data to send), followed by a sleep period. In S-MAC, all nodes are free to choose their own listen/ sleep schedules and share them with their neighbors to enable communication between all nodes. Nodes schedule their transmissions during the listen time of their destination nodes. If a node  $a$  wants to send to  $b$ , it just waits until  $b$  is listening. S-MAC enables multi-hop operation by accepting multiple schedules and encourages neighboring nodes to adopt identical ones for reducing overhead. Collision avoidance uses CSMA with an RTS-CTS-DATA-ACK sequence exchange. However, the low-duty-cycle operation forces the nodes to delay sending a packet until the next listen period of the destination (next hop), which increases latency. In essence, the S-MAC protocol minimizes the energy consumption but at the expense of reduced throughput and increase delay.

The two protocols were specifically chosen because they have a lot in common. Both rely on a sleep-wake up duty cycle to conserve energy. Both adopt a distributed mechanism to synchronize the schedule of neighboring nodes in order to minimize the transmission delay of a packet travelling from hop-to-hop. However, LINE-MAC benefits from the additional knowledge of who the next one- and two-hop neighbors are to ensure longer sleeping times for the nodes and an efficient staggered wake-up pattern. Furthermore in place of the 4-way exchange procedure of S-MAC, LINE-MAC relies on few very short preambles and a single preACK to ensure the transmission of several data packets. Both features incorporate additional knowledge of the linear topology. In the next sections we will discuss in detail how these slight differences

in the MAC protocol significantly affect the delay and the consumed energy of the network as a whole.

### C. Communication Model and Transmission Scenarios

In the following we describe the communication model and scenarios that are used in the simulations for comparing the performance of the two protocols for LWSN.

We consider a linear WSN consisting of  $N$ , equally spaced nodes indexed by  $0, 1, 2, \dots, (N-1)$ , with fixed communication range. The node with index 0 is always the sink, and all nodes except the sink can sense, transmit and receive, i.e. they are both source and forwarding nodes. All nodes are fixed and know their locations as related to their ID. The density of the nodes is high enough, so that a node can directly communicate with two neighbors on each side. A node connects to the sink directly, if within range or by multi-hop communication. The topology of the sensor network examined in this paper is based on a multi-hop linear network. Each source node generates packets and transmits them through a single-hop or multi-hop communication in order to reach the sink. Node  $j$  is called *downstream 1-hop-neighbor of node  $k$*  if its ID number is one less, a *downstream 2-hop-neighbor for  $j$* , is the one whose ID number is two less than the ID  $j$ . Neighbors in the direction further than the sink are upstream neighbors, respectively *upstream 1-hop neighbor* and *upstream 2-hop neighbors*. For identical, off-the-shelf nodes transmission energy is biggest depletion operation. So we focus on regular topology to allow explicit delay and energy consumption comparison while varying the transmission range. The extra level (2-hop transmission) is the simplest hierarchy which allows control of the overall latency.

***N1 scenario:*** all nodes generate packets (with the data they sense) and relay packets from their upward neighbors. Nodes transmit only to their downward 1-hop neighbors.

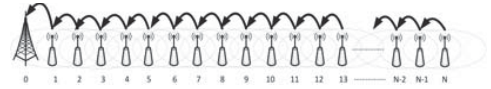


Fig. 2. Schematic representation of N1 scenario

***N1N2 scenario:*** all nodes generate packets and relay packets, however in this case packets can be sent downwards both to 1-hop and to 2-hop neighbors.

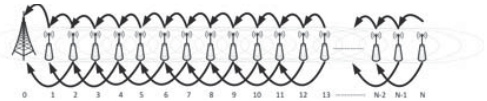


Fig. 3. Schematic representation of N1N2 scenario

## IV. SIMULATION RESULTS AND DISCUSSION

In this section we present the results of simulating the two protocols using Castalia. Our goal is to compare the performance in terms of energy consumption, delay and throughput.

First of all the major goal of these simulations is to test the two protocols for a wider range of traffic intensity and a larger number of nodes as compared to previous works [21,22]. In the

simulations, the traffic intensity is varied based on the packet inter-arrival time, while the maximum number of nodes in the network is 100. Another important point is that the two protocols have different control overheads (in terms of additional control packets sent), so, in order to provide a fair comparison the consumed energy is given per packet sent ( $EC/packet$ ). Fig. 4 shows the consumed energy per packet for the two protocols. In both scenarios LINE-MAC is more energy efficient and has less variance in  $EC/packet$ . S-MAC exhibits higher energy consumption and range variation for the N1 scenario but is less affected by the traffic intensity in scenario N1N2. Another very important conclusion is that restricting the transmission range (1-hop neighbors) does not provide any energy savings - both S-MAC and LINE-MAC have higher  $EC/packet$  for the N1 scenario.

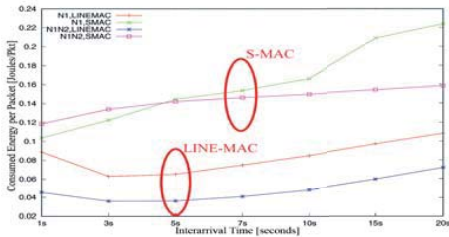


Fig. 4. Consumed energy per packet comparison for scenario N1 and N1N2

Fig.5 and Fig.6 provide detailed results for the  $EC/packet$  per node for the two scenarios. The family of graphs corresponds to different inter-arrival times. The consumption per node for S-MAC does not depend on the scenario, while LINE-MAC shows greater variability especially for the cases of high traffic (1s inter-arrival time).

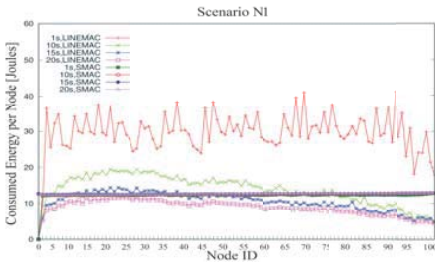


Fig. 5. Consumed energy per node comparison for scenario N1

The performance of LINE-MAC is very dependent on the scenario and N1N2 scenario provides much better results regarding consumed energy per node. An important point is that both scenarios evade the “relay burden problem”, a major reason for depleting the energy of the nodes close to the sink.

A second group of experiments was carried out to examine the delay performance of the two protocols. Fig. 7 shows the CDF of latency for both protocols and both scenarios. All latencies are evaluated at the application layer. It clearly shows

that LINE-MAC provides consistently lower delays, where the majority of the packets arrive with up to 5 s delay.

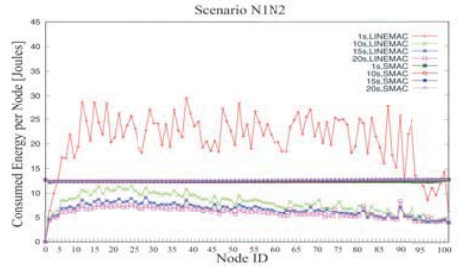


Fig. 6. Consumed energy per node comparison for scenario N1N2

On the other hand, the spread of delay for S-MAC is quite high for lighter traffic ( $\approx 30s$ ) because nodes tend to sleep longer. Again, the N1N2 scenario provides better results. The obtained results for S-MAC, while keeping the general trend, vary slightly in values from previous ones published in [22]. This is due to the fact that in our simulations we are using a Chipcon Radio CC2500, while in those simulations TR3000 was used.

A third group of experiments was aimed at comparing the throughput performance. Previous studies were done with a much smaller number of nodes (10) and much smaller number of packets [22]. In our work with 100 nodes, during simulations  $\approx 180000$  packets are sent in each run in the heaviest traffic case. Table 1 below shows that the packet delivery ratio (PDR) for S-MAC is consistently lower than that of LINE-MAC. Even in the worst case (heaviest traffic) LINE-MAC, with only 18.8% of delivered packets, performs nearly 3 times better than S-MAC. However, for light traffic both protocols performance is in the 90% range (Table 1).

TABLE I. PDR comparison for scenario N1 and N1N2

INTERARRIVAL TIME	Number of Generated Packets	SCENARIO N1		PDR for LINEMAC	PDR for SMAC	SCENARIO N1N2		PDR for LINEMAC	PDR for SMAC
		Number of Received Packets				Number of Received Packets			
		LINEMAC	SMAC	LINEMAC	SMAC				
1s	179960	33832	12168	18.80%	6.76%	46443	10628	25.81%	5.91%
3s	59960	39002	10293	65.05%	17.17%	43591	9264	72.70%	15.45%
5s	35960	32335	8726	89.92%	24.27%	33834	8788	94.09%	24.44%
7s	25700	23494	8197	91.42%	31.89%	24195	8544	94.14%	33.25%
10s	17960	16502	7578	91.88%	42.19%	16869	8342	93.93%	46.45%
15s	11960	10986	6036	91.86%	50.47%	11235	8099	93.94%	67.72%
20s	8960	8252	5623	92.10%	62.76%	8387	7989	93.60%	89.16%

## V. CONCLUSION

This paper presents a comparative study of two duty-cycle based MAC protocols for applications of WSN which specifically exhibit linear topologies – S-MAC is a protocol well-known from different applications of general WSNs, while LINE-MAC is a recently proposed MAC protocol designed to meet the needs of LWSNs. The main purpose of the comparison is to examine the performance of the two protocols in light of the newly emerging M2M applications that have much more stringent delay and energy constraints as compared to well-known WSN.

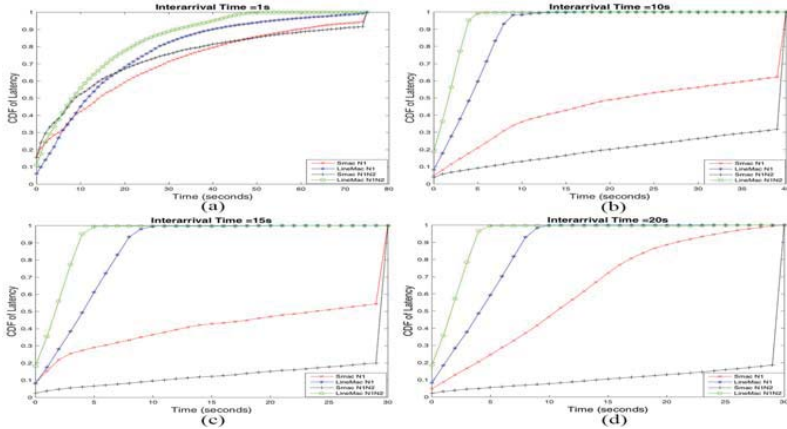


Fig. 7. Latency histogram for (a) 1s interarrival time (b) 10s interarrival time (c) 15s interarrival time (d) 20s interarrival time

The results show that in terms of consumed energy, under varying traffic intensity, both protocols have similar performance; while in terms of delay S-MAC performs quite well under heavy traffic but exhibits unacceptably large delays for lighter traffic. As far as the throughput is concerned LINE-MAC definitely outperforms S-MAC. The obtained results will direct researchers to a more informed decision when the necessity arises for specific implementation in the M2M environment.

REFERENCES

[1] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. D. Johnson, "M2M: From mobile to embedded internet," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 36-43, Apr. 2011.

[2] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: Potential, challenges, and solutions," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 178-184, Mar. 2012.

[3] L. Atzori, A. Iera, G. Morabito, "The Internet of Things: a survey," *Computer Networks*, vol.54, no.15, pp. 2787-2805, Oct., 2010.

[4] K. Chen, "Challenges and opportunities of internet of things", in 17th Asia and South Pacific Design Automation Conference (ASP-DAC), Sydney, Jan, 2012, pp. 383-388.

[5] J. Kim, J. Lee, J. Kim, J. Yun, "M2M Service Platforms: Survey, Issues, and Enabling Technologies," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 61-76, 2014.

[6] Machine-to-Machine Communications (M2M); Functional Architecture, ETSI TS 102 690 v2.1.1, Oct., 2013.

[7] S. C. Jha, A. T. Koc, M. Gupta and R. Vannithamby, "Power Saving Mechanisms for M2M Communication over LTE Networks," in First International Black Sea Conference on Communication and Networking (BlackSeaCom), Batumi, July, 2013, pp. 102-106.

[8] F. Chen, L. Hao, and Q. LiLi, "LC-MAC: An Efficient MAC Protocol for the Long-Chain Wireless Sensor Networks," *Third International Conference on Comm. and Mobile computing*, 2011, pp. 495-500.

[9] I. Stojanov, L. Nachman, S. Madden, T. Tokmouline, and M. Csail, "PIPINET: A Wireless Sensor Network for Pipeline Monitoring," in *IPSN 2007. 6th International Symposium on*, 2007, pp. 264-273.

[10] S. Yoon, W. Ye, J. Heidemann, B. Littlefield, and C. Shahabi, "SWATS: Wireless sensor networks for steamflood and waterflood pipeline monitoring," *IEEE Network*, vol. 25, no. 1, pp. 50-56, Jan. 2011.

[11] S. Kyung Bok, Y. Jae Jun, and K. Do Hyun, "Collision Warning System on a Curved Road Using Wireless Sensor Networks," in *IEEE 66th Vehicular Technology Conference*, 2007, pp. 1942-1946.

[12] K.J. Henry, D.R. Stinson, "Linear approaches to resilient aggregation in sensor networks", *Journal of Mathematical Cryptology*, vol. 9, no. 4, pp. 245-253, Dec 2015.

[13] C.W. Chen, Y. Wang, "Chain-Type Wireless Sensor Network for Monitoring Long Range Infrastructures: Architecture and Protocols", *International Journal of Distributed Sensor Networks*, vol. 4, no. 4, pp. 287-314, 2008.

[14] N. Wang, et al. "Research on linear wireless sensor networks used for online monitoring of rolling bearing in freight train," *Journal of Physics: Conference Series*. vol. 305, no. 1. IOP Publishing, 2011.

[15] R. Sokullu, E. Demir, "MAC Layer Protocols for Linear Wireless Sensor Networks: A Survey", 13th International Conference on Telecommunications and Informatics, Istanbul, 2014 pp, 247-256.

[16] T. Karveli, K.Voulgaris, M. Ghavami, A.H Aghvami, "DiS-MAC: A MAC protocol for sensor networks used for roadside and highway monitoring," *International Conference on Ultra Modern Telecommunications & Workshops*, Oct. 2009, pp. 1-6.

[17] C. Fang, H. Liu, L. Qian, "LC-MAC: An Efficient MAC Protocol for the Long-Chain Wireless Sensor Networks", *Int. Conf.on. Communications and Mobile Computing*, International Conference, 2011, pp. 495-500.

[18] T. Sun, X. J. Yan, Y. Yan. "A Chain-type Wireless Sensor Network in Greenhouse Agriculture". *J of Computers*, pp. 2366-2373, 2013.

[19] E. Lee, J. W. Jwa, and H. Kim, "MFT-MAC: A Duty-Cycle MAC Protocol Using Multiframe Transmission for Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 9, Sept. 2013.

[20] D. Caneva and P. Montessoro, "A Synchronous and Deterministic MAC Protocol for Wireless Communications on Linear Topologies," *Int'l J. of Communications, Network and System Sciences*, Vol. 3 No. 12, pp. 925-933, 2010.

[21] R. Sokullu, E. Demir, "Investigating energy Efficiency and Timeliness for Linear Wireless Sensor Networks with Dense Topologies", *Procedia Computer Science Elsevier*, vol. 37, no. 1, 2014, pp. 24-31, doi: 10.1016/j.procs.2014.08.008.

[22] W. Ye, J. Heidemann, D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks", *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, pp. 493-506, June 2004.

# Decentralized Mobile Cloud Computing using 5G Networks

George Suci  
 R&D Department  
 BEIA Consult International  
 Bucharest, Romania  
 george@beia.ro

Codrin Alexandru Burla, Ioana Marcu, Simona  
 Halunga, Carmen Voicu  
 Telecommunication Department  
 University POLITEHNICA of Bucharest  
 Bucharest, Romania

*Abstract*— Nowadays, a lot of businesses are moving to the cloud, because cloud computing increases efficiency, improves cash flow, offers scalability, flexibility, disaster recovery, security etc. However, there are challenges regarding communication bottlenecks so that cloud resources can be made available on mobile devices. Considering the advantages of the cloud and the exponentially increased capabilities of the 5th generation (5G) mobile networks, MCC (mobile cloud computing) will enable the development of applications which can change people's life styles. In this paper the authors present the implementation of a cloud platform using 5G network capabilities implemented on a USRP (Universal Software Radio Peripheral) hardware equipment with GNU Radio software. We propose a decentralized cloud computing solution that can be used for enterprise data backup or resilience using small devices, like smartphones and tablets. Finally, a series of experiments and tests are conducted in order to evaluate the platform's performance.

*Keywords*- MCC; 5G; USRP; GNU Radio; Cloud.

## I. INTRODUCTION

With the exponentially increased capabilities of 5th generation (5G) mobile networks, Mobile Cloud Computing (MCC) will become even more powerful and will develop to such an extent that it is anticipated that it will change people's life styles and patterns. As of today there are over six billion connected devices that can benefit from cloud-based applications. The evolution towards 5G is considered to be the convergence of Internet services with legacy mobile networking standards, leading to what is commonly referred to as the 'mobile Internet' over Heterogeneous Networks (HetNets), with very high connectivity speeds [1].

In this paper we present a Software Defined Radio (SDR) approach for 5G networks based on Universal Software Radio Peripheral (USRP) and GNU Radio, an open-source software that offers tools for developing software and offers modules for digital signal processing used for implementing radios defined virtually. GNU Radio can be interconnected to real hardware equipment such as RF hardware platforms in order to create SDR equipment. GNU Radio applications are written in Python programming language, and functions are implemented in C++.

The paper is organized as follows: in Section II there are introduced some general aspects of Cloud Computing; Section III contains the description of networking concepts related to 5<sup>th</sup> generation mobile network, while Section IV consists of the

implementation of the cloud platform including results and future applications.

## II. RELATED WORK ON CLOUD COMPUTING

Cloud Computing formal definition enacted by the National Institute of Standards and Technology NIST is as follows "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2]. In this section we present the principles, deployment and service models of cloud computing, as well as a generic architecture.

### A. Principles of Cloud Computing

In this section, the main five essential characteristics of cloud computing are presented:

- On-demand self-service
- Broad network access
- Elastic resources pooling
- Rapid elasticity
- Measured service

### B. Deployment Models

Deployment refers to the way that the cloud services are made available to the end users depending on factors such as location and structure. The main deployment models are:

- Private cloud
- Public Cloud
- Community Cloud
- Hybrid Cloud

### C. Service Models

There are 3 kinds of services described in [2] that are available to end users in regards to cloud computing applications:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

*D. Cloud Computing Architecture*

The Cloud architecture describes the functional mode in which the cloud works. It includes components and services that are used. The cloud is a technology completely dependent on an Internet connection to function.

According to [3], the cloud can be divided into 4 layers, as presented in Fig. 1:

- Layer 1 – User/Client Layer - the place where the client/user initiates the connection to the cloud;
- Layer 2 – Network Layer - facilitates the connection of the user to the cloud;
- Layer 3 – Cloud Management Layer - represents all software used in managing a cloud. The software is typically an operating system.
- Layer 4 – Hardware Resource Layer - represent the actual hardware resources available.

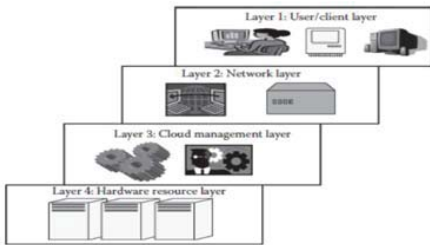


Figure 1. Generic cloud computing architecture

All these layers require virtualization of resources, such as SDR and Software Defined Networking (SDN), in order to provide the deployment and service models of cloud computing.

III. PROPOSED METHODS FOR 5<sup>TH</sup> GENERATION MOBILE NETWORK

Wireless technologies have become an important part of our daily life and have a profound impact in our daily task giving us access to a full range of services for multimedia (videos, video-conferences, images), information services (encyclopedias, academic content) and access to several applications used for e-commerce, health emergency applications. „If analysts’ prognostications are correct, just about every physical object we see (e.g. clothes, cars, trains, etc.) will also be connected to the networks by the end of the decade (Internet of Things)” [1]. 5G gives us not only an upgrade in speed but implements a power and cost effective wide area coverage network powered by HetNet’s.

*A. Standards and Features*

The Next Generation Mobile Network Alliance (NGMN) defines in the white paper for 5G requirements for a full functioning 5G Network [4]. In summary the NGMN suggest the following:

- ✓ Data rates of up to 1Gb/s should be supported in specific environments such as indoor offices, while at least 50 Mb/s shall be available everywhere cost-effectively.

- ✓ The 5G system should provide 10ms E2E latency (duration between the transmission of a small data packet from the application layer at the source node and the successful reception at the application layer at the destination node plus the equivalent time needed to carry the response back) in general and 1ms E2E latency for the cases that require very low latency. The end user should have the perception that he is always connected. The establishment of initial connection to the network should be instantaneous from the perspective of the user.

- ✓ In case of mobility, 5G should not assume mobility support for all devices and services but provide mobility on demand only to those devices and services that need it.

The above parameters are summarized in Fig. 2 below:

Use case category	User Experienced Data Rate	E2E Latency	Mobility
Broadband access in dense areas	DL: 300 Mbps UL: 50 Mbps	10 ms	On demand, 0-100 km/h
Indoor ultra-high broadband access	DL: 1 Gbps UL: 500 Mbps	10 ms	Pedestrian
Broadband access in a crowd	DL: 25 Mbps UL: 50 Mbps	10 ms	Pedestrian
50+ Mbps everywhere	DL: 50 Mbps UL: 25 Mbps	10 ms	0-120 km/h
Ultra-low cost broadband access for low ARPU areas	DL: 10 Mbps UL: 10 Mbps	50 ms	on demand, 0-50 km/h
Mobile broadband in vehicles (cars, trains)	DL: 50 Mbps UL: 25 Mbps	10 ms	On demand, up to 500 km/h
Airplanes connectivity	DL: 15 Mbps per user UL: 7.5 Mbps per user	10 ms	Up to 1000 km/h
Massive low-cost/long-range/low-power MTC	Low (typically 1-100 kbps)	Seconds to hours	on demand, 0-500 km/h
Broadband MTC	See the requirements for the Broadband access in dense areas and 50+Mbps everywhere categories		
Ultra-low latency	DL: 50 Mbps UL: 25 Mbps	<1 ms	Pedestrian
Resilience and traffic surge	DL: 0.1-1 Mbps UL: 0.1-1 Mbps	Regular communication: not critical	0-120 km/h
Ultra-high reliability & Ultra-low latency	DL: From 50 kbps to 10 Mbps UL: From a few bps to 10 Mbps	1 ms	on demand, 0-500 km/h
Ultra-high availability & reliability	DL: 10 Mbps UL: 10 Mbps	10 ms	On demand, 0-500 km/h
Broadcast like services	DL: Up to 200 Mbps UL: Modest (e.g. 500 kbps)	<100 ms	on demand, 0-500 km/h

Figure 2. Summary of 5G parameters

Other considered requirements are that spectral efficiency should be increased significantly compared to the current 4G networks, coverage should be increased and also signal efficiency should be greatly enhanced.

5G is in research stage, but 5G are expected to be operational around Q4 of 2020. According to NGMN [4] the current timeline for 5G is represented in Fig. 3.



Figure 3. 5G Roadmap



Meanwhile several important vendors have begun research and development for 5G in 2013 and in 2015 several 5G laboratory trials have begun [5].

**B. 5G Architecture**

5G will be a fully converged system that will support a multitude of applications ranging from data, voice and multimedia to critical communications, Internet of Things, low latency applications (for example driverless cars) and can function on moving platforms due to increased mobility [1]. A variety of 5G applications are represented in Fig. 4.

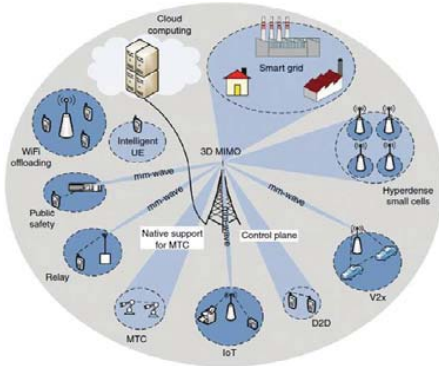


Figure 4. Generic 5G Architecture

The Network architecture of 5G can provide the following capabilities:

- Integrates the Radio Access Network (RAN) in various frequency bands. The radio frequency range will vary from 6GHz up to 100GHz. According to 5G requirements the RAN will provide virtually zero latency.
- Flexible deployments can be implemented using wireless and relying on optical technologies;
- HetNet Implementation;
- Cloud Computing can be applied to the RAN. This capability is combined with the transformation to cloud-based radio access;
- Virtualization of network functions will optimize network resources which improves scalability. This will be done in communication with data centers and will enhance Software Defined Networking (SDN) capabilities;
- Full usage of SDN capabilities;
- Networks will become self-aware, cognitive, and implement extensive automation and continuous and predictive learning [5];
- Internet of Things (IoT) integration;
- Usage of mmWaves;
- Massive MIMO (Multiple Input Multiple Output).

**C. Heterogeneous networks (HetNets)**

A heterogeneous network is a network in which multiple radio access technologies are used (e.g. GSM, WCDMA, and LTE) along with base stations that vary in size. A heterogeneous network is an efficient way of expanding mobile network capacity.

A heterogeneous network is made of two components: small cells (provides mobility) and macro cells (increase coverage and capacity). A HetNet is an evolution of a mobile access network in which an operator can add macro cell capacity as demanded. HetNet can extend closer to the end-user by positioning low cost and low power access nodes indoors or outdoors. (E.g. roadside, posts, corporate buildings) To facilitate deployments, 3G, LTE, 5G and Wi-Fi interfaces can be embedded within cells.

The HetNet access nodes are as follows:

- Macro/Micro Cells - Macro and micro cells provide universal coverage due to the fact that they have an inter-site distance of more than 500 meters;
- Small Cells – Small Cells are better suited for cloud applications due to higher speed demand. Small cells include:
  - Picocells – Picocells must be placed at about 200 meters or less;
  - Femtocells – The Coverage range for a Femtocell is about 100 meters;
  - Distributed Antenna System – A network of spatially-placed antennas connected to a common source via wireless;
  - Relay Nodes – Base stations that provide coverage/capacity to macro cells. Relay Nodes are connected via a Donor eNodeB (through a radio interface).

A generic HetNet architecture [6] that makes use of these access nodes is represented in Fig. 5:

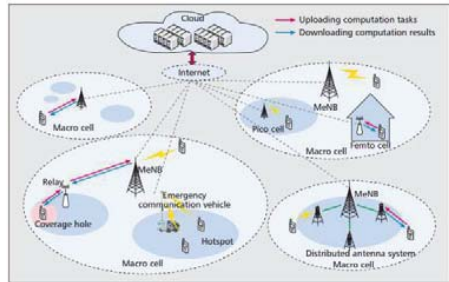


Figure 5. Generic HetNet architecture

**D. 5G Application – Software Defined Network**

The Software Defined Networking approach is composed of a logically centralized entity called the Controller which manages the associated network data plane using an Application Programming Interface (API) that allows

configuration of parameters such as forwarding tables of network equipment. (E.g. router, switch) [7]. A comparison between traditional network architecture and the SDN approach is presented in Fig. 6.

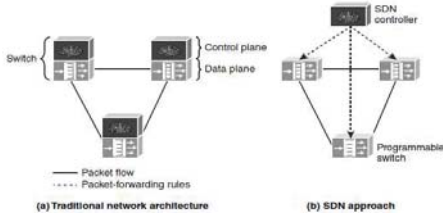


Figure 6. Comparison between traditional networks and SDN Approach

As presented in Fig. 7, the 5G technology can benefit from the programmability and scalability of SDN and NFV (Network Function Virtualization) technologies.

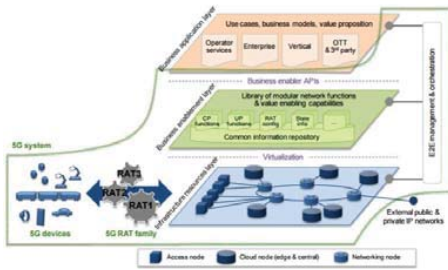


Figure 7. 5G and SDN convergence

As such, the 5G architecture is a native SDN/NFV architecture covering aspects ranging from devices, (mobile/ fixed) infrastructure, network functions, value enabling capabilities and all the management functions to orchestrate the 5G system [4]. APIs are provided on the relevant reference points to support multiple use cases, value creation and business models.

#### IV. MEASUREMENT RESULTS

In this section we present the measurement results from the implementation of a decentralized Cloud platform using 5G networks.

##### A. Proposed measurements

The purpose is to create a 5G Network using the USRP platform and implementing a cloud platform. The 5G Network will be implementing using a BPSK modulation for header and QPSK modulation for data transmission. In order to encode data on multiple carrier frequencies, OFDM (Orthogonal frequency-division multiplexing) will be used. The performance of the Cloud platform will be evaluated by measurements done for the transmission of a UDP data stream.

##### B. Proposed measurement environment

In this section we present the hardware and software connections between the USRP and the PC, which can be a physical machine or a VM in the cloud, as presented in Fig. 8. We ran it on Windows 7.

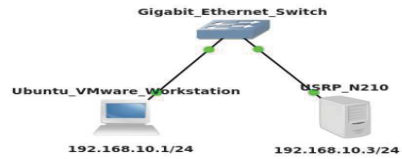


Figure 8. Proposed measurement environment

##### 1) USRP

Universal Software Radio Peripheral (USRP) N210 Networked Series [8] is a hardware designed by Ettus Research and used by research labs and universities to implement software defined radio systems. The USRP connects to a host computer having a software (i.e. GNU Radio) that controls the USRP hardware in order to receive and transmit data.

##### 2) VERT900 Antenna

The VERT900 Antenna is an omnidirectional antenna functioning in 824-960 MHz, 1710-1990MHz Quad-band Cellular/PCS and ISM bands working at 3dBi gain.

##### 3) GNU Radio

GNU Radio is an open-source software that contains a series of tools for implementing software defined radios. It contains blocks for signal processing as well as virtual sources and virtual equipments in order to emulate real equipment. GNU Radio can be used with external hardware (i.e. USRP) in order to create a software-defined radio.

The measurement environment was used in the SaRaT-IWSN project [9], and its main objective is to implement a radio transceiver that is capable of handling multiple communication requirements in a versatile manner, there is a need for a very flexible platform, in which the implementation of physical layer protocols is object-oriented, flexible and easy to modify. The OFDM transmitter and receiver are presented in the Fig. 9, respectively Fig. 10

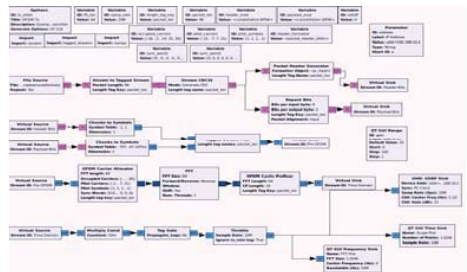


Figure 9. OFDM Transmitter in GNU Radio

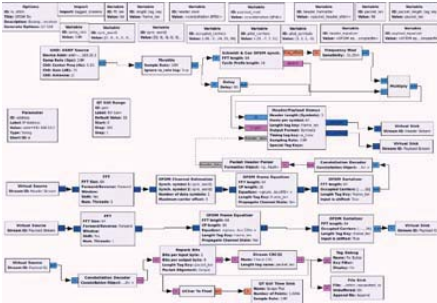


Figure 10. OFDM Receiver in GNU Radio

The results for a 5 GHz frequency (central frequency is 5.1 GHz) and a 25 dB gain at reception and emission are as represented in the Fig. 11, respectively Fig. 12.

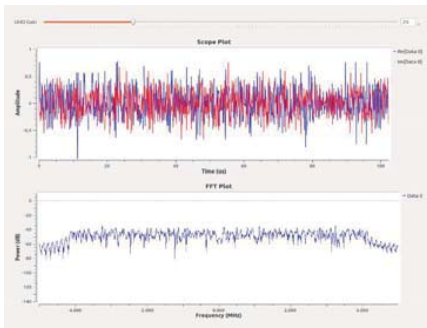


Figure 11. Amplitude/Phase representation at emission

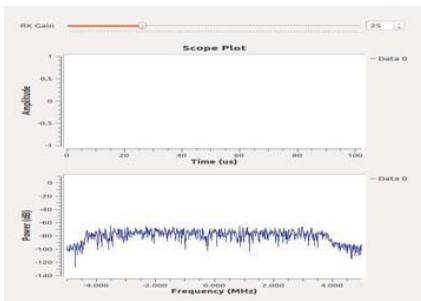


Figure 12. Amplitude/Phase representation at reception

The results for sending a text file using UDP protocol for different transmission and reception gain values are represented in Table 1.

TABLE I. RESULTS BASED ON DIFFERENT GAIN VALUES

Frequency [GHz]	TX Gain [dB]	RX Gain [dB]	Error Rate [%]
5	25	25	8
5	25	20	11
5	25	15	22
5	25	10	28

The error rate was calculated for 100 UDP packets and the results show an high error rate which needs further error detection and correction algorithms.

V. CONCLUSIONS

In this paper we presented an approach of implementing a decentralized Cloud platform using 5G Networks. Also, we presented measurement results, making use of 5G’s capabilities that enabled a resilient and reliable cloud platform. Furthermore, 5G gives the mobility to have a cloud everywhere it is considered to be necessary. As future work we envision possibilities of implementation and development of the proposed system for neutrino detection in the SARAT project, but there are endless other use cases including Internet of Things, M2M and countless Cloud applications.

ACKNOWLEDGMENT

The work has been supported in part by UEFISCDI Romania under grants no. 20/2012 “Scalable Radio Transceiver for Instrumental Wireless Sensor Networks - SaRaT-IWSN”.

REFERENCES

- [1] J. Rodriguez, Fundamentals of 5G Mobile Networks, Wiley, 2015.
- [2] P. Mell and T. Grance, „The NIST definition of cloud computing. NIST Special Publication 800-145, 2011” <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
- [3] K. Chandrasekaran, Essentials of Cloud Computing, CRC Press, 2014.
- [4] NGMN, 5G White Paper, [https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN\\_5G\\_White\\_Paper\\_V1\\_0.pdf](https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf)
- [5] Nokia 5G Whitepaper: [http://networks.nokia.com/sites/default/files/document/nokia\\_5g\\_architecture\\_white\\_paper.pdf](http://networks.nokia.com/sites/default/files/document/nokia_5g_architecture_white_paper.pdf).
- [6] L. Lei, Z. Zhong, K. Zheng, J. Chen and H. Meng, „Challenges on Wireless Heterogeneous Networks for Mobile Cloud Computing,” IEEE Wireless Communications, pp. 1-8, 2013.
- [7] W. Stallings, Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud, Addison-Wesley Professional, 1st edition, 2015.
- [8] M. Lenin, J. Shankar, A. Venkateswaran, and N.R. Raajan NR, “MIMO-Based Efficient Data Transmission Using USRP”, In Proceedings of the International Conference on Soft Computing Systems, Springer, pp. 405-419, 2016.
- [9] G. Suciut, G.R. Tecu, O. Fratu, A. Ochian, and V. Suciut, “Tele-monitoring system for water environments”, In Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies, International Society for Optics and Photonics, pp. 92580X-92580X, 2015.

# Data Recovery For Workgroup Distributed File System (WDFS)

Ravi Upra, Mayoon Yaibuates, Rounsang Chaisricharoen

School of Information Technology

Mae Fah Luang University

Chiangrai, Thailand

ravi@technocom.co.th,

mayoon@outlook.com, rounsang.cha@mfu.ac.th

**Abstract**— Data storages are not reliable. They incline to fail and in operate without any warning or sign. The most critical part is to get the data back to its original as soon as possible. There are many methods in recovering, mostly focusing on personal use with direct connection to the host and restoring from one single source of backup. Network recovery tends to be slow when comparing to direct connection. They depend on many factors such as network bandwidth, node's availability, node's processing power, etc. Most of the research does not concern much on how to speed up the process when restoring through the network. Making use of multiple backup sources is another key factor in achieving faster recovering data through the network. This paper proposes the effective method for selecting and restoring data from multiple backup nodes or sources through local network. Backup data was previously transferred which was part of Workgroup Distributed File System (WDFS). This paper focused on how to achieve the best performance in term of restoring data or recovering time.

**Keywords**— Data Recovery; Network; Node; WDFS;

## I. INTRODUCTION

When storage drives fail, it requires a certain period of time for recovering process. Recovery through the network even requires more extra time due to many factors such as bandwidth, availability of processing power, memory and even spinning speed of storage devices. Multiple data sources were introduced from the previous function of Workgroup Distributed File System (WDFS) [1].

WDFS was introduced on the concept of pooling resources like storage, which were scattered and unused within the devices or machines of a local network or domain. WDFS creates multiple sets of data which was generated from one single node or source and redistributes to multiple nodes within a local domain, for backup purposes. WDFS could keep track of data distributed to other machines and also retrieve data back from those machines or nodes. When there are changes in WDFS drives or folders, files are encrypted [2], compressed [3-4] and distributed. 3 replicas [5] or (n) sets were distributed to be kept on other machines. All participating nodes are independent of each other and the process is transparent to users. Base on the concept of WDFS, most of the work has been done, from selecting multiple reliable devices or nodes

that participates in the WDFS, distributing files or data to be kept on selected reliable nodes.

In this paper, we are focusing on retrieving data back to its original node in time of crisis. Recovering consists of 2 parts, Partial Recovery, which only retrieve certain required files or data and Full Recovery, which is used to restore everything back to its original state. Full recovery is used in case of storage drives fail. This paper will focus on full recovery and more on how to speed up recovering base on Best Node Selection and also file transfer protocol.

## II. BACKGROUND

Transferring data over the network is a challenging process. In this paper, data needs to be transferred or distribute in the network and TCP and UDP [6] are the most commonly use protocols for transferring packets. However, they're a bit different between the two. TCP guarantee recipient of the data transfer by having a tracking mechanism, making sure that all the packets arrive at the destination without any problem. In the case of data lost or corrupted, it will resend the packet to recipient again. UDP on the other hand, focus more on transfer speed by eliminating the tracking mechanism, so that packet could arrive faster at recipient side. No feedback mechanism. If the recipient didn't receive the packet or error, the whole data needs to be resend again. Here we did research few of the protocol that resides on both TCP/UDP. Different Network protocol exists such as Secure Copy Protocol (SCP) which support file transfer between the host and remote using Secure Shell (SSH). Multiple Path Secure Copy (MPSCP) [7] is the enhanced version of SSH which supports parallel file transfer which could increase speed up from 4% to 90%. Transferring large files over the internet is often performed using Peer-to-Peer (P2P) [8] technology. Bit Torrent [9] is the most popular P2P and will be in depth study to see that it could support for local network transfer. FTP [10] or file transfer protocol is also commonly used for transferring data over the network. It does have support for transport layer security which is FTPS or FTP-SSL. File Service Protocol (FSP) is a UDP-based protocol and a replacement for the File Transfer Protocol. FSP performance is much faster when comparing to FTP but as UDP-based are not reliable when comparing to TCP-based protocol. Multicast File Protocol (MFTP) [11] which is also UDP-based protocol. Scalable reliable multicast (SRM) , and

also enhanced version of MFTP called multicast file transfer protocol with erasure correction (MFTP/EC) .More study needs to be done on streaming [12] technology and could this method helps in restoring data at the same time.

Today’s biggest challenges as data are getting bigger and need to properly manage [13].Big Data is the hot issue in today’s problems. How to manage Big Data, and also big size data. How to transfer big size data over the network. Which types of protocols support in transferring big size data over the network.

III. NOVAL PROPOSED METHOD

WDFS offer 2 types of recovery as shown in figure 1.

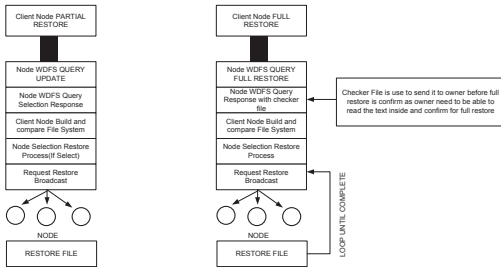


Figure 1. WDFS -Partial Recovery Vs Full Recovery.

- **Partial Recovery:** Is design to restore only required Files or folder. WDFS have a build in File Allocation Table and can send a command to initiate a request for Partial Restore. Nodes which hold data can send back the requested data.
- **Full Recovery:** Is design to fully recover information or data of WDFS drive, back to its original. Full Recovery is used when hardware fail, lost or a newly install storage that needs to recover all of the data back.

This paper focused on the Full Recovery process. Assume that a newly hard drive has been installed with a new operating system and need to restore all the data back. The initiator or requester node will ask for WDFS File Allocation Table (FAT) from other active nodes as they do not have one. After receiving the file, Requester Node will use the correct key to open up the file. Key is the password that was previously setup by initiator and used to protect the information from other person. Only the right owner will be able to open the file. If the correct password is given, requester node will be able to open up the File Allocation table (FAT) and request for Full or Partial Recovery. The node will start request files one by one from related multiple backup. Process for recovering will be split into 2 parts.

1. Finding Best Node Selection for restoring file.
2. Best practice method for speeding file transfer

Finding Best Node Selection for Restoring File

WDFS keeps minimum 3 sets of the same file or 3 replicas. Files are then distributed to be kept on 3 different client nodes. Sets can be of 3 or (n) sets, depends on the calculation of free spaces. For finding Best Node Selection, important factors or attributes will be collected and recalculate to get the best node in term of speed recovery. Process are as show in figure 2.

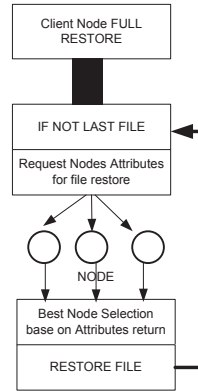


Figure 2. Best Node Selection process.

After receiving a request for restore, agent which resides on node side, pass back calculated attributes. With the help of weighted mean calculation, we put more weight on factors such as available bandwidth, which make it more likely to be selected if it does have more bandwidth.

Attributes Definitions:

Processor Power (P): is the processing power available of the node.

Let P be the average free processing power remain of the Node

Pt be the free processing power remain of the day at that specific time

W1 be the weight of attribute P

$$P = P_t * W_1 \tag{1}$$

Memory (M): is the memory available of the Node.

Let  $M$  be the average free memory available of the Node

$Mt$  be the average free memory available of the day at that specific time

$W2$  be the weight of attribute  $M$

$$M = Mt * W2 \quad (2)$$

Bandwidth (B): is the available bandwidth of the Node.

Let  $B$  be the average available bandwidth of the Node

$Bt$  be the average available bandwidth of the day at that specific time

$W3$  be the weight of attribute  $B$

$$B = Bt * W3 \quad (3)$$

Score (Sc): is the value that indicates the average score of particular node at that given time .Score is denoted by  $Sc$  and is derived from

$$Sc = (P+M+B) / \sum wi = (1+2+3) / \sum wi$$

$$P = Pt * W1$$

$$M = Mt * W2$$

$$B = B * W3$$

$$Sc = ((Pt * W1) + (Mt * W2) + (Bt * W3)) / \sum wi$$

$\sum wi$  = Total weighted mean which is 1 in which we focus more on Bandwidth

$$W1 = .2; W2 = .2; W3 = .6$$

Best node is selected depends on 2 factors. The responding time of the node and also reliability factors of the node. Bandwidth is one common attribute in reliability factors.

Let  $Rt$  be the Response time of a query from backup node to initiator node.

Let  $Sc$  be the score value or reliability history of that particular node which is collected by. The value of  $Sc$  derived from the previous paper.

$Bs$  be best node selection

$$Bs = ((1000-Rt) * W1) + (Sc * W2) / \sum wi$$

$\sum wi$  = Total weighted mean which is 1 in which we focus more on Response time from node

$$W1 = .6; W2 = .4$$

Figure 3. Shows best selected node after calculations.

Computer Name	Status	Response Time (Rt)	Score Value (%)
Ravi-PC	1	1ms	78
MFU-Comp1	1	3ms	56
Thawatchai	1	6ms	72

Figure 3. Node sorted after calculation.

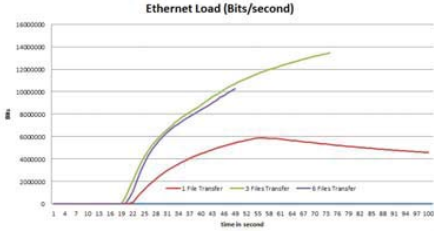
Once the node has been selected, the process for recovery will be started as a thread. The main process will start again for the second file and carry on until the last file been recovered. If the same node been selected over and over again, Response time (Rt) will be more and Score value (Sc) will be less. Score value (Sc) is calculated based on Bandwidth. Too much load on the same node will decrease available bandwidth, which will make that node unlikely to be selected again. The process will finally promote other nodes to be selected node. By this way, all nodes will be equally participate in recovering process and also helps speeding recovery process over a local network.

Best practice method for better transfer speed over network

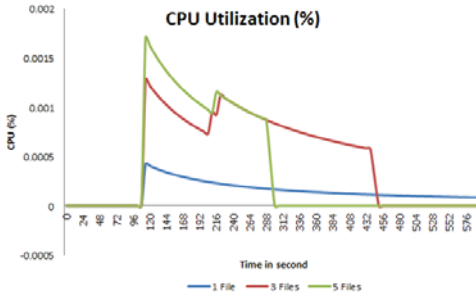
In this section, all related protocol and techniques such as streaming or parallel file transfer protocol will be studied. As with WDFS, Files are distributed over local area network, the concept is how to restore file at a faster speed. Which protocol to be used FTP, SCP, and MPSCP. All of the related work in this section will be in the future work.

### III.A SYSTEM ANALYSIS & DISCUSSION

This section compares the purpose model to the existing solution that is available. Existing models are mostly base on restoring from one single source and are mostly done through attaching storage devices directly to the host. In the proposed model, restoring is done through local network with multiple sources. Restriction for transferring files through a network is network bandwidth of that node and also the availability of that particular node. Graph 1, display a simulation from OPNET network simulation over 100 Mb network. The more files requested from the same node the less bandwidth will be available. Selecting the right node could help in speeding up the recovery process. From Graph 2, the more the node involves in sending or receiving files, the more processing power will be utilized. Recovering Process will be able to switch to the other nodes so that not too much load will be place on the same node. Another consideration is the right protocol for transferring data could also speed up transferring data over network.



Graph 1. Display Ethernet load (Bits/Second)



Graph 2. Display CPU Utilization when sending file over network

IV. CONCLUSIONS

WDFS proves to be a successful method in helping the Small business area in backing up of important information and cutting down the cost or need for backup devices. WDFS pool available and unused resources that are scattered within the domain and create a virtual private backup space. This paper proposes a model to support WDFS in recovering of data. The proposed model could speed up the process of restoring data over the network by restoring from multiple sources of backup. The process also detect, not too much load or too much work will be place on the same node by using Best Node Selection. All Nodes will be equally participating in the recovering process. Transferring Protocol needed to be study in depth and apply for the best transferring protocol over network. This will help in speeding the recovery process. The applying of best transfer protocol will be in the future work.

- [1] R. Upra and R. ChaisriCharoen, "Workgroup Distribution File System (WDFS) for personal cloud system," in *Information and Communication Technology, Electronic and Electrical Engineering (JICTEE), 2014 4th Joint International Conference on*, Chiangrai, Thailand, 2014.
- [2] Q. Q. R. Z. Guang-liang Guo, "Different Implementations of AES Cryptographic Algorithm," in *High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on CyberSpace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on*, New York, NY, 2015.
- [3] H. D. E. v. E. M. Bölümü, "Lossy Medical Image Compression Using Huffman Coding and Singular Value Decomposition," in *Signal Processing and Communications Applications Conference (SIU), 2013 21st, Haspolat*, 2013.
- [4] A. J. P. a. A. J. R. Neves, "LOSSY-TO-LOSSLESS COMPRESSION OF IMAGES BASED ON BINARY TREE DECOMPOSITION," in *IEEE*, Aveiro, Portugal, 2006.
- [5] L. WenHao , Y. Yun and Y. Dong, *Reliability Assurance of Big Data in the Cloung*, Waltham, MA 02451, USA: Elsevier Inc., 2015.
- [6] Shanel Narayan and Yhi Shi, "TCP/UDP network performance analysis of windows operating systems with IPv4 and IPv6," in *2nd International Conference on Signal Processing Systems (ICSPS)*, 2010.
- [7] L. R. H. U. Brian J. Guilfoos, "Improvements to Multiple Path Secure Copy," in *DoD HPCMP Users Group Conference, 2008. DOD HPCMP UGC*, 2008.
- [8] L. Zhong, W. Xiaofan and M. Kihl, "Topological model and analysis of the P2P BitTorrent protocol," in *Intelligent Control and Automation (WCICA), 2011 9th World Congress on*, Taipei,Taiwan, 2011.
- [9] L. Gilhyon, K. Hidehiro and N. Hidenori, "An efficient file transfer using helper in bittorrent," in *Communications (APCC), 2015 21st Asia-Pacific Conference on*, 2015.
- [10] W. Pan, H. Zhiqou and S. Minyong, "Performance analysis of HTTP and FTP based on OPNET," in *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on*, Okayama,Japan, 2016.
- [11] H. C and H. M, "Performance comparison of reliable multicast protocols using the network simulator ns-2," in *Local Computer Networks, 1998. LCN '98. Proceedings., 23rd Annual Conference on*, NJ,USA, 1988.
- [12] E. Bob, "Stream and File Formats—Where are we Now?," *SMPTE Journal* , vol. 111, no. 6-7, pp. 319-322, July-Aug. 2002.
- [13] Zhenju Li, Xuenjun Li, Tao Li and Jianwei Xie, "Towards building a distributed file system for remote sensing image process," in *Big Data Analysis (ICBDA), 2016 IEEE International Conference on*, 2016.

# The Influence of Network Neutrality on CONASENSE Innovation

Yapeng Wang, Ramjee Prasad<sup>1</sup>

**Abstract:** In 2015, the United States and European Commission enacted respective rules in relation to open internet, the network neutrality (NN) gained more attention and was discussed ardently. This paper reviews the NN debate process, and the opinions from different sides, such as the network providers, the service providers, other relevant companies, governments and researchers. This paper also introduces a concept that is CONASENSE (Communication, Navigation, Sensing and Services ) which aims to formulate a vision on solving societal problems with new telecom technique to improve human welfare benefit. This paper focuses on the influence of NN on CONASENSE innovation and summarize the current situation of NN in service innovation era.

**Key words:** Network Neutrality, CONASENSE, Innovation, Long term benefits

## 1. INTRODUCTION

The Information and Communication Technology (ICT) is playing an important role in our life. It supplied us with many various innovative services and applications, from ecommerce, e-health to a real-time telephone meeting and a live video streaming, improving human's Quality of Life (QoL) and benefiting the whole society. We have entered a service innovation era, in which every part of the industry chain is making contribution to the ICT industry. Internet Services Providers (ISPs) are continuing to upgrade the network, the Internet Content Providers (ICPs) are supplying full of various content and services to keep the telecom industry prosperous. How these innovative service run over the telecommunication network is governed by not only technology, but also by the rules as proved by government or authorities in some region. NN, as an important regulation which influences future Internet development, aims that every end user has the equal right to access the internet and use the legal internet content and applications. CONASENSE, as will be discussed later in this chapter, is a telecom convergence concept that will be run above ICT platform. This chapter focuses on the CONASENSE service, and analyzes NN rules' impact on this service.

### (i) Network Neutrality

Network Neutrality rules aims to provide an open internet [1] to the end users. The FCC defined open internet refers to "*uninhibited access to legal online content*

---

<sup>1</sup> Center for TeleInfrastruktur (CTIF), Aalborg University, Denmark



*without broadband Internet access providers being allowed to block, impair, or establish fast/slow lanes to lawful content” [2].*

This means that a legal content, whether it is an application or data, must reach users without the intermediate communication system controlling its flow. Internet Content Providers (ICPs) and Internet Service Providers (ISPs) cannot block, throttle or create the special facilities for a content or application. This gives a kind of liberty to the end users to enjoy the variety of information without bothering about how ICT is dealing with that information. Therefore, users may demand a better network services to enjoy the lawful contents. Section 2 introduces NN in detail.

(ii) CONASENSE [3]

CONASENSE refers to Communication, Navigation, Sensing and Services. In November 2012, CONASENSE foundation was established to support its development, as an integration of communications, navigation and sensing technology. It helps define and steer processes directed towards actions on investigations, developments and demonstrations of service innovation, especially for those services that have high potential and importance for society. Section 3 gives a detailed introduction to this concept.

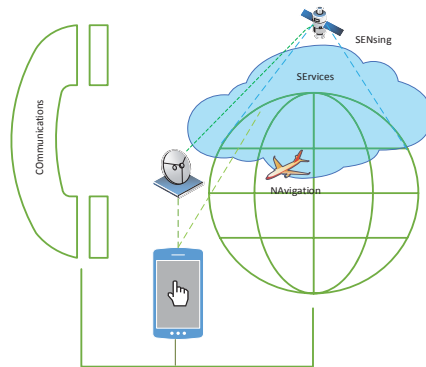


Fig 1: CONASENSE Framework [3]

(iii) Innovative Services

In this chapter, the innovation service refers to over the top service (OTT). OTT implies communications carried over the physical network infrastructure using an IP protocol to reach services available on the internet [4]. An OTT application is any app or service that provides a product over the Internet and bypasses traditional

distribution. Services that come over the top are most typically related to media and communication and are generally cheaper than the traditional method of delivery. As we discussed CONASENSE service belongs to the OTT service domain. Section 4 will discuss NN's impact on it.

This chapter points out that both the opponent and proponent of NN debate agree on the need to keep the Internet open to innovation, and preserve the freedom of users to access the content and services. Meanwhile this chapter also analyses how NN impact innovation in CONASENSE, and meeting the regulation goal. This chapter makes analysis mainly from technical perspective, additionally social and economic aspects are also discussed.

Apart from Section 1, the rest of the chapter is organized as follows: Section 2 reviews the evolution and current discussions of NN, Section 3 briefly introduces CONASENSE, Section 4 qualitatively analyses the impact of NN on Services and finally Section 5 concludes the chapter.

## 2. THE NETWORK NEUTRALITY

As more and more regulators tend to believe that the ICPs are the motivation of innovation, economy and investment. And with the rapid development of the Internet as an ubiquitously available platform and resource, the network infrastructure owners are regarded as to *“have both the incentive and the ability to act as gatekeepers standing between edge providers and consumers. As gatekeepers, they can block access; target competitors, extract unfair tolls”* [2]. But the open internet is regarded as to the guarantee to the innovation, economy and investment, more and more countries government enacted the NN rules. Presently, more than 10 countries enacted relevant rules [5]. Figure 2 shows the countries that either have enacted NN or are in their discussion phase.



Fig 2 : Countries that either have enacted or are in the process of enacting NN

### (i) The Concept of NN

The concept of NN commonly indicates that Internet services providers make or keep

the Internet open, and ensure all the users have same right to access to the network and use the content or services without any discrimination. [2]

### **(ii) NN's principles**

Currently, Network Neutrality is a global debate [5]. FCC released its updated Open Internet Order in 2015, to enact strong, sustainable rules to protect the open internet. The order includes 3 bright-line rules as follows [2]:

- No Blocking (NB), to prohibit the network providers to block the legal applications, contents and devices.
- No Throttling (NT), to prohibit the network providers to degrade the traffic of legal applications and contents.
- No Paid Prioritization (NPP), to prohibit the network providers to provide and charge the differentiated service for the applications and contents.

*“No unreasonable interference or disadvantage to consumers or edge providers [2], and, enhanced transparency between users and ISPs”. “As with the 2010 rules, this Order contains an exception for reasonable network management, which applies to all but the paid prioritization rule” [2] is also emphasized in FCC rules.*

In October 2015, the European Parliament also approved the first EU-Wide NN rules that enshrined its principle into EU law [6]. It indicates that: *“No blocking or throttling of online content, applications and services. Accordingly, every European must be able to have an access to the internet and, all contents and services, via a high-quality service that is provided by an ISP such that all traffic must be treated equally. NN rules seems more inclined to the end users and, equal treatment allows reasonable day-to-day traffic management according to justified technical requirements which must be independent of, the origin or destination of the traffic, and, of any commercial considerations.”*

### **(iii) History of NN**

The phrase of Network Neutrality was first proposed in a law review article [7] by Tim WU in 2003, mainly referred to the concepts of freedom, competition and innovation. NN suggests that each network protocol layer should be independent and perform the assigned duties at the original phase. With the expansion of the commercial internet, the focus of the market competition has been shifting from the connection and network layer to the application and content layer. The main debate of NN also was switched from technological field to commercial field.

Through the key event happened in U.S.A are as follows, we can find the way of NN is not smooth. Meanwhile the term holds different meanings to different expertise, and NN has different debate focuses in different phases. The discussion of NN will be introduced later on.

- In December 2010, The FCC approved the Open Internet Order which was consisted of three items of NN regulations, and they are “*Transparency, No Blocking and No unreasonable discrimination*” [8].
- In January 2014, United States Court of Appeals for the District of Columbia Circuit (D.C. Circuit) overturned the Open Internet Order.
- In February 2015, The FCC issued Open Internet Rules and Order, and in June 2015, The Open Internet Rules and Order came into effect officially.
- In August 2015, The D.C Circuit announced the Open Internet Rules and Order will face an important federal-court test

**(iv) Current discussion on NN**

From the history of Network Neutrality, we can find that NN has been really controversial. And in 2015, European Commission and FCC enacted the NN rules. Most of the goal and the principle in both NN rules are very concurrent, but the distinction still existed. The essential debate between Europe and U.S.A will be discussed below:

**(a) Service innovation:**

The proponents of NN are mainly those enterprises that are related to internet contents. Worriedly, they are stating that actions departing from NN principles could threaten the innovation of the internet content as, ISPs may increase control on the content and applications over the internet. In order to encourage the innovative services with enhance quality of service especially from startups, the new EU net neutrality rules state the following “*enable the provision of specialized or innovative services on condition that they do not harm the open internet access. These services use the internet protocol and the same access network but require a significant improvement in quality or the possibility to guarantee some technical requirements to their end-users that cannot be ensured in the best effort open internet* [9]. *These specialized or innovative services have to be optimized for specific content, applications or services, and the optimization must be objectively necessary to meet service requirements for specific levels of quality that are not assured by the internet access services*”. The rules also urges these services cannot be a substitute to internet access service, can only be provided if there is sufficient network capacity and must not be to the detriment of the availability or general quality of internet access service for end-users.

FCC’s rules refer above mentioned service as non-Broadband Internet Access Service (*non-BIAS*) [1] “*Non-BIAS data services, which are not subject to the rules. According to the rules, non-BIAS data services are not used to reach large parts of the Internet, not a generic platform—but rather a specific “application level” service, and use some form of network management to isolate the capacity used by these services from that used by broadband Internet access services.*”

**(b) The investment on the network infrastructure:**

The opponent of NN which is typically network operators argue that NN regulation will make it more difficult for ISPs and other network operators to recoup their investments in broadband networks weaken the incentives to invest and upgrade the telecom infrastructure. Some ISPs have argued that they will have no incentive to make large investments to develop advanced fibre-optic networks if they are prohibited from charging higher preferred access fees to companies that wish to take advantage of the expanded capabilities of such networks [10]. FCC reclassified the BIAS as telecommunication service in the Open internet Order [2], FCC believed that the reclassification will preserve investment incentives.

**(c) The management of Internet traffic by Internet Service Providers and what constitutes reasonable traffic management.**

*Commonly, “traffic management is used to effectively protect the security and integrity of networks. It helps deal with temporary or exceptional congestion or to give effect to a legislative provision or court order. It is also essential for the certain time-sensitive service such as voice communications or video conferencing that may require prioritization of traffic for better quality. But there is a fragile balance between ensuring the openness of the Internet and the reasonable and responsible use of traffic management by ISPs [11].* The opponents of NN argue that NN may prove ineffective in such a dynamic framework nowadays, leading to welfare-loss caused by congestion problems, arguing in favor of the possibility of differentiation of data packets according to their quality sensitivity [12].

EU urged that all traffic be treated equally but allow the network providers to make reasonable traffic management in consideration of justified technical requirements, so as to preserve the security and integrity of the network or to minimize temporary or exceptional network congestion. According to FCC rules, the no-blocking rule, the no-throttling rule, and the no-unreasonable interference/disadvantage standard will be subject to reasonable network management for both fixed and mobile providers of broadband Internet access service. Figure 3 shows a comparison between EU and the United States rules. We can see that regulators paid attention to the innovation when making their own NN policies. They leave space for the innovative services together with strict constraints.

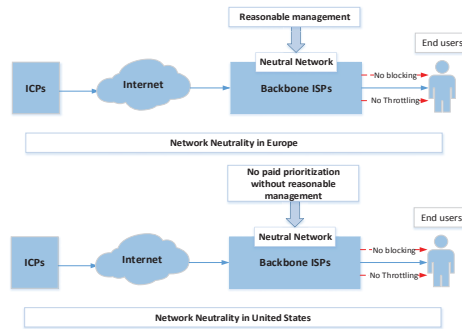


Fig 3: comparison of NN between the US and the EU [8]

### 3. CONASENSE

In the past 10 years, wireless and Internet technology have created an explosive growth in ICT services, supplying a wide range of personal and group data services. The limited data transmission capacity, as the bottleneck in the beginning, was broken little by little. The same wireless channel can now support much higher data rates and thus many more demanding services. There is also now a rapidly increasing demand and innovative application areas for services related to positioning, tracking and navigation. Meanwhile, sensing technology, sensors and sensor networks have experienced an unprecedented development in past several years. A variety of sensors types are now available on the market in many domains. And new sensor types are continuing coming out in different domains. As the integrated provision of these services will obviously raise people’s living efficiency and Quality of Life (QoL), and the ICT network can support much higher data transmission rate, it seems to be the right time to develop integrated CONASENSE services. One problem is that traditional approaches may not be optimal. Different services may have different frequency bands, waveforms and hence different receiver platforms.

Most promising CONASENSE services may be available in 5-20 years. The new CONASENSE services should reflect the trend towards an information society in which applications and services become important likewise, bearing in mind that computing and communications should be integrated so as to save energy, software defined radio combined with cognitive radio technology become increasingly important for new developments.

In order to achieve the goal of CONASENSE service, it is important for researchers and developers to identify the requirements for energy, terminal/platform and receiver/system design concerning diverse application areas, such as e-health, security/emergency services, traffic management and control, environmental monitoring and protection, and smart power grid. They should pay much attention to the novel CONASENSE architecture design so as to minimize the energy

consumption because of requirements for mobility, high data rate communications and green communications. The novel CONASENSE architecture design will help address problems of the existing architectures and be sufficiently flexible for future developments. Consequently, the design of the CONASENSE architecture will be carried out so as not only to integrate existing and novel communications, navigation and sensing services but also to provide smooth transition between existing and new systems in hardware and software. NN impact on CONASENSE service innovation will be discussed in the next section.

#### 4. Network Neutrality Impact on CONASENSE Innovation

Quality of Service (QoS) is an important aspect of CONASENSE. Therefore, this section focuses on QoS and discusses CONASENSE in general.

From technology aspects, no discrimination requirement of Network Neutrality may have some negative impact on the new CONASENSE service innovation. The network nowadays is not neutral (Differentiated Network), it can supply quality of service (QoS) to different applications according to their characters and requirements. But according to the new NN rules, QoS measures will be taken as discrimination, and be banned in the pure neutral network, as shown in Figure 4. It is obvious to lower the network efficiency and will cause congestion easily.

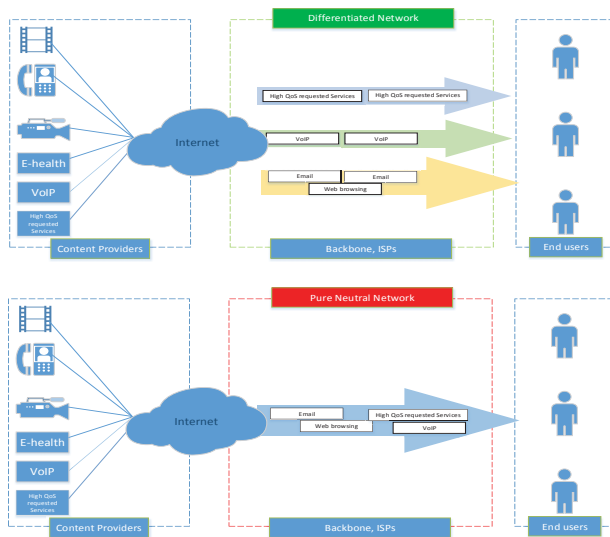


Fig 4: The difference between differentiated network and pure neutral network

In the short term, NN will certainly make Internet content companies have more innovation space, encourage more innovative applications and increase the efficiency

of the society. But in the long run, NN will inevitably weaken the enthusiasm of the investment on the network construction; lower the network quality level gradually. This in turn will influence Internet companies in the end. The reasons are as follows:

1) NN will impact the quality of the Internet service because NN limits the ability to guarantee the network QoS. As of now, there are mainly two types of model of QoS [13].

- Integrated Services (IntServ) uses RSVP (Resource Reservation Protocol) for signaling to invoke a pre-reservation network resource and traffic handling. IntServ can provide end to end guarantee for services and applications. But because Intserv is expensive and time-consuming, it has not been widely used in the Internet.
- Differentiated service (Diffserv) is a mechanism that identifies and classifies traffic in order to determine the appropriate traffic handling mechanism. It can integrate the same type of services and manage them together. It is now used widely.

To exemplify the models of QoS, we can imagine an accident site, there are several wounded persons: some of them got severely injured and need to go to hospital immediately, while others are not so urgent. When the ambulances come, the doctors do not do any diagnosis and ask the severely wounded to get on the ambulance which makes some non-significantly wounded cannot get on the ambulance in time. This case is Best-Effort Model (no QoS). The IntServ model is that the wounded whoever is serious or not would need to reserve the ambulance, if the severely wounded didn't make a reservation, he will not be sent to hospital. In the Diffserv model when the ambulances come, the doctors will distinguish the level of wounded and ask serious wounded persons with similar situation to get on and be treated. So we can find Diffserv is an effective method for application and service transmission on Internet. But with NN, differentiated service is banned in public network.

2) Different Internet services and applications have different requests. QoS uses four parameters to judge the services request and include: bandwidth, delay, delay variation and packet loss rate [14]. Table 1 presents the definition and the impact of the four parameters of QoS.

Table 1: Definition and Impact of the four QoS parameters [14]

Parameter	Definition	Impact on QoS
Bandwidth	The maximum data that can be transmitted per second in the network.	It is to measure the transmission capacity of the network
Delay	The transmission time a service packet takes from one nodes to the other one	If the delay is too long, it will lower the QoS



Delay variation	It means the variation of the packets delays in the same flows	It is a key factor impacting QoS
Packet loss rate	It means the rate of the data packets that lost in transmission	Low packet loss rate will not impact the QoS

Table 2: QoS Parameters for Some Applications [15]

Type	Bandwidth	Delay	Delay variation	Packet loss rate
Email	low	Not sensitive	Not sensitive	Not sensitive
WhatsApp	Medium	Medium sensitive	Medium sensitive	Not sensitive
Video application	High	Sensitive	Sensitive	Sensitive
E-commerce	Medium	Sensitive	Sensitive	Sensitive
IoT, Industry 4.0	High	Sensitive	Sensitive	Sensitive
CONASENSE services	Large	sensitive	sensitive	Not known yet

Table 2 shows that some services are sensitive to delay but not to packet loss, and some services are sensitive to packet loss but not to delay. It is necessary for a network to treat packets belonging to different applications differently in terms of their different requests. For instance, a network should give low-delay service to VoIP packets, but best-efforts service to e-mail packets [15]. Different services also need different QoS provisioning.

(B) NN may reduce the incentive of investment on the network construction. As we know, the funds usually flow to the market with big profits. If there is not enough profit from telecom network, investment on the network will certainly be reduced. Internet companies may also lose their interest to improve the efficiency of transmission, because they do not need to pay for the overuse of bandwidth. Insufficient investment on network with the overuse of bandwidth will certainly result in network congestion and inefficiency.

(C) The large amount of CONASENSE services, to some extent, can be taken as environment-sensitive services. And for people on air, on vessel, or on car, on road, their environmental situation including location, speed, temperature, health condition etc. are constantly changing. All the relevant information should be received correctly and timely for a CONASENSE services will probably use them to make a correct and

in-time decision, to help improve people's QoL. NN rules put relatively strict criteria of specialized services or no-BIAS services, sometimes on a case by case base. That may slow down the CONASENSE services experimenting, developing, commercializing process.

High data transmission capacity is the base for the CONASENSE services. Most of the CONASENSE services focus on the future, on the assumption of Wireless Innovation System for Dynamically Operating Mega Communication (WISDOM) [16] or 6G network deployment and extensive use of sensors. The value of CONASENSE services will not be realized on a congestive and a network without operational guarantees.

## 5. CONCLUSIONS

NN rules, its relevant debates, and, its impact on ISPs and ICPs are discussed in this chapter as a background study. In this discussion, NN rules such as NB, NT, and NPP were covered including their exemplified definitions and impacts on present communication paradigm. These rules are user centric and may hamper the financial benefit of ISPs. The prohibition on any control in the flow of information through a communication network may refrain ISPs from earning quality based revenues. In such case, ISPs may not be willing to enhance or upgrade the infrastructure to improve the QoS of the network. Therefore, although NN rules are beneficial for innovations currently, as campaigned by NN proponents, provided by NN rules, the end user may not "really" enjoy the QoS.

With the convergence of Internet and telecommunication network, basic telecommunication services have been moved to the public Internet network. Many of them have quite strict QoS requirement so as to guarantee the relevant service responsibilities, such as emergency call, the basic service quality level. In this process, the policy maker and regulator should be careful to handle between service innovation, lawful customer right and social responsibility. QoS models have enough reasons to remain as legal network functions.

We have investigated how NN rules, in their present form; will impact CONASENSE service in the future. Through some assumed futuristic scenarios it is discussed that NN rules may not be favorable for innovation service. Banning Diffserv may put innovation service in the category of a usual communication system and ignore the very high data demand of the CONASENSE service. As customers may not welcome the innovative but poorly served new technology, the aspirant companies will struggle capturing market for this innovative technology. Further, NPP may not allow users to choose better services among the choices offered by ISPs.

NN policy may stimulate the development of service innovation in short term but may

be not good for the network base in long term, which may in turn hinder the service innovation, such as CONASENSE services. The relatively strict criteria for specialized services or non-BIAS services may slow down the CONASENSE services experimenting, developing, commercializing process.

Regulators are suggested to make careful decisions on NN policy to guarantee the long term benefit according to their own situation. All the stakeholders should be encouraged to find a way to ensure the prosperous development of the industry in the market on their own.

## REFERENCES

- [1] FCC, Category “For Consumers ” “Open Internet”, 2015
- [2] FCC, “The Open Internet Rules and Order” FCC 15-24, March, 2015
- [3] Leo Lighthart, Ramjee Prasad, “Communication, Navigation, Sensing, and Services (CONASENSE)” River Publisher, 2014
- [4] Cory Janssen, “Over the Top Application (OTT),” Techopedia <http://www.techopedia.com/definition/29145/over-the-top-application-ott>, 2015
- [5] Winston Maxwell, Mark Parsons, Michele Farquhar, Net Neutrality – A Global Debate, Hogan Lovells Global Media and Communications Quarterly 2015, P15-17
- [6] European Parliamentary “Our commitment to Net Neutrality”, EU Actions, October, 2015
- [7] Tim Wu, Network Neutrality, Broadband Discrimination, Journal of Telecommunications and High Technology Law, Vol. 2, p. 141, 2003
- [8] FCC, “The Open Internet Order” FCC 10-201, December 21, 2010
- [9] European Commission,, “Net Neutrality challenges”, October, 27, 2015
- [10] Dong-Hee Shin & Tae-Yang Kim, “A Web of Stakeholders and Debates in the NetworkNeutrality Policy: A Case Study of Network Neutrality in Korea”
- [11]European Commission, “Roaming charges and open Internet: questions and answers” 27 October 2015

- [12] Barbara van Schewick, "Network Neutrality and Quality of Service: What a Nondiscrimination Rule Should Look Like?", *Stanford Law Review* [Vol. 67:1], January 2015
- [13] El-Bahlul Fgee, Jason D. Kenney, William J. Phillips, William Robertson1 and S. Sivakumar, "Comparison of QoS performance between IPv6 QoS management model and IntServ and DiffServ QoS models" the 3rd Annual Communication Networks and Services Research Conference, 0-7695-2333-1/05, 2005, IEEE
- [14] Fabricio Carvalho de Gouveia and Thomas Magedanz "Quality of Service in Telecommunication networks", *Telecommunication systems and technologies*, Vol, II
- [15] Hua wei technologies co(2013), ltd, "QoS Technology White Paper", <http://e.huawei.com/us/marketing-material/onLineView?MaterialID=%7B3623FE01-3572-4413-A71B-EBEBE9F2E141%7D>
- [16] Ramjee Prasad, *5G Revolution Through WISDOM*, Springer Science+Business Media New York 2015, *Wireless Pers Commun*(2015)81:1351-1357 2015 March

# Improving Consideration Performance of Student Loan Fund using Data Clustering Approach

Klangwaree Chaiwut, Worasak Rueangsirarak, Rongsan Chaisricharoen

School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand  
5771501003@lamduan.mfu.ac.th, worasak.rue@mfu.ac.th, rongsan.cha@mfu.ac.th

**Abstract**—The student loan fund is an important mechanism, that can fulfill the quality of education as the standard on an equal basis. It aims to help the needy students to study thoroughly. There are many factors that influence the decision making based on the received information. Therefore, the main factors need to be improved and verified in order to reduce the error and enhance the efficiency of the decision making processes. This paper aims to improve and verify all related factors of the student loan candidates' information. K-mean clustering was used to improve student loan fund consideration performance by validating difference parameters of Thailand national criteria. The result shows that the clustering performance of improved criteria, without bias factors, is better than the tradition criteria on the average distance, no single-member cluster, and cluster's density.

**Keywords**—student loan; data clustering; K-mean; criteria for consideration; bias factors

## I. INTRODUCTION

Education is a key role in determining how you spend your life. It enhances the quality of people's life. The higher education means higher earnings and better health. An individual's level of education is directly correlated to the quality of life the student will live in the future [1]. Many people cannot get the educational opportunities because they don't have enough money to study. Student loan fund (SLF) is a form of financial support for students who are from a poor family background, designed to help students pay for their tuition fees and associated expense [2]–[4]. The student loan is difference from common loan in that the interest rate considerably lower and the repayment schedule after the student graduated. The several countries select the student loan fund for supporting a citizenship to expand education opportunities and fulfill the quality of education as the standards based on an equal basis such as China, United State, Canada, Hungary, United Kingdom and Thailand [5]. In Thailand the student loan is an important mechanism of the Thai government to develop their country. It provides a financial assistance for students who are from the secondary education level. The level includes general and vocational education, and up to higher education. However, the student loan fund is limited in a number of budgets [6]. The most university in Asia used the interview method to screen the students who qualify for getting a loan [7].

Many studies show that there are not all poor students in the accepted list of student loan [5]. Therefore, the loan that is

to be considered should be complied with the standards that are accurate and fair for screening an insufficient students who are in low-income families, by providing a financial aid for those who need it most. Thus this research used the data clustering approach for improving student loan fund consideration performance by eliminating unimportant factors from the criteria.

This paper is organized as following sections. Section II explains the related work. Section III proposes the research methodology. Section IV presents the experiment result and analysis. Section V concludes the paper.

## II. RELATED WORKS

### A. Student Loan

Student loans are used quite extensively in Asia to provide financial support for students in higher education, to enable them to pay the costs of tuition and/or living expenses [7]. These are designed to help the students who are insufficient in financial. The student loans system does not need any loan guarantees or mortgages [1]. Several countries have the student loan such as New Zealand, Chile, South Africa, Ethiopia, Hungary, the UK and Thailand, which are mostly supported by the government [5]. In China, the student loan is very popular. It is designed to help the students from poor families [1], [3]. The government support for poverty stricken students in colleges and universities. A certain number and a limited budget allocation require resources and targeted toward needy students [8]. In the USA, the student loan is applied for a higher education development. It expands the opportunities in education for the United States citizen [7].

### B. Loan Decision Making Techniques

There are many researches study about the techniques to make a decision in loan, especially for banking.

Jozef Zurada used eight models; logistic regression (LR), neural network (NN), radial basis function neural network (RBFNN), support vector machine (SVM), case-based reasoning (CBR), and three of decision tree model (DTs) to test classification accuracy based on a historical dataset provided by a German financial institution. The dataset contain 1,000 samples and 21 attributes. The decision trees models has an overall classification accuracy rates better than those of the other models [9].

Zhenyu, Wei and Yayue used GABP algorithm of neural network to predict the future of national student loan for judging whether the student loan is to be broken or not. The result seems to be able to handle a lot of the sample data with fast operation and accurate. The dataset consists of 18 samples which selected from 480 national student loans. The result shows that it can reduce credit risk and predict the future of national student loan applicant's defaults more quickly and accurately [2].

Aslı Çaliş al. used *K*-mean clustering and decision tree for reducing the rate of risk in decision making. *K*-mean grouped data into the same cluster, after that used decision tree to form the potential credits of customer. The dataset contain 200 samples based on legal follow-up and normal payment records for a period of six months belonging to the individuals in the branch. The result can be estimated and reduced the rate of entrance into legal [10].

Venkata Sreedhar used Rattle for reducing the dimensionality of the dataset, then classified the loan applicants by using a classification based on decision tree method. The result can predict which loan applicants are "risky" and which are "safe" from a loan dataset [11].

Kambal, Osman, Taha, Mohammed and Mohammed used classification techniques, which are decision tree (DT) and artificial neural networks (ANN) for credit scoring. In addition, genetic algorithms (GA) and principal component analysis (PCA) are also applied as feature selection techniques. The dataset is Sudanese credit dataset contains 1300 cases, where 720 are classified as non-defaulters and 580 as defaulters. It consists of 17 attributes. The result reveals that ANN model outperform DT model in most cases [12].

Most of the loan decision making techniques used the data clustering to consider the results [2], [10], [12]. Therefore, this research will apply clustering technique to improve consideration performance of student loan fund.

### C. Data Clustering

Clustering analysis aims to group data on the basis of similarities and dissimilarities among the data elements [13]. Clustering is a partitioning algorithm that distinguishes a dataset into the clusters that contain a similar data according to defined criteria. It is an effective machine learning method and easy tool for classification, decision making and a selecting group of objects from dataset for experimental analysis [13]–[15]. The clustering is also a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics [16]–[18].

The *K*-mean clustering, a popular and simple algorithm, was proposed over 50 years ago. The *K*-mean can group each data point to a member of multiple clusters with a membership [15]. The *K*-mean algorithm is a famous algorithm which is available in Weka tool [19]. *K*-mean algorithm will partition the whole space into different segments and calculate the frequency of data point in each segment. This algorithm calculates the distance for each data point that belong to only one cluster, with increasing in the number of clusters would prove to be more significant [13].

The performance of four clustering techniques, were performed and used to detect the fault in traditional student loan found analysis, confirm that there is a fault in student loan consideration with tradition criteria provided by the government. The *K*-mean algorithm showed the highest percentage of an ambiguous cluster, which confirmed to be an appropriate algorithm for student loan fund analysis [20].

Therefore, the algorithm procedure of *K*-mean can be described as follows;

*Step 1:* Desired number of clusters, *k*. And desired points represent initial group centroids.

*Step 2:* Group the object to the cluster with the nearest centroid by using this equation;

$$Ex = \sum_{i=1}^k \sum_{p \in c_i} dist(p, c_i)^2 \quad (1)$$

where;

*E* = Sum squared error for all objects in the dataset.

*n* = Amount of data.

*K* = Number of clusters.

*p* = The point in space representing a given object.

*Step 3:* Compute the positions of the *K* centroids.

*Step 4:* Repeat *Steps 2* and *3* until the centroids do not change.

### D. Clustering Validation

#### 1) Distance of cluster

The average distance of the corresponding points, between each historical data and the centroid, can be calculated by applying Euclidean distance to measure the difference within the cluster. The equation can be explained as follows;

$$distance = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2)$$

where;

*p* is the historical data.

*p* = (*p*<sub>1</sub>, *p*<sub>2</sub>, *p*<sub>3</sub>, ..., *p*<sub>*n*</sub>)

*q* is the centroid.

*q* = (*q*<sub>1</sub>, *q*<sub>2</sub>, *q*<sub>3</sub>, ..., *q*<sub>*n*</sub>)

#### 2) Single-member cluster

Clustering algorithm will group data on the basis of similarities or same class. The amount of members in each cluster should more than one data. According to a cluster has only one member, the accuracy is higher than normal. Therefore, single-member cluster is commonly used to validate the clusters' quality [15].

#### 3) Cluster density

Cluster density is about the amount of members in the cluster which are grouped by regions where observations are dense and similar. To be the best cluster is based on the density of its' members. The similarity measurement of density is used to reflect the data distribution characteristics

[21]. The density-sensitive similarity measure is a common used for computing the density of the objects [22]. In this research, the density measurement is used to verify members of each cluster.

### III. METHODOLOGY

The proposed research methodology is to improve student loan fund consideration performance with procedures shown in Fig 1.

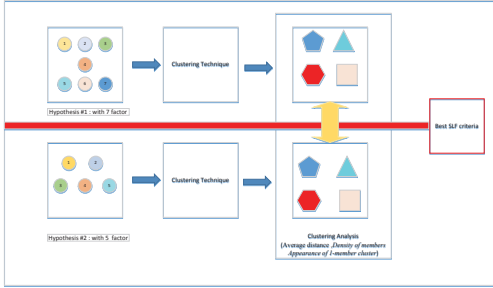


Fig.1. Proposed methodology with two experimental scenarios

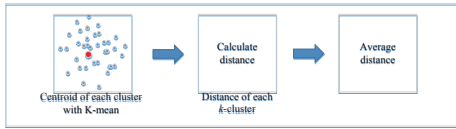


Fig.2. The model construction of  $K$ -mean clustering

#### A. Research Framework

This paper will analyse the informative of several factors using the clustering technique. The analysis is executed on the Weka 3.6.13 to improving student loan fund consideration performance. The research selects the  $K$ -mean clustering technique to enhance the student loan analysis and compare the clustering results based on distance, density, and single-member cluster for both hypotheses.

#### B. Data Collection

The datasets are the historical data of student loan candidates collected between 2014 and 2015. There are 1,146 samples. The factors for student loan decision are shown in the table. I. This data can be processed to be two different scenarios as follows;

*Scenario#1:* The data consists of seven factors; supporter who provides a financial for the student, the family's income, debt, education level of the sibling, the amount of money, the student got from other scholarship, including two bias factors (the positive score from the committee and the negative score from the committee.)

*Scenario#2:* The data consists of five factors; supporter who provides a financial for the student, the family's income, debt, education level of the sibling, the amount of money, and the student got from other scholarship.

TABLE I. EXAMPLE OF DATA SETS

Data set	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
1	2	3.00	3	2	0	6	-2
2	1	14.50	1	2	0	2	0
3	2	1.00	2	2	5	4	0
⋮	1	11.90	1	6	0	8	0
⋮	1	13.00	1	1	0	4	0
n	2	3.00	1	5	2	6	0

Table I shows the examples of all factors from the historical dataset of student loan candidates. The factors in table I consists of;

- $X_1$  = supporter who provide a financial for student,
- $X_2$  = family's income,
- $X_3$  = debt,
- $X_4$  = education level of the sibling,
- $X_5$  = amount of money the student got from other scholarship,
- $X_6$  = positive score from the committee and
- $X_7$  = negative score from the committee.

In this study,  $X_6$  and  $X_7$  are hypothesized as the bias factors which given by the committees' feeling.

#### C. K-mean Clustering

Fig.1 show the proposed methodology with three main procedures for two hypotheses. The clustering technique will process within three steps as Fig.2. Firstly, group the data with the concrete number of cluster ( $k$ ). Secondly, calculate clusters' average distance and overall distance from the centroid. Finally, compare the results of clustering analysis. The improvement will be evaluated by an average distance, density of member, and appearance of the single-member cluster.

## IV. EXPERIMENTAL RESULT AND ANALYSIS

#### A. Average Distance

The overall average distance for each  $k$  clusters can be summarized and shown in table II. The average distances of all seven factors (*Scenario#1*) are greater than average distances of five factors (*Scenario#2*) that means the similarity

between members have less relationship than the second scenario.

After testing the clustering technique with the dataset, the members of the cluster can be partitioned into 16 separated clusters of output as shown in table III and table IV.

TABLE II. THE OVERALL AVERAGE DISTANCES

Number of cluster	Overall average distances	
	All factors	Without bias factors
n = 2	5.58	4.83
n = 4	5.44	5.75
n = 8	7.75	7.41
n = 16	8.10	6.82

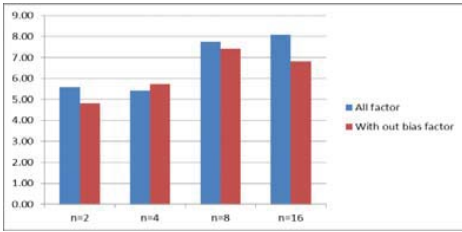


Fig.3. Comparison of the overall average distances of  $k = 2, 4, 8$  and  $16$

The comparison graph in Fig.3 shows the overall average distances of  $k = 2, 4, 8$  and  $16$ . The number of cluster

in *Scenario#2* with  $k = 2$  represents the shortest density using *K*-mean. Thus, this shortest distance represent the highest similarity and best performance to analyze the student loan by this clustering approach.

*B. Appearance of Single-member Cluster*

After used the clustering techniques to partition the dataset. The results are shown in table III and IV. There was no single-member cluster presented within every clusters. If a cluster has only one member, the accuracy will be higher than normal. Therefore, *K*-mean clustering with  $k=2$  can represent the best student loan fund consideration criteria with the *Scenario#2* of data that has no single-member cluster.

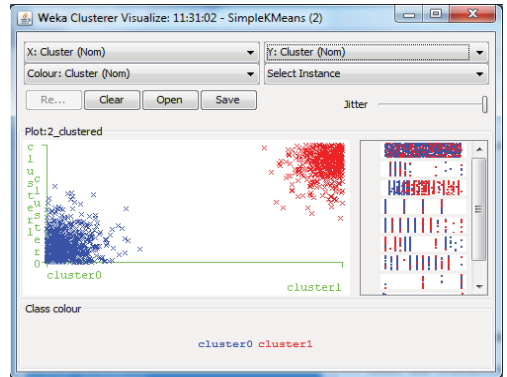


Fig.4. The density of data on *Scenario#1* with  $k = 2$

TABLE III. MEMBERS OF CLUSTER FOR SCENARIO#1

All FACTORS : Members of cluster																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n=2	719	427														
n=4	257	370	319	200												
n=8	214	40	16	147	315	106	80	168								
n=16	124	5	16	39	177	110	63	97	11	104	51	79	84	75	24	87

TABLE IV. MEMBERS OF CLUSTER FOR SCENARIO#2

WITHOUT BIAS FACTORS: Members of cluster																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n=2	732	414														
n=4	249	253	385	259												
n=8	167	40	74	201	265	153	82	164								
n=16	139	8	47	45	159	158	62	28	10	16	67	69	121	77	22	118



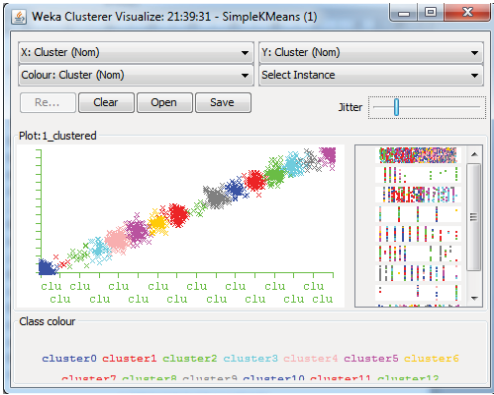


Fig.5. The density of data on Scenario#1 with  $k = 16$ .

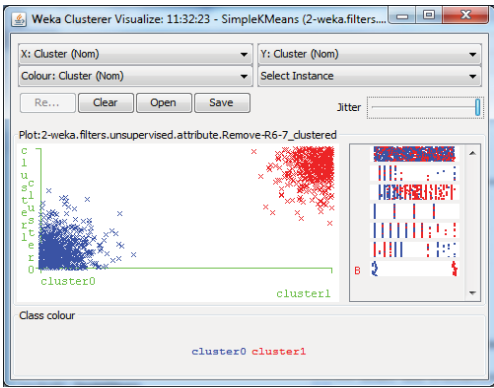


Fig.6. The density of data on Scenario#2 with  $k = 2$ .

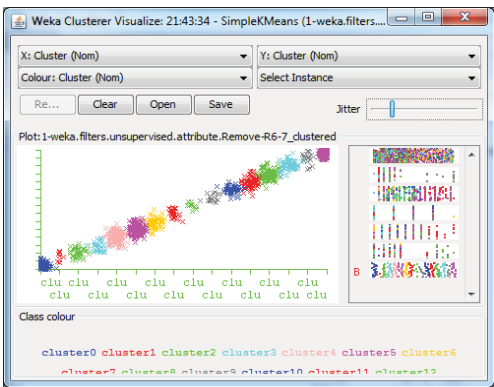


Fig.7. The density of data on scenario #2 with  $k = 16$ .

C. Cluster Density

The best cluster density can be defined as areas of greater amount of members. The density of data sets without bias factors (*Scenario#2*) has higher density than the data sets of all factors (*Scenario#1*) as shown in Fig.4 to Fig.7. Therefore, without bias factors, the quality of clustering was performed better than clustering of all factors.

V. CONCLUSION

This research present the improvement and verification of all related factors of the student loan information by using *K*-mean clustering technique with different numbers of clusters (*k*) as 2, 4, 8 and 16. The method begins with collecting the data which are historical data of student loan candidates applied between 2014 and 2015 and preparing to be two datasets. The first dataset was consisting of all factors hypothesised as parameters with bias factors (*Scenario#1.*) The second dataset, *Scenario#2*, was without bias factors ( $x_6, x_7$ ). This research used the Weka 3.6.13 to calculate the centroid for each cluster. After that, a measurement by using the distance between each member and the centroids by using the Euclidean distance was processed, including calculation of the overall average distance for each *k* cluster. The last process in the proposed methodology was to compare the distances between datasets from two scenarios. The results revealed that the dataset of *Scenario#2* has the shortest average distance with a number of clusters (*k*) equal to 2. These experimental results also show the relation of a shortest distance which represents the highest density and best performance for analyzing the student loan candidates. Thus, the student loan consideration based on analysis of only true data (*Scenario#2*) has better performance than the consideration of data which included the human thought.

ACKNOWLEDGMENT

This research is sponsored by Mae Fah Luang University. The authors of this paper are highly grateful to Dr. Chayapol Kamyod in providing the technical assistance.

REFERENCES

- [1] J. Hu, "Personal credit rating using Artificial Intelligence technology for the National Student Loans," in *4th International Conference on Computer Science Education, 2009. ICCSE '09, 2009*, pp. 103–106.
- [2] Z. Zhao, W. Zhang, and Y. Zhou, "National student loans credit risk assessment based on GABP algorithm of neural network," in *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011*, pp. 2196–2199.
- [3] H. Wei, "An International Estimation on Multiattribute of Student Loan Scheme Based on Entropy Theory," in *2008 International Conference on Information Management, Innovation Management and Industrial Engineering, 2008*, vol. 3, pp. 8–12.

- [4] W.-R. Shi, Y.-H. Lin, Y. Li, T.-N. Yang, and L. Zhang, "A matter-element evaluation model for the individual credit of university students loan," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics, 2004*, 2004, vol. 5, pp. 2853–2858 vol.5.
- [5] B. Chapman, K. Lounkaew, P. Polsiri, R. Sarachitti, and T. Sitthipongpanich, "Thailand's Student Loans Fund: Interest rate subsidies and repayment burdens," *Economics of Education Review.*, vol. 29, no. 5, pp. 685–694, Oct. 2010.
- [6] D. Zhao, "Integrating RFM Model and Cluster for Students Loan Subsidy Valuation," in *International Seminar on Business and Information Management, 2008. ISBIM '08*, 2008, vol. 2, pp. 461–464.
- [7] Maureen Woodhall, "Student loans in higher education 2. Asia," *International Institute for Educational Plannin.*, 1991.
- [8] D. Zhao, "Integrating RFM Model and Cluster for Students Loan Subsidy Valuation," in *International Seminar on Business and Information Management, 2008. ISBIM '08*, 2008, vol. 2, pp. 461–464.
- [9] J. Zurada, "Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions?," in *2010 43rd Hawaii International Conference on System Sciences (HICSS)*, 2010, pp. 1–9.
- [10] A. Çaliş, A. Boyacı, and K. Baynal, "Data mining application in banking sector with clustering and classification methods," in *2015 International Conference on Industrial Engineering and Operations Management (IEOM)*, 2015, pp. 1–8.
- [11] Sreedhar Ventrapragada and Venkata, "Predictive Mining on Loan Data using Rattle.," *International Journal of Advanced Research in Computer Science.*, vol. Sep/Oct2013, no. 10, pp. 223–231.
- [12] E. Kambal, I. Osman, M. Taha, N. Mohammed, and S. Mohammed, "Credit scoring using data mining techniques with particular reference to Sudanese banks," in *2013 International Conference on Computing, Electrical and Electronics Engineering (ICCEEE)*, 2013, pp. 378–383.
- [13] R. V. Singh and M. P. S. Bhatia, "Data clustering with modified K-means algorithm," in *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2011, pp. 717–721.
- [14] N. Chen, Z. Xu, and M. Xia, "Hierarchical hesitant fuzzy K-means clustering algorithm," *Applied Mathematics-A Journal of Chinese Universities.*, vol. 29, no. 1, pp. 1–17, Mar. 2014.
- [15] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
- [16] W. Rueangsirarak, A. S. Atkins, B. Sharp, N. Chakpitak, K. Meksamoot, and P. Pothongsunun, "Clustering the clusters – knowledge enhancing tool for diagnosing elderly falling risk," *International Journal of Healthcare Technology and Management*, vol. 14, no. 1–2, pp. 39–60, Jan. 2013.
- [17] Sonam Narwal and Kamaldeep Mintwal, "Comparison the Various Clustering and Classification Algorithms of WEKA Tools," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2013, pp. 866–878.
- [18] A. Kitipong, W. Rueangsirarak, R. Chairicharoen, and R. Banchuin, "Batik price estimation system for textile production," in *2014 12th International Conference on ICT and Knowledge Engineering (ICT and Knowledge Engineering)*, 2014, pp. 48–53.
- [19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [20] Klangwaree Chaiwut, Worasak Rueangsirarak, and Rounsang Chairicharoen, "Fault Detection in Student Loan Analysis using Clustering Techniques," *International Conference on Digital Arts, Media and Technology.*, 2016, pp. 200–203.
- [21] Q. Yuan, H. Shi, and X. Zhou, "An optimized initialization center K-means clustering algorithm based on density," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2015, pp. 790–794.
- [22] YAO Yue-hua and SHI Xiu-ling, "K-means rough clustering algorithm based on optimized initial center." *Computer Engineering and Applications.*, 2010, pp. 126-128.

# Overview and challenges with HF RFID systems

Kevin D'hoë\*,

Nobby Stevens\*, Jean-Pierre Goemaere\*,  
Bart Nauwelaers† and Lieven De Strycker\*

\*DraMCo research group, ESAT,  
Faculty of Engineering Technology  
KU Leuven,

Gebroeders De Smetstraat 1, 9000 Ghent, Belgium

E-mail: kevin.dhoe@kuleuven.be

†Department of Electrical Engineering,  
ESAT, KU Leuven,

Kasteelpark Arenberg 10, 3001 Leuven, Belgium

E-mail: bart.nauwelaers@esat.kuleuven.be

**Abstract**—In spite of the fact that HF RFID is one of the oldest identification technologies, we do want to draw attention on new opportunities and challenges with HF RFID systems. Thanks to research on wireless power transfer systems and related RFID frequency domains some ideas are worth a try in HF RFID. These HF RFID systems still have their value in industrial environments, in which metals and liquids are frequently present. In this paper, an overview on the main challenges with HF RFID systems is given, together with the latest developments related to these issues. When the innovative research can brought together in commercial products, we can improve the reliability of HF RFID systems, resulting in new applications.

## I. INTRODUCTION

Whether we realize it or not, Radio Frequency Identification (RFID) is a part of our daily life. RFID increases the efficiency and improves the quality of it. This technology is used for hundreds of applications such as collecting tolls on the highway; controlling access to buildings; preventing theft of merchandise; tracking library books; monitoring the air pressure in tires; timing marathons and races; validating rail passes; paying with NFC enabled smartphones and off course, tracking assets in supply chain management. A variety of radio frequencies and techniques are used for these RFID applications.

RFID is generally characterized by two types of devices on each side of a wireless link. The simple devices, i.e. transponders, are small and inexpensive. They are typically attached to objects and operate automatically. Typically, the ID of the transponder is read by a more complex device, i.e. a reader, which is usually connected to a host computer or network.

Two main principles to couple the transponder with the RFID reader are used, namely inductive coupling and electromagnetic coupling. The inductive coupling method is primary used for relative short distances which goes up to 1 meter between the transponder and the antenna, this principle is based on the well-known voltage transformer coupling. The

electromagnetic principle uses microwaves and is used when a higher reading distance is preferred. Depending on the application, different technologies and variants of RFID are used. Besides that, transponders are found in all kind of forms and can be attached to object in different ways, e.g. stickers, glass tubes, smart cards, ...

Within this paper we will focus on inductively coupled systems, this technology is the oldest one in the wide range of RFID technologies but it deserve some new insights on the possibilities of it. After a brief history and an overview on the architecture of inductive coupled systems. We will discuss the capabilities of the HF RFID systems and the main challenges of this technology to reach it full potential.

## II. A BRIEF HISTORY

RFID is an old technology. The principle of identification using radio signals is found in the Second World War, i.e. the use of RADAR (Radio Detection and Ranging). Germans, Japanese, Americans and British forces all used this technology to identify approaching airplanes. The problem was that it lacked to a method to identify friendly from enemy airplanes. But soon, the Germans discovered that a rotating movement of the airplane caused a change in the reflected radio signal on the RADAR, which is actually the same as modulating a signal. This way they could identify their own airplanes. In essence, this was the first passive RFID system. The technology was invented by Watson Watt in 1935.

At the same time the British forces were testing a system that could influence the radar reflection of airplanes by means of a passive dipole antenna on the airplane body. The system worked very well for a single airplane, although when flying in formations they could not distinguish between individual airplanes. To solve this problem, an active transponder was introduced in 1939. The so-called Mk I (and later the Mk II) IFF (Identification of Friend and Foe) sent out pulses at the frequencies of all the radars in the United Kingdom.

Radar and RF communication systems kept developing through the 1950s and 1960s. Scientists and academics did research and presented papers explaining how RF energy could be used to identify objects remotely [1], [2], [3]. Companies began commercializing anti-theft systems that used radio waves to determine whether an item had been paid for or not. These electronic article surveillance (EAS) systems, which are still used in packaging today, have a 1-bit transponder. The bit is either set or not. If someone pays for the item, the bit is turned off, and a person can leave the store. In the other case, readers at the door detect the tag and sound an alarm [4]

In the 1970s, the addition of a unique code in a transponder was a step forward towards the RFID systems we know these days. Over time, companies commercialized 125 kHz systems and then moved up the radio spectrum to high frequency (13.56 MHz). High frequency offered greater range and faster data transfer rates. Companies, particularly those in Europe, began using it to track reusable containers and other assets.

The 1990s were another important decade for RFID since a wide scale deployment took place. The United States introduced electronic toll collection with RFID transponders and at the same time North America installed over 3 million RFID transponder on rail cars. While in that same period, IBM engineers developed and patented an ultra-high frequency (UHF) RFID system. UHF offered longer read range (up to 5 meters under good conditions) and faster data transfer.

In 1999, the Auto-ID Center at the Massachusetts Institute of Technology was set up by a number of organisations which meant the creation of a common standards. Also the International Standards Organisation (ISO) introduced standards for the different elements of RFID. The biggest step forward in the history of RFID took place in January 2005, when Wal-Mart recommended their top suppliers to apply RFID labels on all their products.

### III. ARCHITECTURE OF A HF RFID SYSTEM

There are three main parts which forms a complete RFID system: a transponder, a reader and the antenna.

The word 'transponder' is two-fold: you can recognize *transmitter* and *responder* in it. This means that the transponder not only receive data, but also send data to the reader.

Every transponder has a unique ID which is assigned on the fabrication of the transponder. The antenna of the transponder, which actually is a coil, has a double function. At one side it is used to collect and supply energy to the chip, at the other side the antenna is used to exchange data with the reader.

In HF RFID systems only passive transponder are used which means that there is no power source available on the transponder. All energy to power the chip and to communicate with the reader comes from the reader antenna. The reading distance is low (up to 1 meter) and is depended on the power supplied by the reader and the specification of the transponder. The size and shape of the reader antenna could hereby play an important role in the power delivery to the transponder.

At the interrogator side, we find the reader and the antenna. Basically we have 2 categories, in most cases the reader is

fixed at a certain position with an external antenna. The reader is than connected to a computer or a data center. These types of readers are found at checkpoints in shops and warehouses where items and goods has to be traced. In other cases, we have portable readers (handheld readers) which are easy to carry and where the antenna is integrated in the mobile device. The data is saved locally or wirelessly transmitted to a central unit. They are typically used for inventory management of small stockrooms, e.g. to verify the drug supply in the hospital.

### IV. HF RFID, WHAT TO EXPECT?

High frequency 13.56 MHz technology was one of the first forms of RFID to be successfully commercialized. Therefore, it became the most popular RFID technology in the market and it is still now [5]. Moreover, the frequency provides excellent resistance to common sources of electrical interference found in industrial environments [6]. The only drawback of 13.56 MHz technology is its limited range of about 1 meter.

This frequency is supported by various standards. The most popular ones are the ISO 14443 and the ISO 15963. The first, is used for vicinity cards, which means they operate within 10 cm of the reader antenna. This standard is mainly used for access control and public transportation systems. The ISO 15693 standard refers to proximity cards and transponders, where we aim for a higher range. This standard is seen as an ideal solution for numerous applications because it benefits from its read range.

The popularity and wide applicability of this standard also had some drawbacks. As a lot of companies where not familiar with the limitations and technical aspects of this technology, it led to failures in some industrial environments.

The shortcomings specific to 13.56 MHz technology for identifying goods in distribution and supply chain operations led some to conclude that RFID in general was an unreliable or ineffective technology [7]. These experiences stunted the intense developments of HF RFID for a while now, and consequently a lot of companies moved to the UHF RFID domain which again has his pros and cons when using it in industrial environments.

In the research and development we performed for companies, we also experienced the problems of the companies implementing an RFID system. As they wanted to appeal on our knowledge of wireless and mobile communication systems, we learned from the implementation mistakes they made. It gave us the opportunity to gain insight in the challenges of this technology. Most of the time, the problems were situated at the reader loop antenna. Retuning the antenna, placing the antenna in another place or orientation, changing the size of the antenna or even the power output of the reader are common solutions.

Most users have high expectations of RFID systems, in their opinion it is only useful if a 100% successful reading is guaranteed. Vendors of RFID systems do promise a very reliable system and actually they are not lying. RFID systems can work perfectly in most circumstances and situations. These reliable situations are most likely to be found in a free

space environment with a well controlled detection of the transponders, i.e. the read range, the amount of transponders, the orientation, the speed, etc.

The RFID system is put to the test if one of previous conditions change. This does not mean it will not work, but there is a chance that the transponder is not read by the reader. There are many aspects which influence the proper functioning of the systems and thus many academic work is still on the go to overcome these 'shortcomings' in HF RFID systems.

#### V. GENERAL CHALLENGES WITH HF RFID SYSTEMS

Improvements are possible on different aspects of this RFID technology. We will draw attention on the main challenges within HF RFID systems in order to discuss the latest developments.

First of all, the **limited read range** is worth a closer look. As we rely on the inductive coupling principle, the maximum read range limit is 3.5 m. This is related to the operating frequency and is defined by the field regions of a loop antenna, i.e. where the reactive near-field stops and merges into the Fresnel zone [8]. As most systems only reach a maximum of 1 m, there is some room for improvement. Of course, we can increase the power at the reader but this is limited by the European Telecommunications Standards Institute (ETSI). This means, we have to take a look at the receiver side, i.e. the transponder. The transponder captures the energy out of the magnetic field via an antenna coil. This antenna is connected to the analog front-end of the transponder which delivers energy to the digital blocks. The efficiency of the analog front-end is therefore directly responsible for the read range of the transponder. This can be improved in several ways, in which all of the following work contribute to an increased reading range.

Paixao Cortes et al. [9] presented a low power, low voltage RF/analog front-end architecture for LF RFID tags with a dynamic power sensing scheme. The front-end converts the incoming RF power into DC power using a power management system that adjusts its performance according to the available RF power. The power sensing scheme improves the available power going to the system. Also, the clock extractor optimizes the system with respect to the modulated signal produced during data modulation. It results in an increased reading distance and a gain of around 15% in the available power.

Dongsheng et al. [10] created a high sensitivity analog front-end for a semi-passive HF RFID tag. A rectifier with high power conversion efficiency is presented to provide a stable rectified voltage. Moreover they proposed a novel tag system architecture with a wake-up circuit to improve the sensitivity. Simulation results show that the new analog front-end design has realized much longer recognition distances than regular passive tags, while its power consumption is as low as 129.6  $\mu$ W.

If we look outside the RFID domain, we also find interesting work on the analog front-end design. For example within the biomedical domain, implanted devices with functions such as neural recording and/or stimulation usually need power.

Therefore, biomedical implants with a highly-efficient wireless power transfer scheme that could deliver real-time power in the mW range with a small form factor are in great demand. Hashemi et al. [11] and Lu and Ki [12] both did some work on the rectifier circuit which allows improving the overall power efficiency.

RFID systems come to their full potential when they are able to read a **high amount of transponders** at one time. As anti-collision protocols are integrated into the software of the reader, especially created to read several transponders at once [13]. Yet there is a physical phenomena that has to take in to account and can cause a lot of problems, i.e. the loading effect.

The loading effect takes place at the reader loop antenna, the antenna has to be tuned at the resonance frequency and matched to the characteristic impedance of the reader. This is a critical process to create a proper working system. Each transponder creates an extra load to the system which will cause a change in the impedance of the loop antenna. A single transponder will not have much impact but a lot of them in the vicinity of the loop antenna will cause a major change in the impedance and the resonance frequency, which will lead to a drop in the read range or the malfunctioning of the system. The impact of the loading effect is well described in several papers [14], [15].

Solutions to this problem are situated at the matching unit. Jiang et al. [16] created an adaptive matching network as a solution to mitigate the detrimental effect of the mutual load on the power transfer efficiency. The disadvantage of these adaptive matching networks is the speed of adaptation. As it take some time to find the best matching network, it can not be used when transponders are moving with a certain speed or when time is limited for a correct reading.

Despite the fact that they do not mention the loading effect in their work, Soodman et al. [17] created a loop antenna (matching network) with a wider bandwidth, still having an excellent read range. Initially they wanted to create a loop antenna which is suitable for a wide range of HF transponders, i.e. ISO 14443 and ISO 15693 transponders, as they have different requirements towards the reader loop antenna. Consequently, they created a loop antenna which is much more robust for changes in the amount of transponders which is interesting to create a reliable RFID system.

When a transponder is read, we have to keep in mind that the transponder has a preferred **orientation** relative to the plain of the loop antenna. If the transponder is close to the reader antenna, it is not really an issue. But if the magnetic field is relatively small for whatever reason, the orientation of the transponder could play a decisive roll whether or not the transponder is read.

To prevent these so called null zones, two opposing antennas are used. The field is now created from both sides, allowing bigger distances. Moreover, if we feed one antenna with a

signal that is  $90^\circ$  out-of-phase to the opposing antenna, we will get a 'rotating field' which decrease the chance of missing a transponder [18]. Still it is not a foolproof solution, these techniques are therefore combined with multiple antennas in a gate or tunnel arrangement. Examples of these configurations are found in [19], [20], [21], [22], [23].

The gate and tunnel configurations do enhance the reliability of the total system but when the space is limited, they are not an option. Hirayama et al. [24] proposed a one plain antenna, without expanding the antenna size of a regular antenna. A loop antenna is split into four parts, which work as an array antenna. Phase inverters are introduced on each part of the splitted loop antenna and are controlled to obtain maximum power at the RFID transponder. Validation of their model with simulations, could demonstrate an enlargement of the power at the transponder compared to a traditional loop antenna.

Last but not least, the **environment** has an important impact on the reliability of the system. With limited space available or a metallic environment surrounding the loop antenna, we can not always create the ideal alignment between the loop antenna and the transponders, although not with the commercial available antennas. These antennas are usually round or square with a fixed size, what limits us in the adaptations of the setup.

It is clear that we need to fill the gap between the standard commercial available antennas and the antennas needed for some specific but valuable RFID applications/implementations.

In our own research, we obtained the same performance of multiple antennas with a single loop antenna. Therefore, we need to adapt the shape of the antenna to its surroundings and finally step away from the traditional shapes of commercial antennas. By doing so, we can change the behaviour of the magnetic field and acquire some specific goals with a single antenna [25], [26], [27].

The environment also has its impact on the transponder, especially in metallic environments. The solution to this problem is best known as the mount-on-metal transponder [28]. This type of transponder is made to place on a metallic surface. Where normal transponders wouldn't work, this transponder has a good read range on a metallic surface. This is made possible with an additional layer between the transponder and the metallic surface, a layer of ferrite [29], [30].

Ferrite is an expensive material, therefore it was never considered as a solution for commercial reader loop antennas. Though, it seems to prove its usefulness, as Wielandt et al. [31] combined a copper plate with a thick, highly permeable ferrite layer in an LF RFID system. This combination even led to an increase of the magnetic field strength.

In recent research a new method has been applied to counteract the effect of a metallic environment on an antenna. The idea of using an anti-parallel resonant coil for near-field powering was first introduced by Lee et al. [32]. It is a very interesting method to counteract the degrading effect of the Eddy currents, moreover it is a low cost and easy-

to-implement solution to increase the reliability of inductive coupled systems.

A variant of this techniques was later applied on an HF loop antenna by Mukherjee S. [33], where an additional resonant loop, the compensation loop, is able to mitigate the effect of metal. The use of an extra resonant loop is know for the enhancement of the magnetic field but it is never used to encounter the effects of a metallic environment on the loop antenna. Both, papers showed interesting results for future implementations in commercial loop antennas. These novel techniques can replace expensive ferrite material and guarantee a similar efficiency of the RFID system in metallic environment.

## VI. CONCLUSION

It is clear that there are still some challenges with HF RFID systems and broader than that; inductive powered system. At this moment a lot of the research, based on inductive coupled systems, is concentrated on the wireless power transfer (WPT) technology. This technology is still young and has a bright future in innovative applications.

New ideas, like the antiparallel resonant coil, have their roots in WPT technology. Now, they need to find their way back to HF RFID systems and ultimately lead to a boost of renewed RFID products. A combination of the latest research could open a wide range of new applications for HF RFID systems.

## REFERENCES

- [1] H. Stockman, "Communication by means of reflected power," *Proceedings of the IRE*, vol. 36, no. 10, pp. 1196–1204, Oct 1948.
- [2] A. R. Koelle, S. W. Depp, and R. Freyman, "Short-range radio-telemetry for electronic identification, using modulated rf backscatter," *Proceedings of the IEEE*, vol. 63, no. 8, pp. 1260–1261, Aug 1975.
- [3] R. F. Harrington, "Theory of loaded scatterers," *Electrical Engineers, Proceedings of the Institution of*, vol. 111, no. 4, pp. 617–623, April 1964.
- [4] B. Violino, (2005) The history of rfid technology. [Online]. Available: <http://www.rfidjournal.com/articles/view?1338>
- [5] IDTechEx, "Rfid forecasts, players and opportunities 2016-2026," Webinar, November 2015.
- [6] S. Serkan Basat, K. Lim, J. Laskar, and M. Tentzeris, "Design and modeling of embedded 13.56 mhz rfid antennas," in *Antennas and Propagation Society International Symposium, 2005 IEEE*, vol. 4B, July 2005, pp. 64–67 vol. 4B.
- [7] Intermec, (2007) Uhf vs. hf rfid: New insight on the old debate. [Online]. Available: <http://www.intermec.com/>
- [8] I. Canada, *TN-261 Safety Code 6 (SC6) Radio Frequency Exposure Compliance Evaluation Template (Uncontrolled Environment Exposure Limits)*, Spectrum Management and Telecommunications, March 2015.
- [9] F. Paixao Cortes, J. Martinez Brito, R. Cantalice, E. Ghignatti, A. Olmos, F. Chavez, and M. Lubaszewski, "A low-power rfid/analog front-end architecture for lf passive rfid tags with dynamic power sensing," in *RFID (IEEE RFID), 2014 IEEE International Conference on*, April 2014, pp. 60–66.
- [10] L. Dongsheng, L. Huan, Z. Xuecheng, G. Liang, Y. Ke, and L. Zilong, "A high sensitivity analog front-end circuit for semi-passive hf rfid tag applied to implantable devices," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 62, no. 8, pp. 1991–2002, Aug 2015.
- [11] S. Hashemi, M. Sawan, and Y. Savaria, "A high-efficiency low-voltage cmos rectifier for harvesting energy in implantable devices," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 6, no. 4, pp. 326–335, Aug 2012.

- [12] Y. Lu and W.-H. Ki, "A 13.56 mhz cmos active rectifier with switched-offset and compensated biasing for biomedical wireless power transfer systems," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 8, no. 3, pp. 334–344, June 2014.
- [13] Q. Zhihong and W. Xue, "An overview of anti-collision protocols for radio frequency identification devices," *Communications, China*, vol. 11, no. 11, pp. 44–59, Nov 2014.
- [14] C. Marechal and D. Paret, "The loading effect of proximity contactless smart card. incidences of impedance of the shunt regulator," in *Wireless Communications, Networking and Mobile Computing, 2008. WICOM '08. 4th International Conference on*, Oct 2008, pp. 1–4.
- [15] M. Chaoui, H. Ghariani, M. Lahiani, F. Sellami, R. Perdraiu, and M. Ramdani, "Reducing load effects in inductive links for implantable systems," in *Intelligent Engineering Systems, 2005. INES '05. Proceedings. 2005 IEEE International Conference on*, Sept 2005, pp. 117–122.
- [16] B. Jiang, J. Smith, M. Philipose, S. Roy, K. Sundara-Rajan, and A. Mamishev, "Energy scavenging for inductively coupled passive rfid systems," in *Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE*, vol. 2, May 2005, pp. 984–989.
- [17] S. Soodmand, T. Brown, and A. Gluhak, "Evaluation of hf band nfc/rfid antennas for smart shelf applications," in *Antennas and Propagation (EuCAP), 2013 7th European Conference on*, April 2013, pp. 1895–1898.
- [18] *HF Antenna Design Notes - Technical Application Report*, Texas Instruments, September 2002, 11-08-26-003.
- [19] G. Benelli, S. Parrino, and A. Pozzebon, "Possible configurations and geometries of long range hf rfid antenna gates," in *Wireless Communication Systems, 2009. ISWCS 2009. 6th International Symposium on*, 2009, pp. 46–50.
- [20] J. Bonne, K. D'hoë, T. Hamelincx, N. Stevens, J. Goemaere, and L. de Strycker, "Design of an rfid antenna gate with multiple loop antennas," in *RFID-Technologies and Applications (RFID-TA), 2014 IEEE International Conference on*, Sept 2014.
- [21] S. Kawdungta, C. Phongcharoenpanich, and D. Torrungrueng, "Design of flat spiral rectangular loop gate antenna for hf-rfid systems," in *Microwave Conference, 2008. APMC 2008. Asia-Pacific*, Dec 2008, pp. 1–4.
- [22] S. Kawdungta, P. Wouchoum, C. Phongcharoenpanich, and D. Torrungrueng, "Improvement of communication performance for trapezoidal dual loop antennas by switching techniques," in *Microwave Conference, 2007. APMC 2007. Asia-Pacific*, Dec 2007, pp. 1–4.
- [23] *13.56 MHz Tunnel-Antenna for 13,56 MHz Long-Range reader/writer SHL-2100*, Scemtec.
- [24] H. Hirayama, Y. Satake, N. Kikuma, and K. Sakakibara, "Improvement of null zone avoidance capability for hf-band rfid using diversity combining of loop antennas," in *Antennas and Propagation, 2009. EuCAP 2009. 3rd European Conference on*, March 2009, pp. 1594–1597.
- [25] K. D'hoë, N. Stevens, J. Goemaere, L. de Strycker, and B. Nauwelaers, "Optimization of an rfid loop antenna with smart goal functions," in *RFID-Technologies and Applications (RFID-TA), 2012 IEEE International Conference on*, Nov 2012, pp. 254–258.
- [26] K. D'hoë, J.-P. Goemaere, N. Stevens, B. Nauwelaers, and L. De Strycker, "Automated design of an hf rfid loop antenna based on parametric geometry modification," in *RFID (IEEE RFID), 2014 IEEE International Conference on*, April 2014, pp. 1–7.
- [27] K. D'hoë, J. P. Goemaere, N. Stevens, B. Nauwelaers, and L. D. Strycker, "Optimized shape of a parametric designed hf loop antenna," in *Signal Processing and Communication Systems (ICSPCS), 2015 9th International Conference on*, Dec 2015, pp. 1–5.
- [28] *MOUNT-ON-METAL TRANSPONDER*, Texas Instruments, November 2008.
- [29] S. Bovelli, F. Neubauer, and C. Heller, "Mount-on-metal rfid transponders for automatic identification of containers," in *Microwave Conference, 2006. 36th European*, Sept 2006, pp. 726–728.
- [30] H. Zhu, S. Lai, and H. Dai, "Solutions of metal surface effect for hf rfid systems," in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, Sept 2007, pp. 2089–2092.
- [31] S. Wielandt, D. Mercy, N. Stevens, L. De Strycker, and J.-P. Goemaere, "Evaluation of shielding materials for low frequency rfid systems," in *Electromagnetic Compatibility (EMC EUROPE), 2012 International Symposium on*, Sept 2012, pp. 1–5.
- [32] W.-S. Lee, H.-S. Jang, K.-S. Oh, and J.-W. Yu, "Close proximity effects of metallic environments on the antiparallel resonant coil for near-field powering," *Antennas and Propagation, IEEE Transactions on*, vol. 61, no. 6, pp. 3400–3403, June 2013.
- [33] S. Mukherjee, "Mitigation of proximity to metal for magnetically coupled transponders by use of resonant loops," in *RFID (IEEE RFID), 2014 IEEE International Conference on*, April 2014, pp. 8–14.

# Realization of Global Precision Agriculture Monitoring System using WiFi-Based Sensor Nodes and Internet of Things Open Data Platform: A Case Study in Thailand

Thongchai Yooyativong  
 School of Information Technology  
 Mae Fah Luang University  
 Chiang Rai, Thailand  
 e-mail: thongchai.mfu@gmail.com

**Abstract**— Major industry in most countries still be agriculture. With advancement in network technology and sensor technology, precision agriculture becoming one major driving force to raise the production yield in agriculture. However, implementation of wide area precision agriculture system requires utilization of wide-area sensor network which can be complicated and costly if improper system architecture is selected. Most developing, low and middle-income agricultural countries faced this situation and make it impossible for them to implement, utilize, and benefit the concept of precision agriculture in large farming. However, with the fast research and development of open platform microcontroller unit and Internet of Thing platform including the reduction of cost and size of sensors, the implementation of wide area precision agriculture monitoring and control system become less complicated and potentially more feasible for most low and middle-income agricultural countries. The purpose of this paper is to propose one basic approach for low and middle-income agricultural countries of setting up a precision agriculture monitoring system using WiFi-based sensor nodes interconnected with the internet of things open data platform by using Thailand as a case study.

**Keywords**—Precision Agriculture, Sensor, WiFi, Smart, Internet of Thing, Platform, Monitoring

## I. INTRODUCTION

Agriculture is still being the main industry in most agricultural countries such as Thailand. However, most of the farmers in agricultural countries still are in the low-level income due to the type of crops they are growing and the lack of using appropriate technology in their agricultural process. For the past few years, in term of agricultural process, precision farming seems to be part of major contributions to enhance the efficiency of growing process, increase production yields and also reduce the cost of production [1].

Precision farming sometimes called “Smart Farming” use an embedded system that composed of sensors and microcontroller unit to measure, monitor and control the microclimate of crop environment such that maximum production yield is obtained. With sensor technology, the measurement of farming environment can be very accurate and can be closely monitored and also precisely controlled. Typical sensors in precision agriculture are used to measure temperature, humidity, light intensity, soil moisture, pH, and EC. However, to cover and manage a large area of farming, a

network of sensors is required. The communication between sensors or smart devices leads to nowadays what we called Internet of Things (IoT). With IoT technology, information can be sent or monitor anywhere in the world. This opens up the opportunity for farmers to remotely monitor and control their farm through networked monitoring and controlling system such that the maximum production yield can be obtained.

However, for most agricultural developing countries, wide-area remote monitoring for precision farming seems very far reached and impractical due to the availability of related technology knowledge as well as the high-cost and complexity of system implementation. Fortunately, with the fast growing of IoT technology, there are many open IoT application platforms such as ThingSpeak, Thing Broker, Netpie, H.A.T, etc., that were developed to support IoT worldwide with simplicity and low cost of implementation [2]. This makes the implementation of precision farming monitoring system much more feasible in agricultural developing countries.

This paper is aimed to propose a practical implementation and a simple architecture design of the precision agriculture monitoring system that will integrate the typical agricultural sensor system with open IoT platform so that farmers in developing countries can implement a monitoring system for precision farming that is economically and technically feasible. Thailand is used as the scenario case for this study.

## II. LITERATURE REVIEW

Various fields of research and development efforts are aiming to improve agricultural monitoring and control system. These various fields include sensor technology, microcontroller platform, communication technology, software technology, internet of thing platform as well as system development approach. This due to requirements need to improve agricultural process such as weather data collection, soil and crop information, fertilizer and water requirement of distributed land, crops condition monitoring [1]. Wireless sensor network (WSN) presents itself as a strong candidate to satisfy these requirements. WSN comprises of several devices used for collecting required monitored data called ‘node’. Three fundamental functions of the sensor network are sensing required data, communication between



system components and computation of display data or control signal. For wide and uneven land, wireless communication is more practical for sensor network implementation. Wireless technology like ZigBee, Bluetooth, Wibree, and WiFi are commonly used in agriculture WSN. In designing WSN, critical issues that developers need to consider are energy consumption, data acquisition (sampling and transmission), fault tolerance, the size of housing, sensor placement, and security.

Smart or Intelligent farming is replacing traditional way of farming due to the ability of WSN that help farmers to monitor aware context parameters such as humidity, light intensity, temperature, air quality and soil fertilizer. This enables farmers to effectively manage their farm from anywhere and reduce cost from manpower [3]. The node of WSN comprises of sensors related to required contexts such as temperature and humidity sensor, air sensor, etc., microprocessor like Arduino or Raspberry Pi, and communication module such as ZigBee module [4].

In recent years, research and development of interconnecting smart devices (Things) are growing rapidly due to the foreses of future development such as smart home, smart city, digital health. The interconnection of smart devices with application and service refers to as Internet of Things (IoT). The infrastructure and middleware that enable farmers to interact with smart devices is referred to as IoT platform[5]. Currently, there are various IoT platforms available, both proprietary and open-source, such as LinkSmart, Node-RED, SkySpark, and ThingSpeak. The IoT platform essence is to enable seamless interconnection of heterogeneous sensors and actuators which can be satisfied by a pool of standard communication protocols such as AMQP, HTTP, Web socket, and MQTT [6].

III. SYSTEM ARCHITECTURE

In general, basic components of precision monitoring system platform are sensors, microcontroller, communication devices, communication platform, and display units. Wireless communication is more practical and widely used in a wide area and large farming. With this concept, the simplified diagram demonstrates components of a system architecture that composes of wireless sensor nodes, monitoring devices, wireless communication, and IoT open data platform (Fig. 1).

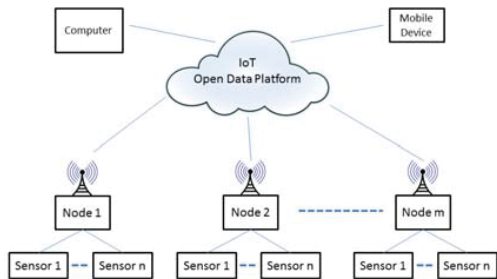


Fig 1. Simplified precision agriculture monitoring system with IoT platform

A. Wireless Sensor Node using open architecture microcontroller unit

Wireless sensor node is the foundation unit of precision agriculture. It generally composes of agricultural sensors (temperature, humidity, soil moisture, light intensity, wind flow, pH, EC, etc.), processing unit or microcontroller, wireless communication module, and power module (Fig 2.). A typical architecture of sensor nodes can be categorized as embedded multi-chip sensor node and system-on-chip (SoC) sensor node.

Typically, an embedded multi-chip sensor node consists of an application related sensor array, a microcontroller unit that is used as a processing plant, a communication transeiver unit and also memory units for data storing. The specific architecture of an embedded multi-chip sensor node varies to meet the application demands such as processing power, data storage size, parameters needed to be measured. This makes it more flexible in design selection. On the other hand, the system-on-chip (SoC) architecture provides an embedded integration of programmable processors, co-processors, memory units, peripheral I/O interface units and application specific units. The overall design is aim to minimize power consumption, support intensive computation and specific application. This design concept results in a smaller size sensor node but higher cost. Even though SoC sensor node has a lot of potential for agriculture application, it very rarely used.

With current sensor technology, most agricultural sensors are easily available, inexpensive and easily interface with microcontroller unit using standard communication protocols and provided interfacing-libraries from the manufacture. Example for varieties of sensor available in the market is shown in Fig 3.

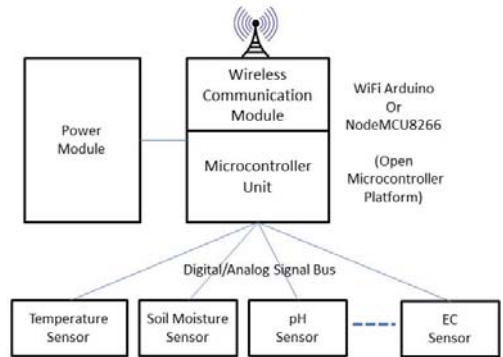


Fig 2. Sensor node architecture using WiFi Arduino or Node MCU8266



Fig 3. Different type of digital sensor modules

Table 1. Characteristics of different wireless technologies.[1]

	ZigBee	Bluetooth	Wibree	WiFi
Frequency Band	2.4 GHz	2.4 GHz	2.4 GHz	2.4 GHz
Protocol	DSS,CSM A/CA	FHSS	FHSS	DSS/CCK, OFDM
Data Rate	250 kbps	1 Mbps	1 Mbps	11-54 Mbps
Range	30 m-1.6 km	30-300 ft	10 ft	100-150 ft
Power Consumption	Low	Medium	Low	High
Cost	Low	Low	Low	High

With the popularity gained and requirement in communication ability, manufacturers have been integrating WiFi communication module into microcontroller unit for convenience of application such as WiFi Arduino and NodeMCU8266 which is Arduino compatible (Fig 4). This makes implementation of wireless sensor node simple and easily done by non-engineering background developers.

Microcontroller unit and communication module could be expensive and technically used by engineers. With open platform microcontrollers such as Arduino, Raspberry Pi, or BeagleBone Black, that are designed for simplicity of use, make them gain popularity and variety of support in term of interfacing library and open source programming. Wireless communication technology that is commonly used in the agricultural field is ZigBee, Bluetooth, Wibree and WiFi. Each of them has characteristic as in Table 1.

ZigBee technology is based on network and application layer protocols regarding the IEEE 802.15.4 standard. Being low power radio-enabled devices, ZigBee is widely used in the agricultural field due to its characteristics in terms of low power consumption and cost, however; it yields a low data rate of only 20-40 kbps at 868/915 MHz and 250 kbps at 2.4 GHz. Therefore, ZigBee is suitable for low data rate or periodic update of data is required.

Bluetooth is a low cost, low power wireless technology based on IEEE 802.15.1 standard used for short range (8 – 10 m) communication with data rate range from 1 to 24 Mbps. With its ubiquitous characteristics, it's suitable for use in multi-node agricultural application.

Currently, WiFi the IEEE 802.11 standard is a standard for information exchange for internet connection. It is the most widely used in most devices ranging from smart sensors, mobile phones to laptops. It provides communication range from 20 m to 100 m with data transmission rate 2 – 54 Mbps at 2.4 GHz. In wireless applications, WiFi broadens the use of heterogeneous smart devices connecting through the internet (Internet of Things). In Thailand, with the national policy to expedite development of communication infrastructure and make it accessible to Thai population in the rural area, WiFi infrastructure availability becomes nationwide and its cost of service reducing rapidly. This makes WiFi technology gaining more potential and popularity in the agricultural application as well.



Fig 4. WiFi Arduino and NodeMCU8266 board

**B. Internet of Things Platform**

With the growing of IoT technology, both proprietary and open platforms for IoT have been developed rapidly. Some example of IoT platforms are H.A.T, Node-RED, SmartThings, Thingworx and Thingspeak. The platform should be selected based on its characteristics that fit appropriately to the application. One of the internationally well-known open platforms is ThingSpeak that let you send data to collect and store in the cloud storage and display your data as an IoT application. The ThingSpeak IoT platform also provides the users with the ability to analyze and display in a mathematical analysis tool such as MATLAB as well. Data can be sent to this open platform from WiFi Arduino, NodeMCU8266, and other hardware [11].

A similar platform is developed by NECTEC, Thailand is called NETPIE (Fig 5.). NETPIE platform is a cloud-based platform-as-a-service (PaaS) that facilitates interconnection of IoT devices in a most seamless and transparent manner possible by pushing the complexity of connecting IoT devices from the hands of application developers to the cloud. NETPIE is a cloud platform for IoT solution development that lets developers connect devices

seamlessly without tedious work, like infrastructure administration, so developers can spend time on their IoT applications. Devices can communicate whenever they want, wherever they are as if they have direct end-to-end connections with one another. With this open and free of charged infrastructure, with the intention of NECTEC, Thailand to promote the application of IoT, the implementation of precision agricultural monitoring system seem easily feasible by using this open platform [12].



Fig 5. NETPIE IoT Open Platform by NECTEC, Thailand

C. Communication Protocol

There are variety of protocols used for interconnection of devices in IoT platforms. However, protocols that are popularly used recently are lightweight protocols like REST and MQTT. NETPIE provide both REST and MQTT, even though it is not managed MQTT brokers. It has its own publish-subscribe communication model that allows instant messaging among IoT devices. This publish-subscribe communication model can support both the MQTT protocol and the HTTP REST protocol. To facilitate developers and make devices communicate over the MQTT protocol, NETPIE libraries are developed and make available for general developers. The libraries is based on open source Microgear’s libraries firmware or SDKs that provides communication channels and other functionalities between devices and the NETPIE platform. Microgear’s libraries are currently available for most popular hardware and operating systems[12]. With these provided standard protocols and open data source as well as widget plugin, this makes it simple for developers to implement their monitoring system interface.



Fig 6. Example of NETPIE Monitoring Dashboard (<http://naringroup.blogspot.com>)

IV. CONCLUSION

Precision agriculture is being strongly promoted in the agricultural countries especially low and middle income ones to enhance the quality of products and production yield as well as to reduce the cost of farming processes. Wide area sensor network and IoT platform are major foundation and infrastructure for precision agriculture monitoring and control system. With advancement in research and development on Internet of Thing Open Data Platform as well open platform microcontrollers such as Arduino, Raspberry Pi, or BeagleBone Black, that are designed for flexibility and simplicity of use along with the reduction of cost and size of sensors, developers can design and implement precision agriculture monitoring system without much difficulty and less expensive. In the case of Thailand, sensor nodes can be implemented using available and low-cost agriculture-related sensors and open platform WiFi based microcontroller units such as Wifi Arduino or NodeMCU8266. NETPIE IoT open platform by NECTEC, Thailand can be used as interconnection middleware between sensor nodes and farmers. This system architecture realization model can be utilized in many low and middle-income agricultural countries.

ACKNOWLEDGMENT

I would like to acknowledge the kindly effort of Prof. Ramjee Prasad , President, CTIF Global Capsule, for advising this work during his visit at Mae Fah Luang University, Thailand in October 2016.

REFERENCES

- [1] Aqeel-ur-Rehman, Abu Zafar Abbasi b, Noman Islam b, Zubair Ahmed Shaikh “A review of wireless sensors and networks' applications in agriculture” Computer Standards & Interfaces 36 (2014) 263–270
- [2] Julien Mineraud , Oleksiy Mazhelis, Xiang Su, Sasu Tarkoma “ A gap analysis of Internet-of-Things platforms” Computer Communications 89–90 (2016) 5–16
- [3] Rupali B. Mahale, S. S. Sonavane “Smart Poultry Farm Monitoring Using IOT and Wireless Sensor Networks” International Journal of Advanced Research in Computer Science Volume 7, No. 3, May-June 2016

- [4] Shubhangi Bhosale, Sonavane S, "Controlling and Environmental Monitoring of Polyhouse Farm Through Internet" International Journal of Advance Research in Computer Science, Volume 7, May-June 2016
- [5] Issa Kamar, Parag Chatterjee " Internet of Things in Learning Systems – A Perspective Platforms" International Journal of Advanced Research in Computer Science Volume 7, No. 2, Mach-April 2016
- [6] Mohanraj I, Kirthika Ashokumar, Naren J, "Field Monitoring and Automation using IOT in Agriculture Domain" 6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 ,September 2016, Procedia Computer Science 93 ( 2016 ) 931 – 939
- [7] Farag Sallabi, Moustafa Fadel, Ahmed Hussein, Ahmad Jaffar, Hazem El Khatib "Design and implementation of an electronic mobile poultry production documentation system" Computers and Electronics in Agriculture 76 (2011) 28–37
- [8] Lihua Zheng, Minzan Li," Development of a smart mobile farming service system" Mathematical and Computer Modelling 54 (2011) 1194–1203
- [9] Gianluca Barbon, Michael Margolis, Filippo Palumbo, Franco Raimondi, Nick Weldin "Taking Arduino to the Internet of Things: The ASIP programming model" Computer Communications 89–90 (2016) 128–140
- [10] Ray, P.P., A Survey on Internet of Things Architectures, *Journal of King Saud University - Computer and Information Sciences* (2016), doi: <http://dx.doi.org/10.1016/j.jksuci.2016.10.003>
- [11] ThingSpeak ; Internet of Thing , <https://thingspeak.com/>, Accessed ; 2016-10-10
- [12] NETPIE ; Network Platform for Internet of Everything ; <https://netpie.io/>, Accessed : 2016-10-10

# *Role of Mobile Technology Evolution in Accelerating Digital Transformation of Traditional Businesses and Deriving a Score Card Model That can Estimate the Growth Factor*

Jeevarathinam Ravikumar  
Digital Innovation Leader  
Stanley Security Europe (Stanley Black and Decker)  
Slough, UK  
Jeevarathinam.Ravikumar@sbdinc.com

Harshini Arumugam  
IB Second Year Student  
ACS Egham International School, UK  
harumugam@stu.acs-schools.com

Ramjee Prasad  
Center for TeleInfrastruktur (CTIF)  
Aalborg University  
Aalborg, Denmark  
Prasad@es.au.dk

Digital technologies have been on the leap forward growth for the past two decades. Notably, mobile technology evolution which is the nucleus of this revolutionary growth is pushing forward the use of digital technologies for industrial evolution that is impacting the dynamics of global economic growth. Rapid changes in the digital technology evolution is impacting the traditional businesses that are entering into the digital tsunami. Digital transformation has become a turning point for all traditional businesses and it is not an option but it has become mandatory for the business has to reach the next level of growth. This paper discusses the impact of digital transformation over the businesses and is also proposing a simple score card model that can estimate the growth impact correlating to the digital transformation.

*Keywords—digital transformation; digital score card;traditional business;digital growth coefficient*

## I. INTRODUCTION

Mobile technologies especially digital cellular networks have evolved from second generation (2G) to fourth generation (4G) just in the span of two decades. This evolution has brought a life changing experience in the human history in using the technology from home automation to industrial automation.

This evolution has in fact empowered the humans with mobile internet, cloud transformation, bandwidth transformation, virtual & augmented reality, Big data analytics. This is not only enriching the day to day life experience but also enabling non-humans in adopting the technologies to provide the

insights of machines & things leading to communications among machines and humans.

## II. TECHNOLOGY TRANSFORMATION, A BUSINESS DISRUPTOR

This paper is focusing on how the technological advancement among inter and intra communication of human and things is leading to business transformation in traditional businesses. Every board room meeting in the corporate is filled with discussion and debate over the use of technologies not just for functional transformation but for business transformation. Traditional businesses are pressurised with the digital economy impact that is affecting the way businesses are done in the past. There is a paradigm shift in the way of doing business as the infrastructure is becoming commodity and in contrast applications that are running on top of these infrastructure businesses are becoming the only business growth factor.[1,2,3] For example with the revolution of mobile internet, OTT (over the top) businesses such as Google, Facebook, Apple have taken control of mobile infrastructures. On the other hand, industrial OTTs such as Uber, Airbnb have been taking the control of industrial infrastructure businesses. It is in fact, a business awakening for traditional businesses to safeguard the control over their infrastructure towards digital way of doing business. This paper is focused on two major topics, firstly starting with the opportunities that are offered by digital technologies that can catalyse the business growth of traditional businesses. Secondly discussing a Digital score card

model that can be used to estimate the growth contribution.

### III. OPPORTUNITIES THAT ARE OFFERED BY DIGITAL TECHNOLOGIES

Digital Technology is transforming the way businesses are done today. For example, transformation in finance technologies is rapid especially mobile is completely replacing the retail banking. Mobile banking is becoming the norm by the way people do banking transactions using their smart phones. M-PESA from Vodafone has revolutionised the mobile money by offering banking to the unbanked in Africa in terms of mobile money transactions. Mobile banking and Mobile money transactions are now getting inroads into the developed countries. Singapore’s monetary authority is in the process of regulating and introducing the money transfer from mobile to mobile using the mobile number rather than the account number [4].

Bitcoin is the revolutionary digital currency that offers the highest level of secured peer to peer transfer. Block Chain computing behind Bitcoin is becoming de facto standard in the banking industry. Many technology companies are promoting block chain as a standard for most secured digital interactions among human, non-human, living, non-

living things.

Edge Computing, Fog Computing and Block Chain Computing are pushing forward to become the next generation internet and cloud computing.

Fig.1 depicts how the computing changing from centralised to decentralised with more power at the edges and offering direct peer to peer digital interactions without any mediatory. Fig.2 depicts how the Mobile bandwidth evolution is changing the world of user experience [5]

Retail businesses are changing from physical stores to virtual stores with the experience of shop anytime, anywhere and anyway. Customer experience is taking the central stage of any applications.

Augmented Reality (AR) and Virtual Reality (VR) will change the user experience and the consumer industry. Pokémon Go on the mobile has proved augmented reality a major game changing and economic growth factor in the consumer industry. VR and AR are not only playing role in the gaming industry but taking the retail experience to the new heights.

Healthcare is going through the major technology adaption from connected patients to robotic surgery to senses induced prosthetic revolution.

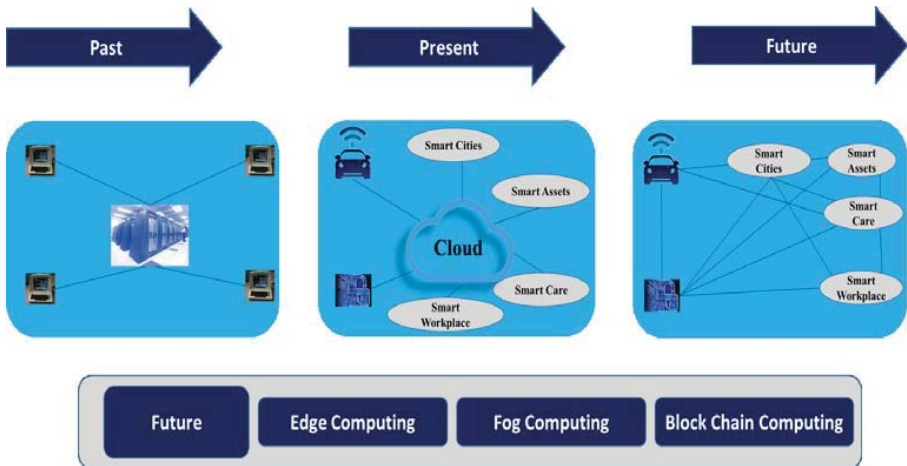


Fig. 1. Mobile Bandwidth Evolution

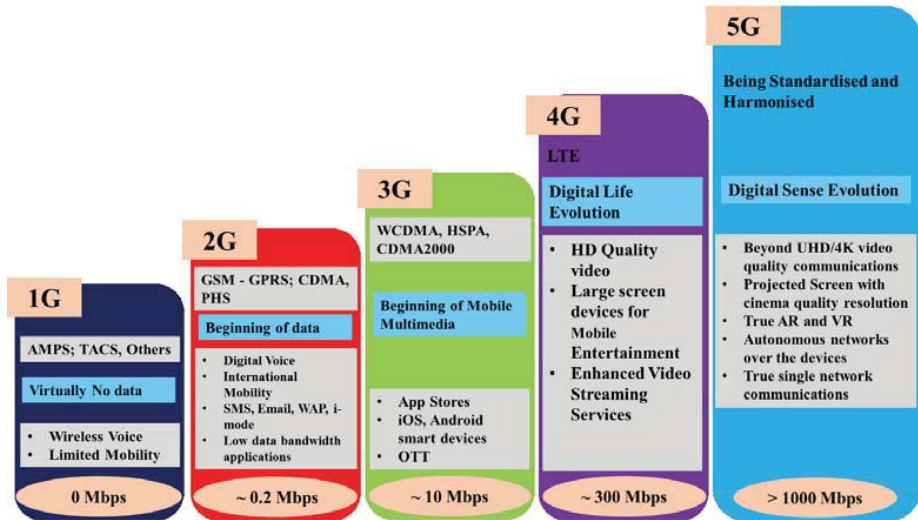


Fig. 2. Computing Evolution: Centralised of the past to most secured and empowered distributed computing of the future

#### IV. HOW DOES THE DIGITAL TRANSFORMATION AFFECT THE TRADITIONAL BUSINESSES?

Apple made a major revolution in mobile user experience by making the mobile internet, a one touch use application by introducing the worlds' first app store. Today, we have over six million apps in the app store, majority are in Google and Apple store.

“The people who are crazy enough to think they can change the world are the ones who do” - Steve Jobs, Ex CEO, Apple Inc.

Another example of the unique digital visionary is the CEO of Tesla Motor; Elon Musk, who had a revolutionary vision for producing eco-sustainable cars.

Businesses that don't believe in digital disruption or stop innovating will become extinct by staying with the complacency of self-perceived monopoly.

#### V. DEFINING DIGITAL TRANSFORMATION

Digital Transformation has become a business mantra in any organisation. Digital transformation is the new normal that is shaking every business to think differently to sustain and grow. CEOs of almost all these organisations will agree that Digital Investments is the way to a successful growth. All of them have their own way of defining and

implementing it. Going digital has become a buzz word but what does it mean to the business? It is a kind of a puzzle in most of the traditional businesses. Lot of these organisations have also got into a dilemma of thinking into a belief that IT transformation is the technology or digital transformation. For clarity, IT is only a function of many functions of an organisation. Digital transformation of industries has become a global phenomenon [6].

In this paper, we start with defining the digital transformation.

*“Digital transformation is defined as the convergence of the two, firstly business transformation through technological innovation, secondly functional transformation through technological automation resulting in unprecedented business growth”.*

First aspect of digital transformation is predominantly an innovation-led growth across the organisation clearly empowered by the business leaders. The second aspect is automating the business functions across the organisation through digital transformation. Thus the digital innovation lead growth should be supported by digital automated functional transformation.

#### VI. DIGITAL SCORE CARD MODEL

In this paper, authors derived a score card model to estimate the impact of digital transformation in a traditional business. This paper has considered

different attributes of various functions of an organisation that contribute to the growth of the business. Objective of this model is to estimate how these attributes of various functions arising out of functional transformation directly contributing to the business growth.

In this paper, simplified attributes are taken for consideration that are derived out of different study and survey. It gives a much generalised model.

This study has considered five major functions of the organisation that can contribute to the overall business growth.

Five functions that are considered:

1. Leadership
2. People
3. Technology
4. Sales & Marketing
5. Support (Finance, Admin, Logistics)

Digital Growth Coefficient ( $D_g$ ) is derived out of these functional organisations by analysing the impact of digital transformations.

A weightage contribution is allocated to each and every function.

In this study, every function under consideration contributes a fractional value which is capped to get the overall value of  $D_g$  equals 1. In this paper, support function has not taken into consideration as it will be included in the comprehensive study phase.

Let us first define the Achievable Growth  $A_g$  as

$$A_g = T_g * D_g \quad (1)$$

$A_g$      *Achievable Growth*  
 $T_g$      *Target Growth*  
 $D_g$      *Digital growth coefficient*

$$D_g = C_L + C_P + C_s + C_T \leq 1.0 \quad (2)$$

Target growth is set by the business which is trying to achieve a particular level of growth in a given duration.

Achievable growth is determined by the digital growth coefficient multiplied by the target growth. At times,  $A_g$  may exceed  $T_g$  based on the maturity stage and certain economic and external conditions. In this paper, we considered  $A_g = T_g$  with the full maturity by achieving the maximum of  $D_g$  equals 1.0

The following sections define the different functional coefficient. Every functional coefficient is assigned with top five attributes that has a

maximum score of 10 each and a weightage is assigned to measure the contribution. Few iterations are made in allocating weightage to different attributes. Attribute scores,  $s_1$  to  $s_n$  are derived from the various surveys and the performance trends. Weightage,  $r_1 \dots r_n$  are assigned with various iterations determined by direct correlation of the attributes to the revenue growth and mean average of the weightage equals 1.0

$$\frac{1}{n} \sum_{k=0}^n r_k = 1.0 \quad (3)$$

All the functional coefficients are capped to a predetermined value thus the sum of the coefficients don't exceed 1.0 at all times.

#### A. Leadership Coefficient ( $C_L$ )

All research shows that leadership of an organisation plays a major role in digital transformation.

$C_L$  is the leadership coefficient and is derived from the related attribute score  $s_k$  and related weightage  $r_k$ .

TABLE I.

No.	Attributes	Score	Weightage
1	Digital Empowerment	s1	r1
2	Digital Visionary	s2	r2
3	Digital Strategy & Roadmap	s3	r3
4	Digital Landscape Understanding	s4	r4
5	Digital Capital Allocation	s5	r5

$$C_L = \frac{1}{n} \sum_{k=1}^n s_k * r_k \leq 0.2 \quad (4)$$

#### B. People Coefficient ( $C_P$ )

People play most important part in any organization as the digital mindset is an important attribute that has a direct correlation on where the organization is leading to.

$C_P$  is the people coefficient and is derived from the related attribute score  $s_k$  and related weightage  $r_k$ .

TABLE II.

No.	Attributes	Score	Weightage
1	Digital Recruitment Strategy	s1	r1
2	Digital Work Environment	s2	r2
3	Digital Transformation Training	s3	r3
4	Digital Message Across Organisation	s4	r4
5	Digital Innovation Culture	s5	r5

$$C_P = \frac{1}{n} \sum_{k=1}^n s_k * r_k \leq 0.2 \quad (5)$$



C. Sales & Marketing (Cs)

Sales and Marketing are front runner of any organisation that has direct contribution to the revenue growth.

Cs is the people coefficient and is derived from the related attribute score sk and related weightage rk.

TABLE III.

No.	Attributes	Score	Weightage
1	Digital Sales Strategy	s1	r1
2	Digital Sales Resources	s2	r2
3	Digital Sales and Marketing Tools	s3	r3
4	Digital Customer Experience Strategy	s4	r4
5	Innovative Pricing Model	s5	r5

$$Cs = \frac{1}{n} \sum_{k=1}^n sk * rk <= 0.25 \quad (6)$$

D. Technology (Ct)

Technology plays the most important role as it directly relates to the digital transformation growth as an organisation is making business transformation through technology solutions offerings.

TABLE IV.

No.	Attributes	Score	Weightage
1	Digital Innovation Strategy & Roadmap	s1	r1
2	Customer Focused Digital Innovations	s2	r2
3	Digital delivery team alignment	s3	r3
4	Digital Ecosystem Partners	s4	r4
5	Patents and IPs	s5	r5

$$Cr = \frac{1}{n} \sum_{k=1}^n sk * rk <= 0.35 \quad (7)$$

VII. DIFFERENT STAGES OF DIGITAL TRANSFORMATION GROWTH ANALYSIS

A trend analysis is run using a simplified score card model. In this analysis, we have considered an organisation that is investing heavily into their digital transformation initiative for creating more digital innovative platforms. This organisation is expecting to disrupt the market through a major business growth in three years by the new initiative. To help the new digital technology disruption, the organisation is also investing in all the functional organisations to help accelerate the revenue growth. This analysis did run five stages of digital growth maturity evolution

Fig.3 shows the trends at different stages of evolution. It is clearly evident that the different coefficients that are directly affecting Dg hence the business growth. Each and every single attribute is directly correlating to the digital growth maturity.

TABLE V.

Current Revenue in Million USD	Target Revenue in Million USD	Years
1000	1500	3

TABLE VI.

Target CAGR	~14%
Targeted Growth in Million USD/Year	500

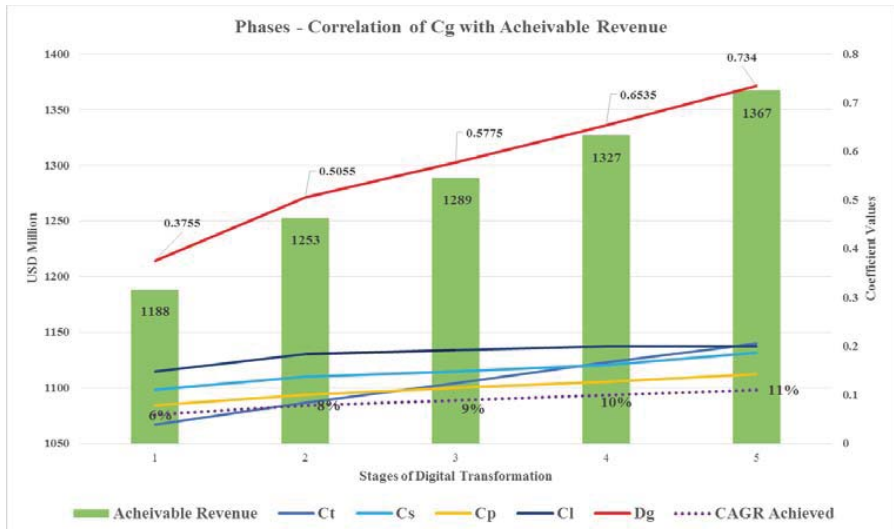


Fig. 3. Five stages of estimation from stage 1 to stage 5 showing the digital transformation growth maturity

### VIII. AUTOMATION USING MOBILE APPLICATION

All the survey and iterations can be fully automated with a mobile application and backend analytics. This is being proposed as of now. But for this study, a mobile wireframe application is being developed to realise the concept of automation.

### IX. CONCLUSIONS

Evolution of mobile technology is completely changing the way businesses are done. It is imperative for all traditional businesses that are embracing digital transformation shall sustain and grow. The score card model proposed here can be further expanded to include the technology evolution and its applications as a major contributor to digital coefficient.

### X. REFERENCES

- [1] Cap Gemini, "The digital revolution and its impact on leadership", 2013
- [2] Cap Gemini, MIT Sloan Management, "DIGITAL TRANSFORMATION: A ROADMAP FOR BILLION-DOLLAR ORGANIZATIONS", 2011
- [3] Deloitte Digital, "Building your digital DNA Digital transformation in progress", 2015
- [4] Monetary Authority of Singapore (MAS) Press Release on Fund Transfers via Mobile Phone Number, 19 Aug 2016.
- [5] Jeevarathinam Ravikumar, Ramjee Prasad, "Adding a new dimension to customer experience, the reality of 6<sup>th</sup> sense - 5G and beyond", 5G outlook – Innovations and Applications, 2016, River Publishers.
- [6] World Economic Forum, Accenture, "Digital Transformation of Industries", Digital Enterprise, January 2016.

### XI. BIOGRAPHIES



**Jeevarathinam Ravikumar** is a technopreneur with 25 years of experience in Mobile technology R&D, heading technology businesses and mentoring early stage start-ups. He was the CTO of Marconi Wireless companies in Asia, heading 3G research at Centre for Wireless Communications (Institute of Infocomm Research, I2R under NUS, Singapore) and Head of Regional Business for various multinationals. He is a senior member of IEEE. He works closely with Professor Ramjee Prasad on next generation technologies.



**Harshini Arumugam**, is in 2<sup>nd</sup> year International Baccalaureate at ACS Egham International School, London. She has immense interest in the subject of digital transformation and exploring more into how artificial intelligence, Big data and robots can aid businesses to digitally transform. She has achieved the Gold CREST award for design and construction of Robot for Eurobot competition by British Science Association. She has high ambition to be a successful entrepreneur in business consulting and empowering women in digital startups.



**Professor Ramjee Prasad** is Director at CTIF, Aalborg. He is one of the most renowned professors in wireless telecommunications research and is currently leading 5G standards for developing countries.

# Distributed Dynamic Backhauling in Self-Itinerant Intelligent Aerial Radio Architecture

Purnima Lala Mehta  
Center for TeleInfrastruktur (CTIF),  
Aalborg University,  
Aalborg, Denmark,  
pla@es.aau.dk

Prof.(Dr.) Troels B. Sørensen  
Department of Electronics Systems,  
Aalborg University,  
Aalborg, Denmark,  
tbs@es.aau.dk

Prof. (Dr.) Ramjee Prasad  
Center for TeleInfrastruktur (CTIF),  
Aalborg University,  
Aalborg, Denmark,  
prasad@es.aau.dk

**Abstract**— Our previous works, subsequently produced the following findings. Firstly, we identified that delivering cellular services to the dynamic hotspots, formed due to random accumulations and movement of ground users as a major challenge to the future generations of mobile communications, as a problem defined as Place Time Capacity (PTC). Our second work proposed utilizing small Unmanned Aerial Vehicles (UAVs) following the user groups, to offload the serving Base Stations (BSs) affected by the PTC. We introduced this concept as the Hovering Ad-hoc Network (HANET). Our recent work extended this concept to give a holistic view to an aerial radio architecture, defining this architecture as a Self-Itinerant Aerial Radio Architecture (SIARA), and investigated the radio access performance of the SIARA. In this paper, we aim to propose and investigate the distributed and dynamic backhaul feature of the SIARA along with an additional intelligence characteristic of the aerial Members. As the nature of investigation is focused on serving dynamically moving hotspots, we propose and evaluate the distributed dynamic backhauling feature through ranking based procedures, at two different user accumulation scenarios and further compare the static and dynamic backhauling for such a deployment. We also perform optimizations on the count and placement of the aerial Members.

**Keywords**— *Place Time Capacity; Unmanned Aerial vehicles; Hovering Ad-Hoc Network; Self-Itinerant Intelligent Aerial Radio Architecture; Distributed Dynamic Backhauling.*

## I. INTRODUCTION

The perpetual growth in wireless data services and subscriptions are encouraging researchers and network operators to shift from conventional and traditional infrastructures and technologies to naive and innovative implementations for the future 5G+ generations (beyond IMT 2020). Data hungry and high bandwidth mobile services like audio and video streaming, cloud based services, and data sharing through social networks, etc., are overloading the existing infrastructure, especially at the hotspots situations and is foreseen to escalate in the coming future. It becomes even more challenging for the service providers to deliver the services to moving user groups characterized with temporary, erratic, and high data rate demands giving rise to phenomenon like Place Time Capacity (PTC) [1].

We have proposed and evaluated using low altitude UAVs to serve the moving group users with an architecture namely, Self-Itinerant Aerial Radio Architecture (SIARA) [6], incorporating multiple Hovering Ad-Hoc Network (HANET)

systems [4][5] and assisting the terrestrial network in offloading the traffic from the Macro base stations in a Heterogeneous (HetNet) based network environment. Our investigations in [5] and [6], focused mainly on the radio access part of the SIARA architecture. However, we believe that one of the limitations of such kind of deployment could be, that the aerial cellular networks do not possess their own independent and dedicated backhaul system, and thus, the downlink serving capacity of the hovering aerial Members might be limited by the existing dedicated backhaul link capacities of the ground base stations. Nonetheless, data transfer is a crucial aspect of the wireless networks, and in our work, under high capacity demand and motion dynamics, we propose the aerial members to be incorporated with self-organizing, adaptive, and intelligent features to manage and maintain the network backhaul through distribution of relaying links among the nearer available gateways for backhauling purposes. This intelligent feature in the individual HANET Members, several HANET sub-systems working together to solve PTC congestions in the entire network region is proposed in this paper as the Self-Itinerant Intelligent Aerial Radio Architecture (SIARA).

Designing a high-performance and efficient backhaul network to carry user data traffic between the aerial networks like HANET systems, and the core network, is a *green field*. Due to the mobility of the ground users in groups, flexible and alternate backhauling schemes needs to be devised to backhaul the offloaded data by the aerial members in an effective manner to the core network. We have previously introduced the concepts of dynamic and distributed backhauling in [6] and in this paper, we will evaluate these concepts in a HetNet based environment to perform the distributed and dynamic backhaul management for different network scenarios and perform optimizations in terms of number of aerial Members needed (for its different roles), and their optimal placements under user motion dynamics.

The rest of the paper is organized as follows: Section II provides a brief overview of the related work. Section III explains the system model including the Dynamic Backhauling concept and simulation assumptions. Then we will show some simulation results in Section IV. Finally, we provide the concluding remarks in Section V.

## II. RELATED WORK

The backhauling aspect of the aerial cellular networks is a state-of-the-art research field and has not been explored much

in the literature. In context to cellular backhauling, Free Space Optics (FSO) [7], microwave, and millimeter wave [8] have been proposed for dense cellular networks comprising of small cells, as regular backhauling solutions, falling under the category of wireless backhaul. Our work concentrates on small cells mounted on UAVs and their usage in cellular backhaul network.

The ABSOLUTE project suggests, Satellite and WiFi as candidates for wireless backhaul [9] [10]. They believe that Satellite backhauling (Ka band) brings the advantage of unlimited coverage offering the possibility of connecting the aerial network for any distance but bring along the drawback of introducing delay by the satellite links, that may affect some real-time services such as voice and real-time video. They have further suggested that alternate solution of WiFi links can be used, but, at the cost of reduced coverage and capacity. In one of their other works, they have introduced a novel cluster-head selection approach, in clustering wireless nodes on ground fields, through the assistance of an aerial base station to form an aerial backhauled network [11]. They have further investigated the possibility of using a WiFi-based link between two Low Altitude Platforms (LAPs) for backhaul and, modelled the Hand Over process over the WiFi backhaul [12]. In [13], authors have used UAVs as relays to improve the reliability of wireless backhaul networks consisting of balloons launched in the stratosphere to carry wireless transceivers owing to the problem of the wireless links (e.g. FSO or MMW links) that can cause the link instability.

### III. SYSTEM MODEL

#### A. SIIARA-HANET Concept

Events like carnivals, parades, festivals, etc. are packed with heavy user accumulations with random and uncertain mobility in groups. We define places of potential accumulations in the network as an Area of Event (AoE). To serve these mass users, we aim to deploy small, low power, and close-to-ground hovering aerial Members. These Members work in teams, coordinate, and cooperate with each other to form multiple ad-hoc networks, and follow the user clusters with the goal of providing coverage and high data rate cellular services. It is a task for the aerial hovering Members, to offload the additional load from the ground macro base stations and assist the network in maintaining the required QoS to the ground users while in motion. In [4][5], we have presented a concept of HANET as a hovering ad-hoc team to serve PTC congestions and further proposed SIIARA as a holistic and independent network architecture comprising multiple HANETs in [6].

The Members are required to communicate with the ground users, ground base stations, and among themselves to carry out the control and signaling for the user data to be served and backhauled to the backbone network effectively. For carrying out the necessary network functions, the Members are required to play multiple roles namely, (as described and discussed in [5]):

1. HANET Serving Member (HSM): Engaged for delivering the cellular services to the ground users.
2. HANET Relay Members (HRM): Participate in relay chain to relay the user data from different HANET sub systems in the network to the nearest ground base

station, named as the HANET Gateway Base Station (HGBS) (in our work), from where the user data is further backhauled.

3. Gateway Aerial Vehicle of HANET (GAVH): Transfer of user data between different HANET networks.

Further, the communication links except the access radio links operate in 60 GHz Millimeter Wave Communication (MMWC).

#### B. Dynamic Backhauling in SIIARA: The Concept

SIIARA as a network architecture, is proposed to manage the calls of SIIARA-bound users without any external assistance. However, the SIIARA to terrestrial network calls, or vice-versa, can only be managed if there is suitable backhaul links available between the two architectures. By suitable we mean that the backhaul links must be resilient, stable, and high capacity links to enable bulk data transmission owing to PTC centric conditions. To enable the inter-network data transfer, the aerial Members are required to communicate with the ground base stations, that are the HGBSs, to route the data to the backbone network.

As we are dealing with the dynamic users, the backhaul links in the SIIARA is an iterative continuum process of releasing the old links and forming new links, so that the data exchange remain uninterrupted throughout the operation. We define here this term of iterative backhauling process as the 'Distributed Dynamic Backhauling for the Aerial Architecture' (DDBAA). The aerial HANET Members are required to determine all the possible and nearest under-utilized HGBSs, that have capacity to backhaul extra user data. Due to the fixed placement of the HGBS and mobile nature of the hovering aerial Members, it is utmost important for the HRMs to dynamically transfer the link from the previous HGBS to the present closest HGBS. Fig 1 demonstrates the DDBAA concept with a scenario where the ground users move from BS1 coverage area to BS2 coverage area. Initially, the BS1 network site is overloaded due to PTC congestion and the sites BS2 and BS3 are the underutilized, having capacity to route the user data through their backhauling channels. The aerial HANET Members are deployed above the PTC users, three HSMs service the users and one HRM, relays the data to the underutilized BS2 and BS3, making them the HGBSs. As soon as the PTC causing users reach the coverage area of BS2, the site becomes overloaded and the sites BS1 and BS3 becomes underutilized. The Members in the HANET change their links and roles, and position themselves as per the present condition to be able to have relay connection with the underutilized BS1 and BS3. The process of choosing more than one HGBS is the dynamic feature of the HANET and to distribute the user data between the chosen HGBSs, depicts the distributed feature of the HANET. The heavy user data can be backhauled through more than one HGBS, preventing the overloading of a single HGBS backhaul pipe. It is to be noted that the user data will be relayed through the 60 GHz MMWC backhaul links and calculations have been performed accordingly.

#### C. Dynamic Backhauling: The Working Procedure

The dynamic backhauling in SIIARA is achieved through the following generic procedure:

- 1) *Identifying the fulcrum: This is the operating network scenario in which the DBAA process is performed and includes user distribution, type of radio environment,*

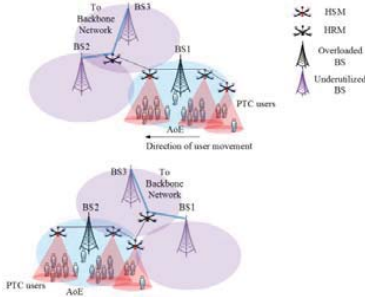


Fig. 1. Distributed Dynamic Backhauling Concept

throughput demand per user etc.

- 2) *Identifying the PTC:* This includes knowing the number of PTC users concentration and their mobility behaviour.
- 3) *Ranking of available resources:* Based on outcomes of (a) and (b), the available under-utilized HGBSSs are searched for. The identified HGBSSs are then ranked based upon (i) their accessibility, (ii) the capacity available in the backhaul channels, and (iii) resilience of the backhaul links with the user dynamics.

The additional users creating PTC are collectively serviced by the macros falling in the AoE, the neighboring sites and the aerial HANET Members. We perform optimizations to find (i) minimum number of HSMs, HRMs and HGBSSs needed and, (ii) optimal placement of all the aerial Members. Firstly, we target to optimize the number of HSMs needed to serve within the AoE, by identifying only those part of areas, which are not covered by the macro sites. This is done by finding the areas covered by the macros based on the thresholds on (a) SINR, and (b) total number of users serviced by the neighboring macro sectors. The chunk of areas that are essentially covered by the macros are subtracted from the AoE region required to be covered by the HSMs. This optimization ensures that minimum number of HSMs are needed to serve only the weakest areas of the AoE region.

In our simulation, for evaluating the DDBAA process, we follow different ranked based procedures. Firstly, we examine the nearby macros, suitable to act as the HGBSSs, by estimating the rank of every macro BS based on the distance of each HANET Member-macro BS pair (rankX- Distance Based Ranking). Secondly, the macros that were chosen based on their distance ranks, are further ranked based on the utilization of their individual backhaul links capacities (rankY- Capacity Based Ranking). We, then calculate the Aggregated Link Capacity (ALC), which is the aggregation of the backhauling capacities of a macro site in consideration with its first tier of neighbors for the given hexagonal deployment. For better resource management, we divide the aggregation in smaller groups, here we consider groups of three including the site in consideration. The next rank procedure involves the ranking of the groups, sorted in descending manner, and the group with

highest ALC is ranked as one and so on. The initial HRMs are strategically placed at the centroid of the HGBSS groups to be able to communicate with all the sites in the group keeping in view the MMWC communication range limitation. The rest of the HRMs will form a relay chain from the strategic HRMs to the HANET team. We find a cost of deployment to identify the best HGBSSs groups for backhauling that eventually benefit in finding the minimum number of HRMs needed to complete the relay chains. We define this cost as the ‘Length of Tolerance (LoT)’, and is given as:

$$LoT = \frac{N_{UHRM} \times (\min(rankX)) \times (\min(rankY))}{(ALC \text{ of the group})}$$

where,  $N_{UHRM}$  are the total number of HANET Relay Members (HRMs) needed, calculated based on the distance between each HSM-HGBSS pair and the maximum distance allowed between the relay members due to MMWC link distance limitation. We target to minimize the LoT and the group with least LoT is chosen for performing the DDBAA and from the remaining groups, the group with least LoT is again selected, and so on till all the PTC users are completely absorbed.

#### D. Simulation Assumptions

We develop a system model for a downlink heterogeneous Urban Macro (UMA), hexagonal cellular network model for evaluating IMT-A as specified in the ITU-R M.2135-1 report to endorse the DDBAA concept of a single HANET system in SIARA at two different placements in the network area. In the simulations, we investigate this process in three different network scenarios (A, B, and C), affected by the PTC user accumulation. Scenario A is when AoE region is created at bottom left of the network region. PTC users while moving create the AoE region at top right in the network region represent the scenario B and scenario C demonstrates scenario B with a building obstruction. The ground users and the PTC users are randomly placed through Poisson’s Point Process (PPP). The AoE region, spans an area of 500x500 m<sup>2</sup> and comprises of 196 PTC users. The simulation assumptions are further summarized in Table I.

## IV. SIMULATION RESULTS

The top projections UMA network scenario with randomly placed ground users (in blue) and PTC users (in orange red) are shown in Fig 3,5 and 7, representing scenarios A, B, and C respectively. Fig 13 describes the legends used. A 3D view of Scenario A is depicted in Fig 2. After following the ranking based procedure described in the previous section, we get the graphs of distance based rankings (example perspective from 4 different HSMs), for all the three scenarios (shown in Fig 4,6 and 8). The seventeen circled macro sites are the HGBSSs that are selected after performing the ranking procedures for the respective scenarios. The pink dots are the HRMs, strategically placed at the centroid of the selected HGBSSs groups, with altitude same as the macro site. The yellow dots are also the HRMs that place themselves to form relay chains between the pink strategic HRMs and selected corner HSMs.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Simulation Area	3000x3000 m <sup>2</sup>
Propagation Environment	Urban Macro
Transmit Power Macro/HANET	49/30 dBm
Simulation Frequency (f)	2 GHz
System Bandwidth LTE/MMWC	20/100 MHz
Number of Macro sites	38
Number of users per site	30
Average DL Spectral Efficiency (SE) BS/HANET	2.2 /3 bps/Hz/cell
Number of Sectors in Macro/HANET	3
Member Altitude	90 m
AoE Region	500x500 m <sup>2</sup>
PTC user distribution at AoE region	PPP ( $\lambda=267$ per km <sup>2</sup> )
Air to Ground Propagation Model	Dense Urban environment [15]
Macro to User Propagation Model	ITU-R Path Loss Model for Urban Macro
Member to Member Propagation Model	Free Space
Maximum Inter Member Distance (MMWC)	283 m
MMWC Backhaul link capacity	220 Mbps

In scenario C, due to a building obstruction, closest HGBSs (circled in pink) to the obstruction are dropped and their load is distributed to the nearby HGBSs (circled in green). Fig 9 shows the ALC group ranking of selected groups for all the three scenarios indicating different ranks owing to the dynamic behavior in the three scenarios. Fig 10 compares the DDBAA process with the static backhauling (BH) case for the user movement iterations. In static BH, only one HGBS was chosen to backhaul the entire user data and due to the PTC users moving from one position to another, more number of HRMs are required to maintain the relay chain with the single selected HGBS. With DDBAA, the number of HRMs required, increases to a certain count until the link with previous HGBSs are not broken and as soon as the new links with the nearby HGBSs is build up, number of HRMs required go down. The number of HRMs required stay with a certain count bracket only in case of DDBAA. We further compare the static BH and DDBAA through a cost of deployment in terms of ratio of HSMs per user (expected to be minimum). Fig 11 demonstrates the comparison, wherein for the static BH case, the ratio HSMs/user go till infinity with increasing user influx for two different HGBSs chosen. This is because a single HGBS can transmit/receive user data upon a certain limit only, so, even if HSMs increase indefinitely, the services are not delivered to the extra users. On the other hand, for the DDBAA, the HSMs/user ratio becomes almost constant (continued in Fig 12) for increasing user distributions. This is because, for effective service delivery, multiple backhauling links to multiple HGBS groups and optimizations, tend to limit the number of HSMs at the AoE for any given user distribution and hence the ratio of HSMs/user stays constant on adding extra users.

V. CONCLUSION

In this paper, we have proposed the distributed dynamic backhauling process of an aerial radio architecture in

millimeter wave communication to effectively backhaul the offloaded data from the macro sites during heavily crowded conditions. We have described multiple roles of the aerial Members and performed some optimization procedures to achieve minimum number of aerial members required for different operations as well their optimal placements in the network region under user motion dynamics. It was observed through the simulations that number of relay aerial Members required to backhaul the data in the distributed dynamic backhauling process is less than the number of Members required in the static backhauling case. We will further formulate artificial intelligence based algorithms to perform the optimizations.

REFERENCES

- [1] Kumar, A.; Mehta, P.L.; Prasad, R., "Place Time Capacity- A novel concept for defining challenges in 5G networks and beyond in India," Wireless Computing and Networking (GCWCN), 2014 IEEE Global Conference on, vol., no., pp.278,282, 22-24 Dec. 2014.
- [2] H. Shimodaira et al., "Optimization of picocell locations and its parameters in heterogeneous networks with hotspots," 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC), Sydney, NSW, 2012, pp. 124-129.
- [3] A. Jaziri, R. Nasri and T. Chahed, "Offloading traffic hotspots using moving small cells," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6.
- [4] Mehta, P.L.; Sorensen, T.B.; Prasad, R., " HANET: Millimeter wave based intelligent radio architecture for serving place time capacity issue," Wireless VITAE, 2015 Global Wireless Summit, Dec. 2015 (In Print).
- [5] Mehta, P.L.; Sorensen, T.B.; Prasad, R., "SINR Based Capacity Performance Analysis of Hovering Ad-Hoc Network," Wireless Personal Multimedia Communication (WPMC), Nov 2016 (Presented).
- [6] Mehta, P.L.; Sorensen, T.B.; Prasad, R., " A Self-Itinerant Aerial Radio Architecture For Serving Place Time Variant User Accumulations," Wireless World Research Forum (WWRF), Oct 2016 (Presented).
- [7] Y. Li, N. Pappas, V. Angelakis, M. Pioro and D. Yuan, "Optimization of Free Space Optical Wireless Network for Cellular Backhauling," in IEEE Journal on Selected Areas in Communications, vol. 33, no. 9.
- [8] U. Siddique, H. Tabassum, E. Hossain and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," in IEEE Wireless Communications, vol. 22, no. 5, pp. 22-31, October 2015.
- [9] S. Chandrasekharan et al., "Designing and implementing future aerial communication networks," in IEEE Communications Magazine, vol. 54, no. 5, pp. 26-34, May 2016.
- [10] K. Gomez et al., "Aerial base stations with opportunistic links for next generation emergency communications," in IEEE Communications Magazine, vol. 54, no. 4, pp. 31-39, April 2016.
- [11] A. Al-Hourani, S. Chandrasekharan, A. Jamalipour, L. Reynaud and S. Kandeepan, "Optimal Cluster Head Spacing for Energy-Efficient Communication in Aerial-Backhauled Networks," 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, 2015, pp. 1-6.
- [12] R. Fedrizzi, L. Goratti, K. Gomez and T. Rasheed, "On the feasibility of handover over WiFi backhaul in LTE-based aerial-terrestrial networks," 2014 IEEE Wireless Communications and Networking Conference (WCNC), Istanbul, 2014, pp. 2196-2201.
- [13] F. Ahdi and S. Subramaniam, "Using unmanned aerial vehicles as relays in wireless balloon networks," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 3795-3800.
- [14] Rep. ITU-R M.2135-1, Guidelines for evaluation of radio interface technologies for IMT-Advanced, Dec. 2009.
- [15] Al-Hourani, S. Kandeepan and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," in IEEE Wireless Communications Letters, vol. 3, no. 6, pp. 569-572, Dec. 2014.

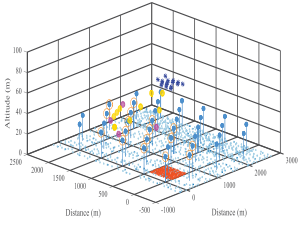


Fig 2. DDBAA- Scenario A in 3D

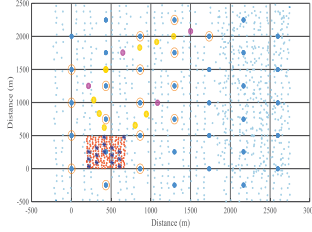


Fig 3. DDBAA- Scenario A in 2D

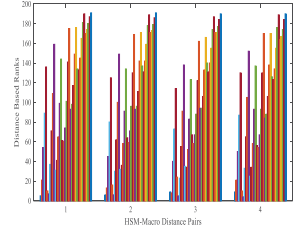


Fig 4. Distance based ranking- Scenario A

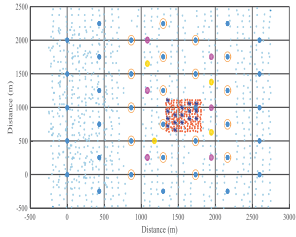


Fig 5. DDBAA- Scenario B

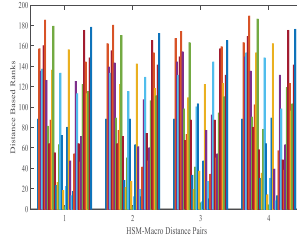


Fig 6. Distance based ranking- Scenario B

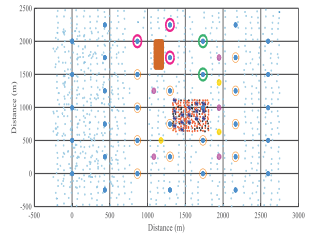


Fig 7. DDBAA- Scenario C

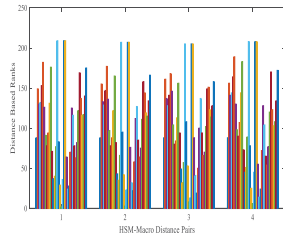


Fig 8. Distance based ranking- Scenario C

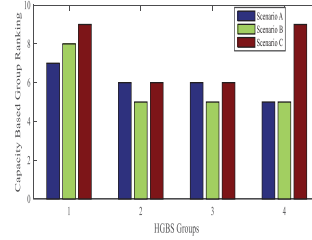


Fig 9. ALC ranking of HGBS Groups-Scenarios A, B, and C

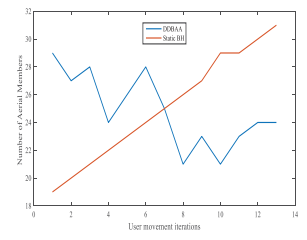


Fig 10. Comparison of Static BH and DDBAA

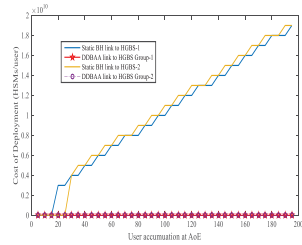


Fig 11. Comparison: Cost of Deployment

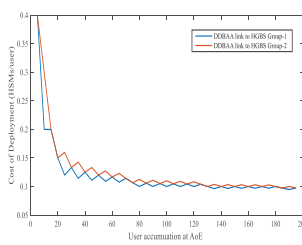


Fig 12. Comparison: Cost of Deployment

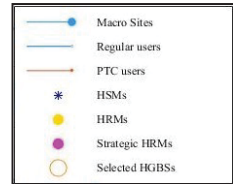


Fig 13. Overall Legends

# Hierarchical Network Design and Optimal Deployment of LTE Enabled RSUs in VANET

Feng Jiayong, Wu Muqing, Zhao Min  
 Beijing Laboratory of Advanced Information Networks,  
 Beijing University of Posts and Telecommunications  
 Beijing, China 100876

**Abstract**—In this paper, we provide an optimal deployment algorithm to design an efficient data transmission network architecture, in order to realize real-time information transmission in VANET. Our first aim of this work is to formulate a three-layer network structure that can provide cost efficient data transmission by the meantime, the bandwidth, latency, and connectivity constraints will be well met. Furthermore, the optimal deployment of the LTE enabled Road Side Units (RSUs) leads us to a second aim which it is the access to the vehicular communications, either directly Vehicle to Vehicle (V2V) or Vehicle to Infrastructure (V2I). In our simulation, the performance of our proposed network and traditional network is compared and then we analyze the effect of each constraint on the network construction cost. Our analysis shows that hybrid three-layer wireless network used in VANETs can solve the problem of real-time data transmission and reduce the overall cost of the network by the improved algorithm proposed in this paper.

**Index Terms**—VANETs; Hierarchical network design; RSUs; Optimal deployment.

## I. INTRODUCTION

The Vehicular Ad-Hoc Networks (VANETs) are an envision of the Intelligent Transportation Systems (ITSs). Vehicles communicate with each other via Vehicle to Vehicle Communication (V2V) or Vehicle to Infrastructure Communication (V2I) to supervise all the vehicle states and provide comprehensive services depending on the different functional requirements. Nowadays, great endeavor has been made by automotive manufacturers, governments, and research universities toward efficient vehicular communications, which significantly improves the development of intelligent transportation systems. In industry, the current generation of vehicles already equips with devices with certain computation and communication capacities, such as MyFord Touch from the Ford Motor Company, Entune from the Toyota Motor Company, etc [1].

The optimal goal of the VANETs is to make safer and more efficient roads in the future by providing real-time information to concerned authorities and vehicles [2]. Thus the overall network design between vehicles and control center is the core and focus of real-time data transmission system. But most of related works [3]–[7] are based on very abstract assumption. The proposed network in [4] established the V2V and V2I2V communication network by WiFi, which provides the high speed links but does not solve the RSUs congestion problem. [5] offered a deployment scheme of RSUs for low latency information transmission from vehicles to infrastructure. But it is highly cost and inflexible because most of the RSUs in

[5] connect to control center by wired link. [7] designed an optimal roadside AP placement for delay-tolerant network. It can guarantee the general quality of service (QoS) requirement but not apply to the emergency information.

Our research proposed a hierarchical network architecture composed of existing optical fiber, LTE network and wireless sensor network in order to ensure timely data transmission. All vehicles in the VANET are equipped with various wireless sensors, and a variety of traffic or vehicles information can be transmitted [8]. All the massive of information should be delivered to the control center by a highly efficient and reliable way [9]. For example, considering the traffic accident scenarios, the emergency information on the road should be transmitted to control center timely. It is obvious that the quality of those information needs to be guaranteed by the communication between vehicles and control center. In this paper, we use the RSUs equipped with LTE transceiver to provide high bandwidth low latency links. Furthermore, we present a cost optimum deployment algorithm of LTE enabled RSUs in consideration with transmission latency and bandwidth constraints.

This paper is organized as follows. In Section II, we analyze the vehicular ad-hoc network structure and propose a hybrid three-layer network architecture. Then in Section III, we establish the three-layer network model and propose the optimal deployment algorithm of LTE enabled RSUs. Section IV presents the network performance analysis and Section V concludes our paper.

## II. NETWORK DESIGN

### A. Network Analysis

The VANETs need a robust wireless sensor network with respect to many constraints such as bandwidth, link unreliability and transmission latency. We present an illustrative example to show the concerned application scenario in Fig.1.

Roadside infrastructure using a big sensor network consists of many vehicle nodes to transmit data to the control center through stations. From Fig.1, we can find RSUs on both sides of the road form a linear network topology, which is considered as a great challenge for real-time data transmission in VANETs. The performance of the linear network model was analyzed in [10]. The result in [10] shows that data from the RSUs far away from the station has to travel a longer distance than that of RSUs near the station. So the successful



data delivery ratio of the RSUs far away from station is much lower, and the possibility of packet collision in RSUs near station is much higher.

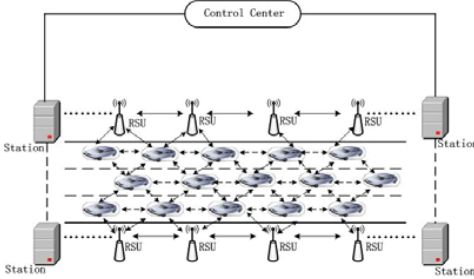


Fig. 1. Traditional Network Structure in VANET

In Fig.1, The RSUs far away from station will use the others which near station to forward data packets. Thus, in this network structure, the RSUs which near station will transfer much more data than others and then cause data congestion. This problem mentioned above will cause the network performance degradation such as increase packet loss rate, decrease bandwidth utilization and raise communication delay [11]. Therefore the RSUs near station will become the bottleneck of the VANET.

In order to solve the existing problems, we have studied a hybrid hierarchical network architecture to transmit several types of data in the vehicular ad-hoc network with respect to some constraints, such as transmission latency, link unreliability and bandwidth.

*B. The Hybrid Three-Layer Network Structure*

In this subsection, we proposed a hybrid three-layer wireless network structure to transmit multiple different data in the vehicular ad-hoc network. This infrastructure of the hybrid network in the paper is a combination of optical fiber, LTE and WSN to improve the performance of the whole network. Nowadays, the optical fiber has been connected to the station for the infrastructure. We can take use of the optical fiber to reduce the cost of the overall network construction. TD-LTE is the next generation of wireless broadband technology and has many advantages, such as high bandwidth, low communication delay, convenient deployment and so on [12]. The reasonable deployment of LTE enabled RSUs as the data relay node, can effectively reduce the data traffic in the wireless sensor network. In this paper, the vehicle nodes are divided into clusters, and the data of each cluster can be transmitted through the LTE enabled RSUs, so that the data stream of each RSU is balanced.

Fig.2 shows our proposed three-layer vehicular ad-hoc network structure. The design involves the low cost wireless sensor network, the LTE network which has high data rate

links and existing optical fiber network. The wireless sensors are installed on each vehicle to form the WSN. The existing fiber network and LTE wireless network work as the link between control center and layer one.

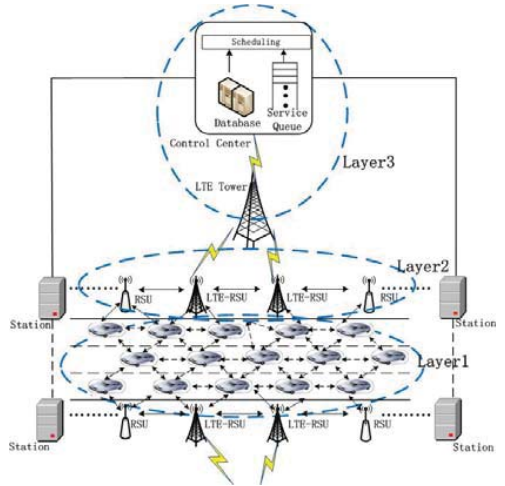


Fig. 2. Three Layer Vehicular Ad-Hoc Network Structure

The first layer is the vehicle nodes layer, composed of sensors installed in each vehicle. Each cluster of vehicle nodes has a cluster head, which is responsible for collecting the data of other vehicle sensors in the cluster and transmitting those data to the RSUs. The second layer of this network is responsible for RSUs communication, composed of RSUs and LTE enabled RSUs. This layer is responsible for the transmission of the data from layer one. Consider a packet composed of vehicle nodes data in the middle of the RSUs linear network. The data packet from these RSUs cannot reach either end station of the linear sensor network because of the limited bandwidth of the wireless sensor network. In these cases, the RSUs equipped with LTE transceivers offer an alternative way to transmit the packet far away from the station by the high bandwidth, low latency LTE links. The third layer network is the convergence layer mainly consists of stations, optical fiber and LTE towers. Stations will transmit the data received from vehicle nodes to the control center through the optical fiber, and the LTE base station will transmit the data through the LTE network. In this paper, we will use the three layer network to improve the overall performance of data transmission in VANET, such as throughput and the number of drop packets.

III. A RSUS DEPLOYMENT SCHEME DESIGN

In this section, we propose a new RSUs deployment scheme which can solve the problem mentioned above.

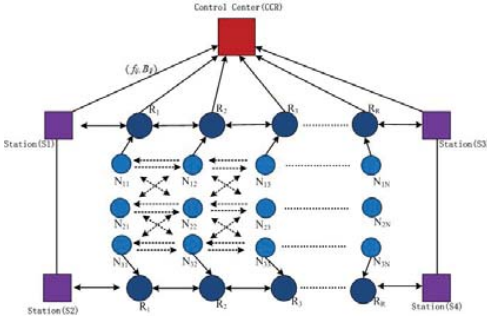


Fig. 3. The Model of Three-Layer Network Structure

### A. Network Model

As the Fig.3 shows, the whole vehicular ad-hoc network is modeled as a directed graph  $G = (V, E)$ , where  $V$  represents the set of vertices and  $E$  represents the set of edges. The set of vertices consists of vehicle nodes ( $N$ ), RSUs ( $R$ ), four stations ( $S$ ) and a control center ( $CCR$ ) and the sum of vertices is  $N_{1N} + N_{2N} + N_{3N} + 2 * R_R + 4 * S + CCR$ . Edge set represents the communication links which can be wired links ( $s, CCR$ ) where  $s \in S$ , LTE links ( $r, CCR$ ) where  $r \in R_R$ , wireless links ( $r, n$ ), where  $r \in R_R$  and  $n \in N_{1N} \cup N_{2N} \cup N_{3N}$ , and wireless ( $u, v$ ), where  $u, v \in N_{1N} \cup N_{2N} \cup N_{3N}$ .  $l_{ijk}$  is transmission latency of the  $k$ th flow incurred on the  $(i, j)$  link. A tuple  $(d_{ij}, b_{ij})$  is used to describe the performance of each link  $(i, j)$ , where  $d_{ij}$  is the delay of link  $(i, j)$  and  $b_{ij}$  is the bandwidth of the link  $(i, j)$ .  $COST$  means the construction cost of a LTE transceiver on a RSU.  $F_{ij}$ ,  $Y_i$  and  $L_{ijk}$  used in the formulation are binary variables. Each flow in the graph can be characterized by the tuple:  $F = (Source, Destination, c, b, D)$ , where  $Source \in N_{1N} \cup N_{2N} \cup N_{3N}$  is the data source in the flow,  $Destination$  is the data destination in the flow,  $c$  is cost of the flow,  $b$  is bandwidth required by the flow, and  $D$  is the transmission latency deadline of the flow.

### B. Deployment Problem Formulation

 TABLE I  
 SYMBOLS USED IN DEPLOYMENT FORMULATION

Symbol	Notation
$l_{ijk}$	Latency incurred by the $k$ th flow on link $(i, j)$
$L_{ijk}$	Binary variable. Is 1 if node $k$ selects edge $(i, j)$ as a link
$WIC$	Per Wireless link cost
$LTC$	Per LTE link cost
$b_{ij}$	Total bandwidth of the link $(i, j)$
$d_{ij}$	Total delay of the link $(i, j)$
$F_{ij}$	Binary variable. Is 1 if link $(i, j)$ used by any flow.
$LC_{ij}$	Total link $(i, j)$ cost
$D$	Transmission latency deadline
$COST$	Construction cost per LTE enabled RSU
$Y_i$	Binary variable. Is 1 if RSUs is enabled LTE.

According to the Fig.3, the optimal deployment problem can be modeled as the calculation of the weight average shortest

path in a directed graph  $G = (V, E)$  while considering the bandwidth and transmission latency constraints of each links. Thus, the input to the LTE enabled RSUs deployment algorithm is the  $N$  vehicle nodes where  $N \in N_{1N} \cup N_{2N} \cup N_{3N}$ ,  $R$  RSUs where  $R \in R_R$ , bandwidth constraint  $B$ , and the transmission latency constraint  $D$ . When the minimum cost path is chosen, the deployment algorithm outputs the network construction cost and a list of  $Y_i$  where  $i \in R_R$ . If  $Y_i$  is 1 means a LTE transceiver should be placed on the RSU  $i$ .

The different parameters used in the formulation are shown in table I and the LTE enabled RSUs deployment problem can be formulated as

Minimize:

$$f(F_{ij}, Y_i) = \sum_{(i,j) \in E} LC_{ij} F_{ij} + \sum_{i=1}^N COST \cdot Y_i \quad (1)$$

Subject to:

$$LC_{i,j}(n, k) = \sum_{v=1}^k WIC \cdot l_v + \sum_{u=k+1}^n LTC \cdot l_u \quad (2)$$

$$(i, j) \in E, v \in N, u \in R$$

$$\sum_{(i,j) \in E} L_{ijk} \cdot l_{ijk} \leq D, \forall k \in N \quad (3)$$

$$\sum_{(i,j) \in E} F_{ij} = 1, \forall i \in N \quad (4)$$

$$LC_{i,j,k} - LC_{i,CCR,k} \neq 0, \forall i, k \in R, \forall j \in N, Y_i \neq Y_k \quad (5)$$

$$LC_{i,j,k} - LC_{i,CCR,k} < 0, \forall j \in R, \forall k \in N \quad (6)$$

$$N = N_{1N} \cup N_{2N} \cup N_{3N} \quad (7)$$

$$R = R_R \quad (8)$$

$$F_{ij}, Y_i, L_{ijk} \in \{0, 1\} \quad (9)$$

From the equation (1), we can find our total cost  $f$  includes two parts: link cost and construction cost. Link cost is sum of wireless link cost  $WIC$  and LTE link cost  $LTC$ . Construction cost is a one-time cost of installing LTE transceivers on selected RSUs. As the equation (2) shows, each link  $(i, j)$  has two parts: one is using wireless link and the other is using LTE link. In hybrid three-layer network, there are  $k$  RSU nodes using wireless link and  $r - k$  nodes using LTE link.  $l_u$  and  $l_v$  in are both binary variables that  $l_v$  is 1 when node  $v$  is a vehicle node in link  $(i, j)$ ,  $l_u$  is 1 when node  $u$  is LTE enabled node in link  $(i, j)$ .

Equation (3) indicates the transmission latency of each link can not be more than the transmission latency deadline  $D$ . Equation (4) restricts each flow to only one source. Equations (5) depicts that the link cost  $LC$  changes when the LTE transceiver is installed on RSU. Equations (6) guarantees every flow of the network is between the vehicle nodes and  $CCR$ . Equation (7) and (8) explain that all the vertices must choice from the vehicle nodes or RSUs. Equation (9) ensures that  $Y_i, L_{ijk}, F_{ij}$  are binary variables. As the deployment formulation shows, our goal is to minimize the total cost  $f$  given in (1) under constraints (2) – (9).

IV. THE NETWORK PERFORMANCE ANALYSIS

To validate the data transmission performance in our proposed network architecture, we use the MATLAB simulator to establish a city scenario. The road is modeled as a linear network with a distance of 1 km. The distribution of vehicle nodes and data transmission mechanisms are simulated in IEEE 802.11p.

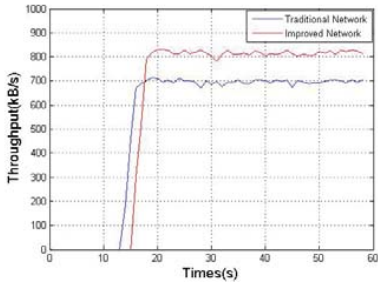


Fig. 4. The Throughput in Two Networks

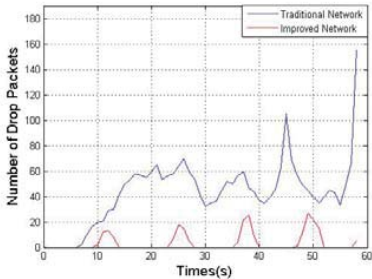


Fig. 5. The Number of Dropped Packets in Two Networks

Fig. 4 depicts the receiving throughput for UDP packets by vehicles of traditional VANET and improved network in the city environment. It is shown that the network proposed in this paper obtained around 830 kB/s for throughput while the traditional network obtained around 710 kB/s when pause time set between 13s and 60s. Fig. 5 shows the average number of dropped packets by two networks in the city scenarios. It is shown that the improved network has much less number of drop packets than traditional network. The reason for the higher packet drop in traditional network is due to the data congestion in RSUs near the station. And the use of LTE enabled RSUs will balance the data stream and reduce the drop packets in improved network. We can remark that the improved network performs better than traditional network.

Fig. 6 shows the average cost comparison between our improved deployment algorithm and the random deployment.

In the graph, as flow bandwidth increases, the random deployment cost increases, but sometimes the improved deployment cost may be equal, such as the flow bandwidth is 192 kbps and 224 kbps. This is because the same deployment of LTE enabled RSUs can meet the flow bandwidth requirements, so the network design does not change and the network construction cost remains equal. From the Fig. 6, we can learn the improved deployment algorithm can save much more cost than random deployment under the same bandwidth constraint.

Fig. 7 depicts the effect of transmission latency constraint on average cost. As mentioned before, the transmission latency deadline can be easily met by increasing the LTE enabled RSUs, so we change the price ratio of LTE in the whole construction. And we proposed a price scheme which can be described as a ratio of LTE enabled RSUs cost to other cost. For example, a price scheme(5:3) means the price ratio of LTE enabled RSUs to other things is 5:3. In Fig. 7, we assume three price schemes  $P1, P2, P3$ . The results show that as the transmission latency increases, the average cost decreases for all three price schemes due to the less number of LTE enabled RSUs. When the transmission latency constraint is bigger than 6.8s, the average cost becomes constant, which is because the overall network is limited by bandwidth rather than transmission latency.

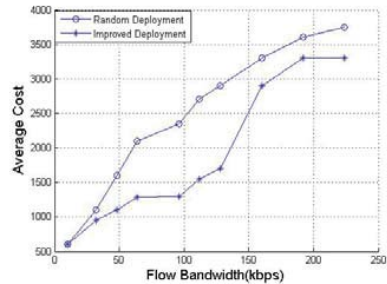


Fig. 6. Comparison between Two Deployment Schemes with Respect to Variation in Flow Bandwidth

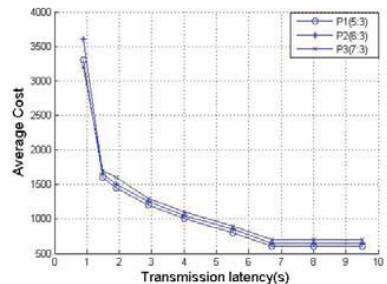


Fig. 7. Effect of Variation in Transmission Latency

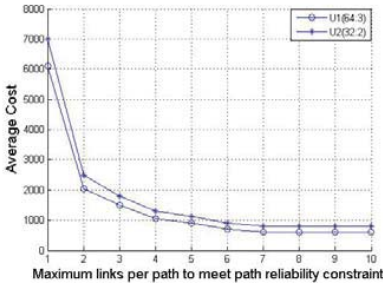


Fig. 8. Effect of Link Unreliability

Fig. 8 shows the network average cost in consideration of link unreliability. In the formulation proposed before, we consider the conditions of wireless link is ideal. However, wireless links usually are unreliable in reality due to random loss, multi-path fading, channel interface, etc [13]. In this simulation, we consider two constraints schemes named  $U_1$ ,  $U_2$  and a scheme (64:3) means the flow bandwidth is 64 kbps and the transmission latency deadline is 3s. At low link reliability, in order to maintain path reliability constraint, fewer nodes should be connected to the path. This leads to more LTE enabled RSUs being deployed on the side of road in order to increase the number of links. As link reliability increases, less LTE enabled RSUs need to be deployed. When 7 or more links per path in Fig. 8, the cost remains constant. This is because the system is limited both by bandwidth and transmission latency.

V. CONCLUSION

In this paper, we have researched the vehicular ad-hoc network(VANET). In order to improve the performance of the data transmission in VANET, we proposed a hybrid three-layer wireless network architecture. Furthermore, the trade-off between cost and performance was studied. We formulated a deployment algorithm of LTE enabled RSUs to optimize the construction cost while respecting all the constraints, such as bandwidth, transmission latency and link unreliability. In order to verify the performance of the hierarchical network proposed in our paper, we have presented a performance comparison between the improved architecture and traditional VANET architecture, using MATLAB. From results, we can come to a conclusion that the improved deployment algorithm performs better in cost reduction with considering of transmission latency and bandwidth. That suggests the hybrid three-layer wireless network used in VANETS can solve the problem

of real-time data transmission and reduce the overall cost of network. In our future work, we plan to carry out site implementation of the hybrid three-layer network as well as continuous performance study of this improved deployment algorithm against some other realistic data traffics.

ACKNOWLEDGMENT

This work is supported by Beijing Laboratory of Advanced Information Networks and National Science and Technology Major Project of China under Grant No.2016ZX03001009-003.

REFERENCES

- [1] K. Liu, J. K. Y. Ng, V. C. S. Lee, S. H. Son, and I. Stojmenovic, "Cooperative data scheduling in hybrid vehicular ad hoc networks: Vanet as a software defined network," vol. 24, no. 3, June 2016, pp. 1759–1773.
- [2] S. Benkirane, S. Mostafa, M. L. Hasnaoui, and A. Beni-Hssane, "A new comparative study of ad hoc routing protocol aody and dsr in vanet environment using simulation tools," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, Dec 2015, pp. 458–461.
- [3] H. Zhang, X. He, T. Luo, and W. Shi, "Transmission opportunity of spectrum sharing with cellular uplink spectrum in cognitive vanet," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.
- [4] V. Bondre and S. Dorle, "Design and performance evaluation of aomdv routing protocol for vanet," in *Computer, Communication and Control (IC4), 2015 International Conference on*, Sept 2015, pp. 1–4.
- [5] B. Ayyappan and P. M. Kumar, "Vehicular ad hoc networks (vanet): Architectures, methodologies and design issues," in *2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM)*, March 2016, pp. 177–180.
- [6] R. Cumbal, H. Palacios, and R. Hincapie, "Optimum deployment of rsu for efficient communications multi-hop from vehicle to infrastructure on vanet," in *2016 IEEE Colombian Conference on Communications and Computing (COLCOM)*, April 2016, pp. 1–6.
- [7] L. Cai, Z. Xu, Y. Wu, and X. Lin, "Transmission sequence reconstruction and allocation for vanet," in *2016 International Conference on Computer, Information and Telecommunication Systems (CITS)*, July 2016, pp. 1–5.
- [8] T. Karimireddy and A. G. A. Bakshi, "A hybrid security framework for the vehicular communications in vanet," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, March 2016, pp. 1929–1934.
- [9] H. Chour, Y. Nasser, H. Artail, and A. Kachouh, "Freddy: A framework for vanet aided d2d discovery," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.
- [10] C. C. Lin and D. J. Deng, "Optimal two-lane placement for hybrid vanet-sensor networks," vol. 62, no. 12, Dec 2015, pp. 7883–7891.
- [11] W. Li, X. MA, J. Wu, K. S. Trivedi, X. L. Huang, and Q. Liu, "Analytical model and performance evaluation of long term evolution for vehicle safety services," vol. PP, no. 99, Oct 2016, pp. 1–1.
- [12] S. Meng, D. Zhang, and B. Chen, "Research and implementation of dsp based on lmmse algorithm based on qr decomposition channel estimation in lte system," in *Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on*, Dec 2013, pp. 959–961.
- [13] K. Liu, S. H. Son, V. C. S. Lee, and K. Kapitanova, "A token-based admission control and request scheduling in lane reservation systems," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2011, pp. 1489–1494.

# Scalable Software-Defined Networking Based on Heap Structure

Yang Liu<sup>1</sup>, Zhuqing Jiang<sup>1</sup>, Yi Zhao<sup>2</sup>, Shanbao He<sup>3</sup>, Chan Wang<sup>1</sup>, Aidong Men<sup>1</sup>, Beihang Zhang<sup>1</sup>

<sup>1</sup>Beijing University of Posts and Telecommunications, Beijing, China

<sup>2</sup>Beijing Institute of Spacecraft System Engineering, Beijing, China

<sup>3</sup>China Academy of Space Technology, Beijing, China

**Abstract**—The new paradigm of Software-Defined Network (SDN) has emerged as a promising direction for next-generation network design. The concept of decoupling control plane from data plane enables the controller to acquire real-time information about the whole network, in other words, the configuration and management of data plane are centralized. However, a centralized control plane can cause problems in scaling up the network.

In this paper, we propose a new mechanism in scaling up SDN networks. Inspired by the form of mergeable heap in data structure, we propose a new Scalable SDN architecture which first abstracts the network layer by layer into a mergeable Fibonacci heap instead of a map, and we mainly focus on the scaling procedure in this new design. Next, we show the advantages brought by this design and we extend Dijkstra shortest-path algorithm which consider not only the edge weights but also the node weights to obtain better load-balancing. We also give a strategy to maintain node weights and edge weights dynamically for further implementation of our algorithm. Lastly, we evaluate our algorithm through comprehensive experiments and the results show that our algorithm outruns the original Dijkstra algorithm greatly.

**Keywords**—SDN; Network Scalability; Shortest Path; Fibonacci Heap; Extended Dijkstra

## I. INTRODUCTION

With the design of decoupling the control plane from data (forwarding) plane of network device, Software-Defined Networking (SDN) is believed to bring traditional network architecture with revolutionary transformation. By breaking the mode of traditional design of network, SDN simplifies network management and enables building networks that meet specific, end-to-end requirements. On one hand, the control plane is completely decoupled from the data plane, all the intelligence and calculation ability are integrated into the software in a logically centralized controller. On the other hand, SDN provides open programmability, which improves the flexibility and manageability of network, which are the most two incomparable advantages brought by SDN. OpenFlow, meanwhile, provides open interfaces that defines the structure and communication mechanism between the two decoupled planes [1]. The data plane is liberated from bondage to heavy tasks and now it only manages forwarding traffic that allocated by the controller. In addition, the operation of SDN network is based on open software instead of specific hardware that differs from each manufacture, which means a heterogeneous data plane composed of network elements from multiple vendors and encouraged by SDN as a contribution for reduction in capital and operating expenditures (CapEx/OpEx) [2].

Despite all the advantages, a number of factors argue against a single monolithic control point for the entire global network, among which scale is the obvious reason, along with related factors such as sheer propagation delay [3]. No matter in current commercial network or SDN testbed, it is well known that modern control plane requires flexibility, manageability, and scalability [4]. A centralized controller is somehow constrained to connect limited switches which means that the control plane is difficult to scale up. Scaling up SDN network not only requires to consider the logic connection between control plane and physical switches in forming up the network hierarchy, but also the increased cost of the expanded network scale. Therefore, scalability of control plane is one of the major challenges that is hindering the development of SDN.

Minlan Yu et al. proposed a scalable solution DIFANE that keeps all traffic in the data plane by selectively directing packets through intermediate switches that store the necessary rules [5]. DIFANE is a partitioning scheme to offload resource pressure in data plane. The logically centralized controller manages network partition information besides generating flow configuration rules as traditional controllers and continuously updates the network/traffic partitioning information to the switches. In this way, the administration pressure from missed flows in the data path can be partitioned and balanced among authority switches, thus scaling the control plane across the network. However, assumption of rules can be computed before the arrival of actual sample packets is too strong to be held in wide-area networks and the demand for proactive rule installation is not applicable for many reactive applications, either.

HyperFlow proposed by Tootoonchian et al. improves the scalability of software defined networks by partitioning the network to several physically distributed controllers [6]. Each sub-network can then be controlled by a controller, and the connections among sub-networks have bandwidth reserved for control messages. However, the bandwidth reserved for controller synchronization and the network scale is still limited by the achievable performance of one single controller. Another concern is about network consistency, because the worse-case delay in the network increases as the network scales.

Inspired by Fibonacci series, it is proposed in the work of Dai et al. that the SDN network can be abstracted layer by layer recursively [7]. From the perspective of scalability, the whole network is first abstracted as a single logical circle, and when a network task is initiated, sub-circles is derived from the global circle in a layer-by-layer way until network strategy is deployed at physical switches. This top-down abstraction view

can enhance scalability of SDN network. Especially when the network is undertaking functional tasks. However, the recursive abstraction design has focused too much on network scalability that some of the elementary network factors are neglected, which could be more than likely to aggravate network function disorders such as congestion, packet drop, etc. Therefore, based on the work of Dai et al., we propose a new design of scalable SDN and we extend Dijkstra algorithm that considers not only edge weights, but also associates node with weights in order to present the physical constraint of each node in route computation.

The rest of this paper is organized as follows: in section II we firstly introduce scalable SDN from the perspective of network scalability; in section III, the routing algorithm in scalable SDN is analyzed; in section IV, we introduce our extended Dijkstra algorithm in scalable SDN; in section V, the effectiveness of the algorithms are compared with each other through simulation; and section VI is conclusion and we indicate our future work.

## II. SCALABLE SDN

### A. Abstraction of The Network

In a software defined network, the whole network is controlled by a controller, or in many cases, by multiple controllers [8]. In scalable SDN, the whole network is abstracted as a single heap, which logically represents a super switch that connects all end systems. Then, the original heap can be divided into several sub sets as needed, which means to divide the super heap into several sub heaps logically. By repeating this process recursively, the whole network is divided into several parts according to specific demands such as routing. The recursive process ends when the division cannot be continued any further, i.e., physical switches are extracted as the last layer of this hierarchy. From a reverse perspective, this design of abstraction means that physical switches can be abstracted layer by layer until the whole network is virtualized as a single switch. This process can simplify many complicated network functions. For instance, each time when there is a request that is initiated by a source node in the network, the controller begins to abstract the network, all those nodes that are directly connected to the border router of the source node are firstly abstracted as one set, and then the other nodes connected to the previous set are abstracted as another set. When every physical node in the network has been abstracted, the abstraction of first logical layer ends and the abstraction of the second layer begins. This recursive procedure terminates when the whole network is finally abstracted into one single heap.

In this hierarchy architecture, several inter-connected physical switches are aggregated as a first-layer logical switch, then likewise, several first-layer logical switches are aggregated as one second-layer logical switch. By abstracting the network layer by layer, the global view of data plane from control plan becomes simpler and simpler. At the top layer, the super switch represents the whole data plane managed by the control plane. In a word, the upper the layer is, the more abstract the topology is, while the lower the layer is, the more concrete the data plane is.

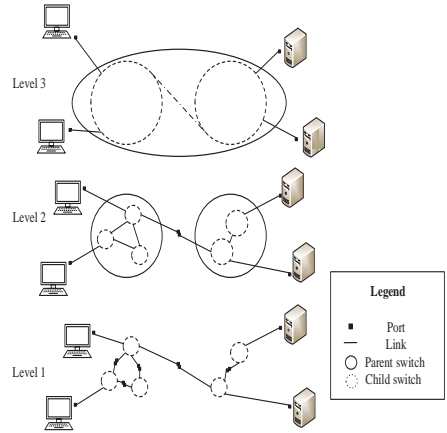


Fig. 1. Abstraction of a network.

### B. Scaling in Recursion

The core of Scalable SDN is to implement a recursive hierarchical architecture. According to the logical property of our hierarchical design, Fibonacci series, which is naturally a recursive series, is used as reference in building up Scalable SDN architecture. The Fibonacci series can be expressed in the following integer sequence.

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144 \dots \quad (1)$$

$$F_n = F_{n-1} + F_{n-2} \quad (2)$$

In a Fibonacci series, the sum of previous two numbers equals to the third number. This recursive relation can be expressed in Eq. 2. While in a Scalable SDN network, the super switch is the sum of several logical switches next layer below. And each logic switch at each layer itself is the sum of several logical switches next layer below. At the bottom of this hierarchy are the first layer of logical switches, each first-layer logical switch consists of several inter-connected physical switches. Inspired by this structure, we apply Fibonacci heap to abstract the Scalable SDN network. A Fibonacci heap is a collection of rooted trees that are min-heap ordered. That is, each tree obeys the min-heap property: the key of a node is greater than or equals to the key of its parent. Figure 2 shows an example of a Fibonacci heap. We use Fibonacci Heap to abstract the Scalable SDN network not only because of the mutually geometric similarity between the network and Fibonacci series but also the convenience brought by Fibonacci heap. For one thing, Fibonacci heap supports a set of operations that constitutes what is known as a *mergeable heap*, which we would apply in scaling SDN networks. For another, several Fibonacci-heap operations run in constant amortized time, which makes this data structure well suited for applications that invoke these operations frequently such as, in our case, UNION. The structure of Scalable SDN based on Fibonacci heap is shown in Figure 3.

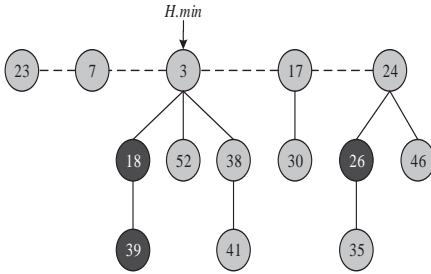


Fig. 2. A Fibonacci heap consisting of five min-heap-ordered trees and 14 nodes. The dashed line indicates the root list. The minimum node of the heap is the node containing the key 3. Black nodes are marked.

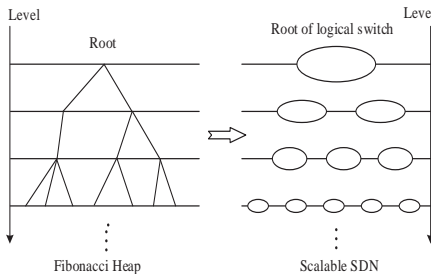


Fig. 3. Scalable SDN and spanning tree.

### C. To Scale is to Merge

We simplify the scaling of a SDN network as a process of adding several new nodes to a established network. In perspective of Fibonacci heap, it means to merge two heaps into one new heap. As we mentioned before, Fibonacci is a merge-able heap. A merge-able heap is any data structure that supports the following five operations, in which each element has a key:

MAKE-HEAP() creates and returns a new heap containing no elements.

INSERT( $H, x$ ) inserts element  $x$  into  $H$ .

MINIMUM( $H$ ) returns a pointer to the element in heap  $H$  whose key is minimum.

EXTRACT-MIN( $H$ ) deletes the element from heap  $H$  whose key is minimum.

UNION( $H_1, H_2$ ) creates and returns a new heap that contains all the elements of heaps  $H_1$  and  $H_2$ . In our work, we primarily focus on the UNION operation of Fibonacci heap because it is directly related to the scaling-up procedure in Scalable SDN.

### D. Scalable SDN Applied in a Routing Scenario

The fundamental function of Scalable SDN is to solve issues caused both during and after scaling up a SDN networks. After abstracting the whole network into a Fibonacci heap, logical switches at each level are responsible for computing shortest path and transforming the result into flow entries that are passed down in this hierarchy. When this information is

reaches the bottom of the hierarchy, it means that all the flow entries have been installed in related physical switches that form the path between source node and destination node. Thus routes are computed in terms of a top-down logical view.

In fact, not only the design of scalable SDN structure has made reference to Fibonacci series, but also that this hierarchical architecture utilizes Fibonacci heap to improve the calculation in routing, compared to conventional shortest-path algorithms. Technically, it is because of the fact that the Fibonacci heap have some valuable characteristics. Firstly, Fibonacci heap is quite similar to Binomial heap in terms of data structure, meanwhile the latter is in fact another form of spanning-tree. To abstract the network into a spanning tree not only simplifies networking functions, but also indicates that there is no loop when information flow transmits in this recursive hierarchical architecture. Secondly, the use of heaps to implement priority queues plays an increasingly important role in modern applications, because it provides an easy way to guarantee logarithmic running time for dynamic situations where large numbers of insert and remove the maximum operations are intermixed. Therefore, we apply our design in a routing scenario to study its efficiency.

### III. FIBONACCI-HEAP OPTIMIZED DIJKSTRA ALGORITHM

Dijkstra's algorithm is one of the most classical single-source shortest-path algorithm in graph theory, and also generally adopted by many prevalent routing protocols, such as OSPF. Suppose the graph  $G = (V, E)$ , where  $V$  is the vertex set of  $G$  and  $E$  is the edge set of  $G$ . Dijkstra algorithm maintains a set  $S$  of vertices whose final shortest-path weights from the source  $s$  have already been determined. The algorithm repeatedly selects the vertices  $u$  from a priority queue  $Q = V - S$  with the minimum shortest-path estimate, adds  $u$  to  $S$ , and relaxes all edges leaving  $u$ . The pseudo codes of Dijkstra shortest-path algorithm is shown as follows.

---

#### Algorithm 1 DIJKSTRA( $G, w, s$ )

---

```

1: INITIALIZE-SINGLE-SOURCE( $G, s$ )
2:  $S \leftarrow \emptyset$ 
3: PRIORITY QUEUE  $Q \leftarrow V[G]$ 
4: while  $Q \neq \emptyset$  do
5:    $u \leftarrow$  EXTRACT-MIN( $Q$ )
6:    $S \leftarrow S \cup \{u\}$ 
7:   for each vertex  $v$  in  $G.Adj[u]$  do
8:     RELAX( $u, v, w$ )
9:   end for
10: end while
    
```

---

Dijkstra shortest-path algorithm uses a greedy strategy to choose the "lightest" or "closest" vertex in  $Q$  to add to set  $S$ . But this step is very time-consuming because it spends too much time traversing the adjacency list and maintaining the priority queue every iteration. Let  $|V|$  be number of vertices, while  $|E|$  represents the amount of edges. The running time of Dijkstra algorithm depends on how the priority queue is implemented. Consider first the case in which we maintain the priority queue by taking advantage of the vertices being numbered 1 to  $|V|$ . Each INSERT and DECREASE-KEY (implicit in RELAX, which is called in line 8) operation takes  $O(1)$  time, and each EXTRACTMIN operation takes  $O(|V|)$

time. Since we have to search through the entire graph, for classical Dijkstra algorithm, the complexity can be expressed as in Eq.(3).

$$O(V^2 + E) \quad (3)$$

While in R-SDN, Fibonacci heap is introduced to implement the priority queue. The amortized cost of each of the  $|V|$  EXTRACT-MIN operations is  $O(\lg|V|)$ , and each DECREASE-KEY call, of which there are at most  $|E|$ , takes only  $O(1)$  amortized time. The overview of pseudo codes at this phase is shown as follows.

---

**Algorithm 2** EXTRACT-MIN WITH FIBONACCI HEAP
 

---

```

1:  $z \leftarrow \min[H]$ 
2: if  $z \neq \text{NIL}$  then
3:   for each child  $x$  of  $z$ 
4:     do add  $x$  to the root list of  $H$ 
5:    $p[x] \leftarrow \text{NIL}$ 
6:   remove  $z$  from the root list of  $H$ 
7:   if  $z = \text{right}[z]$  then
8:      $\min[H] \leftarrow \text{NIL}$ 
9:   else  $\min[H] \leftarrow \text{right}[z]$ 
10:    CONSOLIDATE( $H$ )
11:     $n[H] \leftarrow n[H] - 1$ 
12:  end if
13: end if
14: return  $z$ 
    
```

---

And the complexity of optimized Dijkstra algorithm is shown in Eq.(4).

$$O(V \lg V + E) \quad (4)$$

#### IV. EXTEND THE OPTIMIZED DIJKSTRA ALGORITHM

##### A. Motivation to improve Scalable SDN

Though scalable SDN design has many advantages in scaling up a SDN network, it still has some shortcomings. Most importantly, it is designed mainly based on unicast situation, but multicast cases are much more complicated. When generating the first layer of logical switches, the aggregation of physical switches only takes the connection among each node into consideration. Specifically, when calculating a route, only those nodes that are connected either to the source node or the destination node are chosen. But under practical circumstances, it is more than likely that a certain intermediate node with high degree may participate in multiple sessions, which means that this node can be added into different routes in a short period. Considering its limited physical resource and processing capacity, this node can be overloaded. This could increase processing delay and queuing delay, and in some worse cases, packets may be dropped according to SDN protocol[5]. From a global perspective, end-to-end delay has increased much more than the amount decreased in routes calculation.

##### B. Extended Dijkstra Algorithm

In the original Dijkstra algorithm, nodes are associated with no weight. To avoid consuming the physical resource of each

node, we propose to extend Dijkstra algorithm by adding node weights in RELAX operation. This thought is mainly inspired by the work of Jehn-Ruey Jiang et al. [9]. Jiang et al. had added node weights in Dijkstra algorithm in a purpose of solving single-source shortest path problem. Instead, our motivation is to improve load-balancing in scalable SDN. In addition, we devise the node weight  $nw[u]$  to be renewed after each iteration to obtain dynamic management of scalable SDN. We show the pseudo codes of our algorithm below including the extended part of Dijkstra algorithm along with Fibonacci Heap as follows.

---

**Algorithm 3** OPTIMIZED DIJKSTRA ALGORITHM
 

---

```

1: Input:  $G = (V, E), ew, nw, s$ 
2: Output:  $d[|V|], p[|V|]$ 
3:  $d[s] \leftarrow 0; d[u] \leftarrow \infty$ 
4: for each  $u \neq s, u \in V$ 
5:   insert  $u$  with key  $d[u]$  into the Fibonacci Heap
6: while  $Q \neq \emptyset$  do
7:    $u \leftarrow \text{Extract} - \text{Min}(Q)$ 
8:   for each  $v$  adjacent to  $u$  do
9:     // REALX
10:    if  $d[v] > d[u] + ew[u, v] + nw[u]$  then
11:       $d[v] \leftarrow d[u] + ew[u, v] + nw[u]$ 
12:       $p[v] \leftarrow d[u]$ 
13:      update  $nw[v]$ 
14:    end if
15:  end for
16: end while
    
```

---

The extended Dijkstra algorithm is similar to the original Dijkstra algorithm but we have added the node weight to RELAX procedure as shown in line 10 and line 11 of the pseudo codes above. In light of Dijkstra work, we can prove that the extended algorithm indeed return the shortest path from the source node to every other node with the consideration of the edge weights and node weights. To save space, we omit the proof of the above statement.

Our algorithm is very useful in deriving the best routing paths in a concurrent communication SDN network (i.e., there are more than one session initiated in the network during a same period) in which significant latency occurs when the packet goes through intermediate nodes that is repeatedly occupied by different routes. Below, we show how to define the node weights.

Assume we can derive from the SDN topology a graph  $G = (V, E)$ , which is weighted, directed, and connected. For a node  $v \in V$  and an edge  $e \in E$ , let  $Flow(v)$  and  $Flow(e)$  denote the set of all the flows passing through  $v$  and  $e$ , respectively, let  $Capacity(v)$  be the capacity of  $v$  (i.e., the number of bits that  $v$  can process per second), and let  $Bandwidth(e)$  be the bandwidth of  $e$  (i.e., the number of bits that  $e$  can transmit per second). The node weight  $nw[v]$  of  $v$  is defined according to Eq.(5), and the edge weight  $ew[e]$  of  $e$  is defined according to Eq.(6).

$$nw[v] = nw\_penalty\_coeff \cdot \frac{\sum_{f \in Flow(v)} Bits(f)}{Capacity(v)} \quad (5)$$

where  $Bits(f)$  stands for the number of bits in  $f$  processed



by node  $v$  per second.

$$ew[e] = ew\_penalty\_coeff \cdot \frac{\sum_{f \in Flow(e)} Bits(f)}{Bandwidth(e)} \quad (6)$$

where  $Bits(f)$  stands for the number of bits in  $f$  passing through edge  $e$  per second.

Note that node weight  $nw[v]$  is updated in each iteration as in line 12. This is reasonable in a large-scale dynamic network because any busy node may idle down before another session request comes in. Also, in our definition, we use bit rate to calculate node weights, and this could be replaced by other parameters such as **port-usage rate** ( $num(ports\_activated)/num(ports\_available)$ ) or **buffer-occupation rate** ( $length(queue)/buffer\_size$ ), etc. In doing that, the network is endowed with more flexibility and service request can be arranged with more granularity. In addition, this design copes better with one principle of Software Defined Networking that applications can request needed resources from the network via interfaces to the control plane.

As for edge weight  $ew[e]$ , we do not define the update process because compared to physical switches, edges are relatively static. But still, we recommend  $length$  of  $e$  divided by *average edge length* as a potential substitute for the fraction in Eq.(6).

Two penalty coefficients, which originates from **empirical minimum Bayesian risk theory**, are added to each equation respectively in order to balance the influence on  $nw[v]$  and  $ew[e]$  caused by each other. For instance, if several intermediate nodes are severely occupied while most edges are not, the  $nw\_penalty\_coefficient$  should be much larger than the  $ew\_penalty\_coefficient$ , thus forcing the model inclined to return a detour around those intermediate nodes.

## V. SIMULATION

### A. Simulation Environment

Running time of computing shortest path plays a key role in evaluating algorithms. Therefore, in this section, we mainly choose computation time as the criterion to compare the modified routing algorithm with other algorithms. Since there is no real SDN network that we can apply our model, all algorithms are simulated in Matlab(R2015b) which running on a desktop computer. The datasheet of the desktop computer is: 3.3GHz of quad-core processor; 8G memory; Windows 7 OS.

First of all, we explain how to generate the topology we used in our simulation. In our simulation, random graph is used to simulate a network topology. By "random", we mean not only the topology is randomly set, but also that both the edge weights and node weights are created randomly from a uniform distribution ranged in the interval of [1,100]. Two numbers are required as input: positive integer  $n$  as the number of nodes in total; and decimal number  $d$  between 0 and 1 represents density of the graph. All random graphs are generated in the following steps.

Firstly, an  $n$  by  $n$  matrix  $M$  is generated with diagonal elements set to 0, while the value of any other element is generated from a uniform distribution in [0,100]. Secondly, we generate another  $n$  by  $n$  **connection-matrix**  $C$  with only

either 0 or 1 assigned to each element. To assign each element in  $C$ , we first generate a random decimal  $t$  between 0 and 1, and if  $u < d$ , we assign a 1 to this element, otherwise we assign 0 to it. Then we get a transition matrix  $TW$  which is the dot product of  $M$  and  $C$ . Lastly, we obtain the edge-weight matrix  $W$  by setting the value of all zero elements in matrix  $TW$  to  $inf$ . Note that each  $w_{ij}$  in  $W$  represents the weight of edge  $i \rightarrow j$ , and if  $w_{ij} = inf$  means that there is no edge connects node  $i$  and  $j$ . Approximately, the weight of each node is randomly created from a uniform distribution in [20, 50], and all the node weights are restored in a 1 by  $n$  vector. However, if the graph generated is a connected graph, the above would be repeated again and again until it meets the requirement. As a result, we get the random graphs we used to test operation time of above mention algorithms.

### B. Simulation Results

We simplify the concept of unicast as finding the shortest path between the source and a given node. Similarly, multicast is interpreted as the calculation of shortest paths between source node to multiple nodes. We set density to 0.03 and 0.09 and change the number of nodes in each graph we test with. After running our simulation on random graphs with vertices ranging from 10 to 900, we record the running time in each algorithm. The results are shown from Fig.4 and Fig 5.

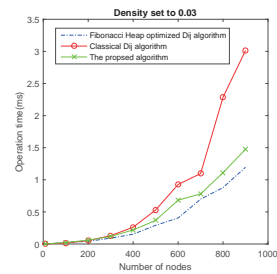


Fig. 4. Density set to 0.03 in each graph

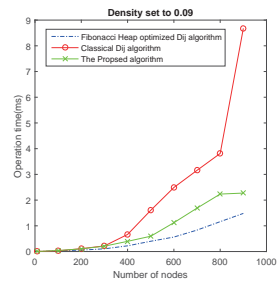


Fig. 5. Density set to 0.09 in each graph

In each graph, X-axis stands for the number of nodes in each graph, Y-axis represents the operation time in microseconds. Red line stands for the results of classical Dijkstra

algorithm, green line represents our proposed algorithm and blue line depicts the result of Dijkstra algorithm optimized by Fibonacci heap.

### C. Analysis

From the above two graphs, we can see that operation time of the proposed algorithm is always shorter than that of classical Dijkstra algorithm, and the more nodes within the random graph, the more obvious the difference is. In fact, as shown in Eq.3 and Eq.4, the complexity of each algorithms leads to their different performances. Nevertheless, we have also noticed that, our algorithm spends slightly longer time than Fibonacci Heap optimized Dijkstra algorithm in each test. This is because we have added an updating procedure in each iteration to maintain node weights, which in turn consumes more time.

## VI. CONCLUSIONS

In this paper, we propose a new design of SDN network in order to solve scaling problems, and we have discussed both the advantages and disadvantages of this design in details. The proposed Scalabl SDN architecture primarily derives from the data structure of Fibonacci heap and based on that, we optimize the shortest-path algorithm in a further step. We associate nodes with weight in order to include the constraints of each physical switch into consideration. We also give strategies to update node weight. The above simulation reflects our model outruns the original Dijkstra algorithm significantly in terms of operation time, though our algorithm is slightly more time consuming than the algorithm optimized by Fibonacci heap in unicast situations.

The algorithm proposed in this paper can also be applied to other networks to help solve traffic engineering problems. Our future work is to apply our model on Floodlight project to test our algorithm in a simulated SDN network, and to evaluate its performance in load-balancing in a more practical simulation environment.

## VII. ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (61671077 and 61671264).

## REFERENCES

- [1] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.
- [2] R. Udechukwu and R. Dutta, "Extending openflow for service insertion and payload inspection," in *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, pp. 589–595, IEEE, 2014.
- [3] "Sdn architecture," pp. 67–70, ONF, 2016.
- [4] C. Elliott, "Geni: Opening up new classes of experiments in global networking," *IEEE internet computing*, vol. 14, no. 1, pp. 39–42, 2010.
- [5] M. Yu, J. Rexford, M. J. Freedman, and J. Wang, "Scalable flow-based networking with difane," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 351–362, 2011.
- [6] A. Tootoonchian and Y. Ganjali, "Hyperflow: A distributed control plane for openflow," in *Proceedings of the 2010 internet network management conference on Research on enterprise networking*, pp. 3–3, 2010.
- [7] W. Dai, G. Shou, Y. Hu, Z. Guo, and J. Liu, "Extending sdn network with recursive architecture," in *Wireless Personal Multimedia Communications (WPMC), 2014 International Symposium on*, pp. 491–496, IEEE, 2014.
- [8] H. Yang, J. Zhang, Y. Zhao, J. Han, Y. Lin, and Y. Lee, "Sudo: software defined networking for ubiquitous data center optical interconnection," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 86–95, 2016.
- [9] J.-R. Jiang, H.-W. Huang, J.-H. Liao, and S.-Y. Chen, "Extending dijkstra's shortest path algorithm for software defined networking," in *Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific*, pp. 1–4, IEEE, 2014.

# *Towards a configurable state model for 5G radio access networks*

Sofonias Hailu<sup>1</sup>, Mikko Säily<sup>2</sup>, Olav Tirkkonen<sup>1</sup>

<sup>1</sup>Department of Communications and Networking, Aalto University, Finland

<sup>2</sup>Nokia Bell Labs, Mobile Networks, Nokia, Finland

{Sofonias.hailu;olav.tirkkonen}@aalto.fi, mikko.saily@nokia-bell-labs.com

**Abstract**— The 5th Generation of mobile communications (5G) is expected to have a large number of use cases with diverse and sometimes contradictory requirements. This needs to be taken into account in state model design for 5G. Recent advances show that there is a need for a new power efficient state, sometimes called RRC Inactive state, in addition to RRC Idle and RRC Connected. In this paper, we argue that RRC Inactive needs to be highly configurable in order to address the diverse 5G use cases and requirements without further increasing the number of states. A list of potential configurations are discussed that could help the network to tune the behavior of a User Equipment (UE) during RRC Inactive. The configurations are performed such that the service requirement of the UE is fulfilled. It is shown via an example how the network can fulfil the requirements of two UE groups with diverse requirements without the need to have separate inactivity states.

**Keywords**—5G architecture; State model; Radio Resource Configuration protocol;

## I. INTRODUCTION

Mobility state handling is one of the crucial functions of 3GPP wireless technologies. In Universal Mobile Telecommunications System (UMTS), the state machine consists of one idle mode state and four connected mode states [1]. The idle mode state is optimized for low power and memory consumption. There is no User Equipment (UE) Radio Access Network (RAN) context stored in the UE nor the network. The connected mode states are optimized for fast connection re-establishment. For this reason, the interface between RAN and Core Network (CN) is kept active and the UE's RAN context is stored in the UE and the network. The differences among the connected mode states are first in the physical channels allocated to the UE, and the transport channels that can be used, second in the type of Radio Resource Control (RRC) connection mobility activity, and finally in the level of UE activity. The connected mode states have several sometimes significantly overlapping characteristics. For example, URA\_PCH and CELL\_PCH have almost identical characteristics except that in URA\_PCH, the location of a UE is tracked on UMTS Routing Area (URA)-level that consists of multiple cells, while it is tracked on a cell level in CELL\_PCH. Such significant overlap among the characteristics of the states increases the overall number of states. Having multiple states is complex to implement and increases the standardization effort.

In Long Term Evolution (LTE) releases 8 to 12, the number of states is scaled down to two; the RRC Idle and RRC\_Connected states, see releases 8 to 12 of [2]. This reduces the standardization effort and simplifies the state machine. RRC Idle is optimized for low UE power and memory consumption. No UE Access Stratum (AS) context is stored in the UE nor the network. The RRC Connected state is optimized for high UE activity where the UE AS context is stored in the UE and the network, and the RAN/CN connection is active. Connected mode Discontinuous Reception (DRX) is adopted in order to enable different levels of UE power saving in RRC Connected. The LTE's two-state model works well for the Mobile Broadband (MBB) use case [3].

However, the LTE's two-state model is shown to be inefficient to handle use cases that transmit frequent small packets, e.g. sensors or smartphone keepalive messages from social network applications [3][4][5][6]. This is due to the heavy signaling procedure to transition from RRC Idle to RRC Connected which is costly to perform for the transmission of a small packet. It is also costly to keep many UEs without active data transmission in RRC Connected, e.g. due to signaling overhead from handovers and measurement reports. The inactivity timer is typically configured to be between 10 and 60 s which forces the network to release the UE to RRC Idle in most cases.

In order to address these issues for Cellular Internet of Things (CIoT), studied under NB-IoT, RRC Suspended has been introduced in LTE release 13 [2]. In this state, the UE AS context is stored both in the UE and the network. This reduces the number of messages required to establish the radio connection; the transition from RRC Suspended to RRC Connected requires less messages compared to the transition from RRC Idle to RRC Connected. However, the RAN/CN interface is released when a UE enters RRC Suspended. Thus, the number of S1 messages remains the same for both transitions. This prompts a further study to enhance or replace RRC Suspended to a state where the RAN/CN interface is kept, see e.g. the LTE light connection work item for release 14 [7].

The discussion on state machine design for 5G is currently ongoing both in academia [8] and in industry [9]. Most, if not all, of these studies propose a three state model which consists of LTE's RRC Idle and RRC Connected, and a new state which sometimes is referred as RRC Inactive. The role of RRC Idle is expected to be limited. It is seen as a bootstrap state during Public Land Mobile System (PLMS) selection and as a

recovery state during faults. RRC Inactive and RRC Connected are expected to be the main mobility states of the UE. In RRC Inactive, the UE's AS context is stored in the UE and the network, and the RAN/CN connection is kept. This reduces the state transition signaling overhead and latency [8]. UE autonomous mobility is used in RRC Inactive to minimize UE battery consumption. However, it is not clear yet how the state machine with limited number of states can handle UEs with highly diverse and sometimes contradictory requirements [10].

In this paper, we argue that RRC Inactive should be configurable for the state machine to be able to handle the mobility state of UEs with diverse requirements without the need to increase the number of states. With this approach, the UE behavior during RRC Inactive is configured according to the requirements of the UE services, e.g., battery life, latency, bandwidth, mobility, etc. A list of possible procedures that can be configured are discussed. Examples of configuration for two groups of UEs with diverse requirements are given to show how the requirements of the UEs can be fulfilled without the need to have separate RRC states during UE inactivity.

The paper is organized as follows: Section II summarizes the state machine currently under discussion for 5G. Section III discusses the motivation for the need to have a configurable state, and potential configurable procedures that can tune the behavior of a UE in the RRC Inactive. Section IV gives examples of configuration for two groups of UEs with diverse service requirements, as well as performance results indicating that the service requirements of the UE groups are fulfilled without the need to have a separate RRC state for the UE groups. In Section V, conclusions are drawn.

## II. POTENTIAL STATE MACHINE FOR 5G

The standardization of a state machine for 5G is currently ongoing in 3GPP. The state machine design is also ongoing in the academics, and discussed in the literature [8][10]. Based on these discussions, the state machine similar to the one shown in Figure 1. The model consists of RRC Idle, RRC Connected and RRC Inactive states. State transitions from RRC Idle to RRC Connected are expected to occur mainly during initial access (e.g. when the UE attaches to the network) or as a fallback case (e.g. when the devices and/or network cannot use the previously stored RAN context). Consequently, this transition is not expected to occur as often as in LTE. On the other hand, transitions from RRC Inactive to RRC Connected are expected to occur quite often and should be optimized to be lightweight and fast.

Some of the characteristics of RRC Inactive includes [8]:

### A. Maintenance of context information

When the UE moves to RRC Inactive (e.g. via an inactivity timer or via explicit network signaling) both the UE and the network will keep RAN context information that has been obtained during the previous RRC connection setup e.g. when the UE has attached itself to the network and moved from RRC Idle to RRC Connected. Context information consists of, e.g., the UE capabilities and the security context.

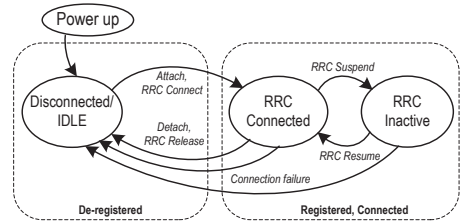


Figure 1: Potential state model for the 5G architecture

### B. Fast and lightweight transition to RRC Connected

It is envisioned that this state transition is handled by a procedure called RRC Resume and RRC Suspend, inspired by [2]. The RRC Resume procedure involves the resumption of Signaling Radio Bearers (SRB) and Data Radio Bearers (DRB). The procedure is triggered by the UE either in response to a page, when the UE has uplink data in buffer, or when it needs to send location updates. Upon receiving RRC Resume Request, the network retrieves the UE context based on the RAN Context ID, performs RAN context fetching and responds with RRC Resume to reconfigure SRBs and DRBs.

### C. Transparent inactivity to the CN

In the proposed state model the CN/RAN context is stored, i.e., the transitions between RRC Connected and RRC Inactive are transparent to the CN. Therefore, incoming packets from the CN may be forwarded to the latest mobility anchor point at the RAN so that transitions from RRC Inactive to RRC Connected do not involve CN signaling such as the setup of a CN/RAN interface connection.

### D. UE-controlled mobility and RAN-based paging

It is envisioned that in RRC Inactive, that the UE can be configured to monitor the paging channel(s) so that the UE may be reached by the network. As the RAN/CN interface is kept, RAN may potentially control UE mobility and reachability in this state.

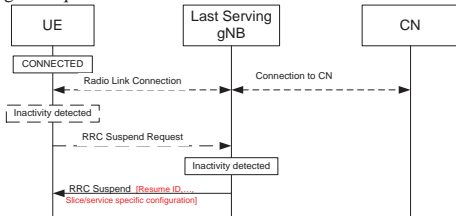
## III. CONFIGURABILITY OF RRC INACTIVE

The 5G use cases have diverse service requirements in terms of battery life, latency, mobility, security, bandwidth, etc [11]. These use cases are often categorized into three service types: Enhanced Mobile Broad Band (eMBB), Massive Machine Type Communication (mMTC) and Ultra Reliable and Low Latency (URLL) [10]. Rather than having independent networks optimized for each service type, single 5G architecture is envisioned that addresses these diverse and contradictory requirements. This design requirement, in fact, also applies to the UE state model.

In addition to the common characteristics and default procedures that are applicable to all scenarios, use cases, etc, the RRC Inactive state is envisioned to include procedures that are configurable based on, e.g., services that are provided to the UE. This can be done, for example, by including service tailored configurations in RRC suspend messages in addition to common configurations, see Figure 2. The Next Generation

NodeB (gNB) in the figure represents a base station that supports 5G radio interface(s) protocol(s).

A service could be characterized based on its requirements on, e.g., mobility, security & privacy, reliability, bandwidth, latency, battery life, etc. The service tailored configuration could then be prepared such that the requirements for the service(s) provided to the UE are fulfilled with minimal signaling overhead impact on the network. For example, a power optimized configuration might be used for a UE with a requirement for a long battery life; whereas a configuration optimized for latency can be used for UEs with applications requiring low latency. This way, the RRC Inactive state is flexible enough to handle the state of UEs with different service requirements. If the UE has multiple purposes, e.g. if it is capable of running multiple concurrent services, the configuration may be based on the service with the most stringent requirements.



**Figure 2: RRC suspend procedure with service specific configuration included in the RRC Suspend message.**

Potential configurable procedures in RRC Inactive state are discussed below.

**A. Widely configurable DRX**

The wide diversity of 5G use cases will also lead to devices with very different traffic patterns and battery requirements. For example, smart meters may transmit very small packets very seldom while requiring long battery life. In addition, there will be use cases requiring quite low control plane latency i.e. fast transmission of first packets while battery life time might not be an issue (e.g. cars sending safety related information). In order to cover such disparate cases, RRC Inactive, envisioned to be the primary sleeping state, has to support widely configurable DRX. Some devices may need to sleep for hours and minutes, while others need to wake up only once a day but still benefit from the fact the RAN context is stored e.g. for quick connection establishment. As an example, mMTC devices performing infrastructure metering and environmental monitoring can indicate to the network with their service identity that they are stationary and can be operated with long DRX cycles for very long battery life.

**B. Single/Multi-RAT camping, access and UE-based mobility**

To benefit from its wide coverage in the 2020 time frame, the evolution of LTE should be tightly integrated to the 5G RAN [10][12]. A solution relying on a common PDCP and RRC frameworks has been recently proposed [12], inspired by previous research in the area of multi-RAT integration [13].

This will enable the dynamic usage of all available resources in the access networks. If connection setup takes long time, the free resources may vanish. These scenarios also motivate dual radio UEs have a common CN connection for both access methods.

A common control plane framework where the evolution of LTE is part of 5G RAN demands common state handling for dual radio UEs. Otherwise, toggling between the novel 5G RAT and LTE coverage would lead to signaling to update the LTE state. In RRC Inactive a dual-radio UE can be configured to either camp on LTE or on a novel 5G RAT (and monitor the respective paging channels accordingly), as in current system. However, an alternative would be to configure dual-radio UEs to camp simultaneously on both RATs, i.e., monitor paging channels of both RATs and possibly try to access both RATs simultaneously. This could be beneficial for use cases requiring a very fast establishment of multi-connectivity. Another possible alternative could be to configure the UE to only monitor the paging channel of one of the RATs, e.g., known to have a better downlink coverage within a certain area, while random access requests can be sent towards both RATs in order to increase the diversity.

If a UE is configured to camp on both RATs at the same time there is no need to inform the network which of the RATs is providing the best coverage since the network knows that the UE is configured to monitor both paging channels. However, for battery efficiency reasons some UEs may be configured to only camp in one RAT at the time so that the UE can be configured not to report every time it reselects from one RAT to another. That can be used to save UE battery with the expense of spending more downlink resources to page these UE. One consequence is that this may increase the network-based access latency.

**C. Potential optimizations for the state transition**

There is potential for optimization of the state transition depending on the service characteristics of a UE. For example, a UE with applications transmitting frequent small packets may use a different state transition approach than a UE with applications transmitting large packets. A UE with the small packet might send the packet as a payload of the RRC connection resume request message using the already established security context [14] or using a contention based transmission [15]. However, a UE with a large packet cannot do that simply because its packet is too large to be added as a payload to the RRC connection resume message. Further MAC-level optimizations may also exist such as MAC control element based signaling for state transition, for example, for semi-static devices.

**D. Measurement configuration**

Measurement configuration can be tailored to the characteristics of the service provided to a UE. For example, a semi-static UE can be configured to skip neighbor cell measurements and related reports. A UE with applications requiring ultra-reliable communication, on the other hand, might be configured to monitor the channel from possibly multiple-RATs for quickly establishing multi-connectivity upon connection resumption. Another example could be to

provide a power optimized measurement configuration for a battery operated UE with long battery duration requirement.

In order to address the envisioned frequencies (up to 100 GHz), the new 5G radio will massively rely on beamforming technologies. A quite wide range of transmission schemes can be considered, from very wide to very narrow beams. Therefore, the network can configure UEs to possibly monitor dedicated reference signals of specific beams. RAN may thus configure some UEs to follow very narrow beams while other UEs could be configured to measure signals transmitted in wide beams.

#### E. UE paging and location tracking

As CN/RAN interface during RRC Inactive is kept, the RAN may control the paging and location tracking of the UE. The gNB that terminates the CN/RAN interface for the UE may act as an anchor gNB, serving as a paging agent and mobility anchor. The anchor gNB is required to maintain the inter gNB interface relationship with all the gNBs in the tracking area of the UE. It may have to buffer and forward Mobile Terminating (MT) data if the UE were camping in another gNB when an MT packet arrives. These requirements may limit the typical paging area to be small. If the tracking area is large, the anchor gNB might not have an inter gNB interface available with all the gNB in the tracking area. Therefore, RAN controlled location tracking and paging is mostly suitable for low mobility UEs. For such UEs, the paging area could be configured to be small, i.e. a single cell or a group of cells that are under the anchor gNB.

However, for UEs that require high mobility, defining a larger tracking area is beneficial. In this case, a high mobility UE may be configured for CN initiated paging and hybrid CN/RAN location tracking.

#### F. Synchronisation configuration

In LTE, UL synchronization is released when a UE enters an RRC Idle state. One of the reasons for the need of a UE to carry out a RACH procedure when transitioning to RRC connected state is to get UL synchronization. In RRC Inactive, most UEs are not expected to keep UL synchronization. In this case, RACH will be required to resume the RRC connection.

However, some UEs, e.g. mMTC devices with low latency requirements, may need to be configured to keep UL synchronization during RRC Inactive state and to monitor a dedicated channel. This enables seamless resumption of the RRC connection fulfilling tight latency requirement or for low latency contention based transmission.

#### G. Potential optimizations for small data transmission

The new 5G radio needs to support efficient small packet transmission. This would be crucial as many smartphone applications and mMTC devices are expected to generate small packets. In particular, the transition from RRC Inactive to RRC Connected needs to be realized with a light-weight signaling procedure. Thus, the transmission of packets from RRC Inactive would not be costly.

However, for some use cases, it might be beneficial to use optimized solutions for small packet transmission, e.g. see in [14][15]. In such cases, the behavior of a UE in RRC Inactive state can be configured such that the preferred small packet transmission method is used. For example, some mMTC devices that require very long battery life might be configured to transmit data using contention based transmission while still in RRC Inactive Connected. This would prolong their battery life by reducing the time that the mMTC device is kept awake; which in turn reduces device power consumption.

### IV. PERFORMANCE ANALYSIS

We consider two example use cases and their potential configurations for analyzing the performance of RRC Inactive. In particular, we show how the configurability of the RRC Inactive can be used to fulfil the diverse requirements of use cases without the need to have separate inactivity states for each use case. The two use cases considered in this example are mMTC and eMBB.

#### A. An mMTC device

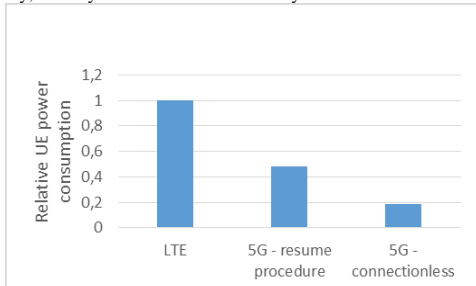
There are several use cases of the mMTC service type. In this example, we consider an mMTC device with the following characteristics: 10 to 15 years of battery life, no mobility, transmission of intermittent small UL packets. Such a device could be, for example, a sensor that transmits intermittent measurement reports. It is logical to configure the behavior of the mMTC device in RRC Inactive such that its battery life is prolonged. Such a power efficient configuration may include

1) *DRX configuration*: As the mMTC device is assumed to have only UL transmissions, it would be natural to configure the DRX cycle of the UE to be long and equal to the tracking area update timer duration to avoid battery drainage. Since the UE has only UL transmissions, the UE might be configured not to monitor the paging channel for paging but only to update its registration during the ON duration of the DRX cycle.

2) *Measurement configuration*: Since the UE is static, the mobility functionality might not be needed at all. Thus, the UE is configured not to perform cell reselection to avoid UE battery consumption.

3) *Small packet transmission configuration*: During every UL transmission, the mMTC device has only one small packet to transmit. From the battery consumption perspective, it would be preferable to transmit the small packet and return to DRX mode as quickly as possible. Therefore, it would be more power efficient to transmit the packet using contention based access, see e.g. [15], than transmitting the packet by transitioning to the RRC Connected. This conclusion can be observed from *Figure 3* where the power consumption model of [16] is used and transmission latencies are taken based on [17, Annex B]. A connectionless transmission is based on a two-stage contention based procedure where data is sent after random access message 2 is received [15]. It should be noted that the power consumption may further be reduced with a one-stage contention based transmission if it has low collision probability. It is worthwhile to note also that the power

consumption to transmit a small packet is the main, if not the only, activity that consumes the battery of such a device.



**Figure 3: Power consumption to transmit a small packet (with almost a zero-byte size).**

### B. eMBB device

The eMBB device typically has a large amount of data to transmit unlike the mMTC device. Taking this into account, the configuration of the eMBB device in RRC Inactive may be as follows.

1) *Camping configuration*: An active eMBB device typically has a large amount of data to transmit. Thus, it will be beneficial if the eMBB device is configured to camp on a cell that can provide a relatively large transmission bandwidth. For example, the frequency layers that provide higher bandwidth are given higher priority during cell reselection. However, when preparing such an optimized priority list, the UE's battery life requirement should be taken into account as well.

2) *State transition configuration*: As an eMBB device typically has a large packet to transmit, it should be configured to first resume the RRC connection before transmitting its packet. The state transition should be kept as light as possible. In [8], it is shown that a state transition using resume procedure is light compared to LTE's state transition procedure from RRC Idle to RRC Connected.

3) *Paging channel monitoring configuration*: An eMBB device can be configured with a DRX cycle that takes into account the trade-off between the latency for 1<sup>st</sup> packet transmission and the UE power consumption. This can be done based on the traffic pattern of the UE. Such configuration typically leads to a DRX cycle that is less than 1 second unlike for the mMTC device which could be in the order of hours or even days.

## V. CONCLUSION

In this paper, we discussed the need for a configurable and flexible state for inactive 5G UEs. We take the three-state model that is currently under discussion as a baseline. We argue that the RRC Inactive state should be configurable in order to address 5G use cases having diverse and sometimes contradictory requirements. The behavior of a UE during RRC Inactive is configured such that its service requirements are fulfilled. This avoids the alternative approach

of increasing the number of states which would be complex to implement and would increase the standardization effort. This approach is future proof and reduces the time to market for new use cases.

## ACKNOWLEDGMENT

Part of this work has been performed in the framework of the H2020 project METIS-II co-funded by the EU. This information reflects the consortium's view, but the consortium is not liable for any use that may be made of any of the information contained therein. The authors would like to acknowledge the contributions of their colleagues from Nokia, Icaro Da Silva, and Gunnar Mildh.

## REFERENCES

- [1] 3GPP, "Interlayer procedures in Connected Mode", TS 25.303, Jan. 2016
- [2] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," TS 36.331, Oct. 2016.
- [3] 3GPP, "LTE Radio Access Network (RAN) enhancements for diverse data applications (Release 11)", TR 36.822, Sep. 2012.
- [4] J. Puttonen, E. Virtej, I. Kesitalo, and E. Malkamaki, "On LTE performance trade-off between connected and idle states with always-on type applications," IEEE PIMRC, Sep. 2012.
- [5] Z. Zhang, Z. Zhao, H. Guan, D. Miao, and Z. Tan, "Study of signaling overhead caused by keep-alive messages in LTE network," IEEE VTC Fall, pp. 1-6, Sep. 2013.
- [6] T. Taleb, and A. Kunz, "Machine type communications in 3GPP networks: potential, challenges, and solutions," IEEE Communications Magazine, vol 50, no 3, pp. 178-184, Mar. 2012.
- [7] Ericsson, "Signaling reduction for UE state transitions", 3GPP SA2 WG, Tdoc S2-152207, June 2015.
- [8] I. Da Silva, G. Mildh, M. Säily, S. Hailu, "A novel state model for 5G radio access networks," IEEE ICC Workshops, pp. 632-637, May 2016.
- [9] Nokia, "Discussion of RRC States in NR", 3GPP R2 WG, Tdoc R2-163441, May 2016
- [10] P. Marsh, O. Queseth et al., "5G radio access network design – a brief overview on the 5G-PPP Project METIS-II", EuCNC, Jul. 2015.
- [11] NGMN Alliance, "5G White Paper", Feb. 2015
- [12] I. Da Silva, G. Mildh et al., "Tight integration of new 5G air interface and LTE to fulfill 5G requirements," IEEE VTC, pp. 1-5, May 2015.
- [13] M. Johnsson, J. Sachs, T. Rinta-aho, T. Jokikyynty, "Ambient networks – a framework for multi-access control in heterogeneous networks", IEEE VTC, Sep. 2006.
- [14] 3GPP, "Study on Machine-Type Communications (MTC) and other mobile data applications communications enhancements," TR 23.887, Dec. 2013
- [15] S. Saur, A. Weber, G. Schreiber, "Radio access protocols and preamble design for machine type communications in 5G", IEEE ACSSC, pp. 8-12, Nov. 2015.
- [16] Nokia, "Evaluating DRX concepts for E-UTRAN," 3GPP WG R2, Tdoc R2-071284, Mar. 2007
- [17] 3GPP, "Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN)", TR 25.912, Sep. 2014

# Enhancing Peer-to-Peer Communications for Future Wireless Systems

Ugljesa Urošević  
University of Montenegro  
Faculty of Electrical Engineering  
Podgorica, Montenegro  
ugljesa@ac.me

Zoran Veljović  
University of Montenegro  
Faculty of Electrical Engineering  
Podgorica, Montenegro  
veljovic@ac.me

Ramjee Prasad  
Aalborg University  
Center for TeleInfrastructure  
Aalborg, Denmark  
prasad@es.aau.dk

**Abstract**—The future generation of wireless networks is expected to support a significantly large amount of mobile data traffic, massive number of wireless connections and devices, achieve better cost, increased energy and spectral efficiency, improved quality of service (QoS) in terms of communication delay, capacity, reliability and security. The main techniques that will enable these features are massive multiple-input, multiple-output (MIMO), utilization of higher frequencies, particularly millimeter-wave (mmWave) frequencies, super dense and deployment of cells, peer-to-peer (P2P) communications, heterogeneous network (HetNet) implementation, software defined wireless network (SDWN), etc. Here we present solutions for improving P2P communications through distributed MIMO systems and SDWN control.

**Keywords**—P2P, MIMO, 5G

## I. INTRODUCTION

The exponential growth of wireless data services driven by mobile Internet and smart devices has motivated the investigation of the 5G cellular network. Future networks will have to support multimedia applications with a wide variety of requirements, including x10-100 increased user data rates, end-to-end latency smaller than 5ms, enhanced indoor coverage, increased spectral and energy efficiency and so on. [1],[2].

There are many various promising enablers for 5G wireless communication systems, such as massive multiple-input multiple-output (massive-MIMO) and millimeter-wave (mm-wave) communications, spectrum and energy efficient communications, cognitive radio networks, visible light communications, densification of existing cellular networks with the massive addition of small cells and a provision for peer-to-peer (P2P) communication like device-to-device (D2D), machine-to-machine (M2M), vehicle-to-vehicle (V2V)-enabled, multi-tier heterogeneous networks (HetNet), simultaneous transmission and reception (full-duplex communication), energy harvesting, cloud-based radio access network (C-RAN), software defined network (SDN), and virtualization of wireless resources, [3].

P2P communication refers to a radio technology that enables peers to communicate directly with each other, that is without routing the data paths through a network infrastructure.

P2P communication helps increase spectral efficiency, enhance user experience, and expand communication applications. In P2P communications, user data is directly transmitted between peers without routing through a cellular network and thus results in hop gain. Moreover, resources between P2P users can be reused, and this results in resource reuse gain. With the hop gain and resource reuse gain, wireless spectral efficiency and network throughput can be increased. Also, the communication system may collapse if core network facilities or access network devices are damaged. However, P2P communication makes it possible for communication terminals to set up ad hoc networks. If the wireless infrastructure is damaged or terminals are not covered by a wireless network, multi-hop P2P can be used for P2P communication or even access to cellular networks. In this way, the number of wireless applications can be expanded.

Potential applications of P2P include local service, emergency communication, public safety, proximity based games and social networking, advertisement for by passers, traffic control and safety, intelligent transport system, Internet of things (IoT) enhancement, etc., [4]-[6].

Here, we present solutions for improving P2P communications through distributed MIMO systems and SDWN control. The simulation results show benefits of this approach.

The remainder of this article is organized as follows. After describing P2P communications we propose scheme for improving P2P communications for future wireless systems. Finally, conclusions are drawn.

## II. P2P

P2Peer communication in licensed band is one of the key enablers toward a more innovative and cost-effective communication systems. A key motivation for D2D connectivity is the potential for operators to offload traffic from the core network. A major breakthrough was achieved when 3GPP (LTE-A release, 12 June 2012) agreed on starting a study item for D2D technology. In the past, cellular operators did not consider P2P communication as an approach to enhance the performance of network because the benefits are



limited to local communication services. However, as mobile applications based on proximity of mobile devices has become increasingly popular, cellular operators are considering introducing P2P communication into the networks. Implementation of P2P communication in the system introduces many benefits compared to the conventional infrastructure-based communication, such as improved spectrum and energy efficiency, increased overall system capacity, decreased traffic load, etc.

Potential applications of P2P include local service, emergency communication, public safety, proximity based games and social networking, advertisement for by passers, traffic control and safety, intelligent transport system, Internet of things (IoT) enhancement, etc., Fig. 1.

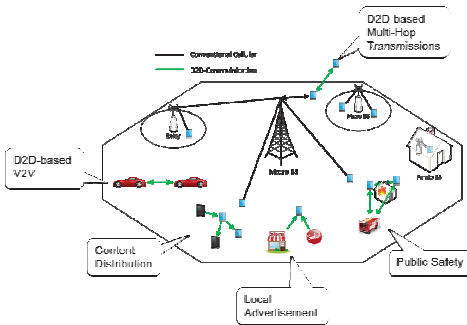


Fig. 1. Examples of a P2P communications

To fully realize these benefits in practice, the architecture for P2P communication should be flexible and powerful to meet the needs of commercial cellular scenarios as well as public safety applications. More demanding requirements are for public safety applications regarding high service reliability with ultra-low delays. Even when network infrastructure becomes overloaded or partially unavailable (e.g., in case of extraordinary events such as earthquakes, floods, fires, terrorist attacks, etc.), basic communication services should still be made available to public safety agencies, police rescue services and so on. Furthermore, these networks should incorporate mechanisms to seamlessly integrate with emerging technologies designed to further enhance public safety such as Wireless Sensor Networks (WSNs).

A typical application of P2P in intelligent transport system is V2V communication or Internet of Vehicles (IoV). When running at high speeds, a vehicle can warn nearby vehicles in P2P mode before it changes lanes or slows down (Fig. 2). After receiving warnings, nearby vehicles alert drivers or even automatically control the driving in an emergency situation so that drivers can react more quickly to reduce the number of traffic accidents. Furthermore, using P2P discovery technology, vehicles can reliably detect and identify specific vehicles nearby, such as those vehicles that may cause danger

at intersections and those specific vehicles (school buses or vehicles carrying dangerous goods) that need special attention.

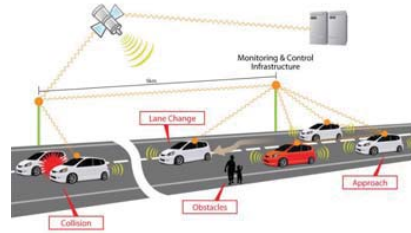


Fig. 2. IoV security - P2P communications

In local service, user data is directly transmitted between terminals and does not route through the network side. It's usually used for social applications. Social applications based on the proximity feature are a basic P2P application.

Another local service is local data transmission. E.g. local advertising service based on proximity can accurately target people to maximize its benefits. A shopping mall can send commercials, discounts and promotions to people who walk into or around the mall, a cinema can push movie information and show-times to people nearby.

Furthermore, there is cellular traffic offloading. Multimedia services like HD videos are becoming more popular so their large traffic flows put tremendous pressure on core networks and spectrum resources. P2P-based local media services can help operators save their core network and spectrum resources. In hotspot areas, operators or content providers can deploy media servers that store popular media services. These media servers deliver media services in P2P mode to users. Alternatively, users can use P2P to get the media content from nearby user terminals that have obtained media services. In this way, the downlink transmission pressure of operator cellular networks can be eased. Moreover, the cellular communication between short-distance users can be switched to the P2P mode to offload cellular traffic.

P2P can also be applied in more advanced scenarios, such as multiuser MIMO (MU-MIMO) enhancement, cooperative relaying, and virtual or distributed MIMO. E.g. in the MU-MIMO, BS determines pre-coding weights based on channel state feedback of terminals to create nulls and eliminate interference between users. With P2P introducing paired peers can directly exchange information about channel states, so terminals can feed the joint channel states information to BS and improve the performance of MU-MIMO.

### III. PROPOSED SOLUTIONS

In this paper, we propose SDWN architecture for improving P2P communication, which can meet the above-mentioned requirements of public safety applications as well as commercial applications such as proximity-based social networking (e.g., online gaming, video streaming), advertisements for by-passers, public safety (devices provide at

least local connectivity in the case of damage to the radio infrastructure), intelligent vehicle communication, efficient content distribution, etc.

The basic idea behind the proposed solution is to associate a P2P controller to a hierarchy of SDWN controllers in the network, in such a way to couple the formation and management of the mobile devices with the centralized control, resource allocation and routing features of SDWN. SDWN controllers can be placed locally and in hierarchical nature which makes the process of control scalable, energy efficient and robust to network infrastructure failure. In future wireless system, more effective technologies have to be employed for spectrum utilization, traffic control, resource allocation, density management, security, etc., so as to support the interconnection of more diversified user equipment (UE) and devices. The exploitation of SDWN technology can empower the network with intellectualization and reduce the complexity of 5G networks, decrease the cost on network deployment and maintenance, and facilitate the future network evolution. The current SDN standardizations have not taken the 5G RAN into consideration, so it is necessary to extend current SDN standards and propose novel SDWN architecture, Fig 3.

The spatial domain information management and complicated MIMO coordination problem can be overcome with SDWN controller that will manage spatial information and conduct massive MIMO coordination for future wireless networks. Nodes like base stations (BS), relay stations (RS), low power nodes (LPN), smartphones, can be extended to have a huge number of antennas, and SDWN controller can be responsible for the massive data processing constraints. E.g. BS can collect the user information and reports it to SDWN. SDWN controller performs the information processing like channel states analysis, weights, null-space calculation, etc. Then SDWN controller sends the processed results and instructions back to BS and LPNs so as to achieve better coordination regarding the global knowledge of network spatial information. SDN takes over the heavy operation from BS and LPNs to release the efficiency of 5G networks, [7]-[8].

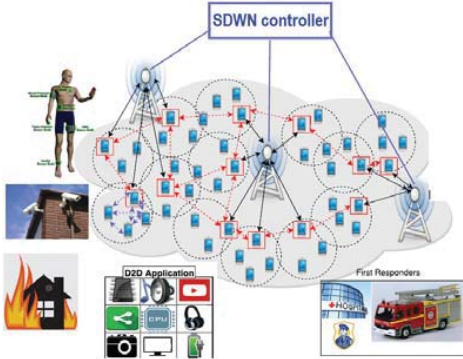


Fig. 3. Software-Defined P2P communication .

Furthermore, SDWN controller can decide which is the best path between two UEs, or which antennas should be used for considered connection, which MIMO technique is optimal, etc., based on many parameters like channel states, UEs battery powers, type of applications, priority, etc. The complicated spatial domain information management and coordination of massive MIMO can be facilitated with SDWN controller that will manage information and perform coordination.

Here, we analyse scenario where one UE transmits via two UEs, i.e. P2P relaying scenario, Fig. 4.

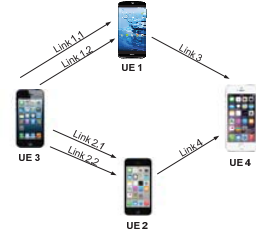


Fig. 4. P2P relaying scenario

It is assumed that UE3 uses two transmit antennas, while UE1, UE2 and UE3 use one antenna. Also, for this scenario it's assumed that Quasi Orthogonal Space-Time Block Code (QOSTBC) is used as a MIMO technique. We propose new solution for improving BER performance of the considered scenario. QOSTBC for system with four transmit antennas is given with, [9]-[10]:

$$\text{QOSTBC} = \frac{1}{\sqrt{4}} \begin{bmatrix} a_1 & -a_2^* & -a_3^* & a_4 \\ a_2 & a_1^* & -a_4^* & -a_3 \\ a_3 & -a_4^* & a_1^* & -a_2 \\ a_4 & a_3^* & a_2^* & a_1 \end{bmatrix} \quad (1)$$

Here,  $a_n, n=1, \dots, 4$ , are input symbols in QOSTB encoder. As it can be seen, for every four input symbols there are four output symbols at every antenna, i.e. the code rate is 1. This code can be virtually realized with this relaying, [11], [12]. The main disadvantage of QOSTBC is that there is no full orthogonality among symbols at output of the decoder, [13]. Here, we propose scheme for decreasing inter symbol interference at the output of QOSTBC decoder, i.e. for increasing orthogonality between output symbols.

For every four data symbols  $a_i, i=1, \dots, 4$ , UE3 generates two symbol sequences:

$$\text{UE3}_1 = \frac{1}{\sqrt{2}} [a_1 \quad -a_2^* \quad -a_3^* \quad a_4] \quad (2)$$

$$\text{UE3}_2 = \frac{1}{\sqrt{2}} [a_2 \quad a_1^* \quad -a_4^* \quad -a_3] \quad (3)$$

The average power spent per data symbol is fixed with scaling factor of  $1/\sqrt{2}$ . It is assumed that all the symbols in a single sequence suffer from same fading condition.

The channel coefficient between the  $i$ -th,  $i=1,2$ , UE and the  $j$ -th,  $j=1,2$ , antenna at UE3 is  $h_{ij}$ . The channel coefficients on the links UE1 – UE4 and UE2-UE4 are  $h_3$  and  $h_4$ , respectively. Received symbol sequences at  $i$ -th,  $i=1, 2$ , UE is:

$$\mathbf{UE}_{in} = \sum_{j=1}^2 \mathbf{UE3}_j \cdot h_{ij} + \mathbf{N}_{R_i} \quad (4)$$

$\mathbf{N}_i$  is additive white Gaussian noise (AWGN) matrix at  $i$ -th,  $i=1,2$ , UE:

$$\mathbf{N}_{UE_i} = [n_{UE_i,1} \ n_{UE_i,2} \ n_{UE_i,3} \ n_{UE_i,4}] \quad (5)$$

We propose new coding procedure with aim to restore orthogonality between symbols, and to keep code rate of 1 at the same time. It means that UE1 and UE2 generate next sequences:

$$\mathbf{UE1}_{out} = \frac{1}{\sqrt{2}} \mathbf{UE1}_{in} + \begin{bmatrix} R_1 & R_2 & R_3 & R_4 \\ 2h_3 & 2h_3 & 2h_3 & 2h_3 \end{bmatrix} \quad (6)$$

$$\mathbf{UE2}_{out} = \frac{1}{\sqrt{2}} \left[ -\mathbf{UE2}_{in}^*(3) + \frac{R_1}{2h_4} \quad -\mathbf{UE2}_{in}^*(4) + \frac{R_2}{2h_4} \right] \quad (7)$$

$$\mathbf{UE2}_{in}^*(1) + \frac{R_3}{2h_4} \quad \mathbf{UE2}_{in}^*(2) + \frac{R_4}{2h_4} \left. \vphantom{\frac{1}{\sqrt{2}}}\right]$$

where SDWN controller calculates and forwards  $R_1, R_2, R_3$  and  $R_4$  to UE1 and UE2:

$$\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & H_3 & H_4 \\ H_2^* & -H_1^* & H_4^* & -H_3^* \\ H_3^* & H_4^* & -H_1^* & -H_2^* \\ H_4 & -H_3 & -H_2 & H_1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} -k \cdot A_4 \\ k \cdot A_3 \\ k \cdot A_2 \\ -k \cdot A_1 \end{bmatrix} \quad (8)$$

$$A_i = \frac{1}{2} \left( \frac{\mathbf{UE1}_{in}(1) \cdot h_{11}^* + \mathbf{UE1}_{in}(2) \cdot h_{12}}{|h_{11}|^2 + |h_{12}|^2} + \frac{\mathbf{UE2}_{in}(1) \cdot h_{21}^* + \mathbf{UE2}_{in}(2) \cdot h_{22}}{|h_{21}|^2 + |h_{22}|^2} \right) \quad (9)$$

$$A_2 = \frac{1}{2} \left( \frac{\mathbf{UE1}_{in}(1) \cdot h_{12}^* - \mathbf{UE1}_{in}(2) \cdot h_{11}}{|h_{11}|^2 + |h_{12}|^2} + \frac{\mathbf{UE2}_{in}(1) \cdot h_{22}^* - \mathbf{UE2}_{in}(2) \cdot h_{21}}{|h_{21}|^2 + |h_{22}|^2} \right) \quad (10)$$

$$A_3 = -\frac{1}{2} \left( \frac{\mathbf{UE1}_{in}(3) \cdot h_{11} + \mathbf{UE1}_{in}(4) \cdot h_{12}^*}{|h_{11}|^2 + |h_{12}|^2} + \frac{\mathbf{UE2}_{in}(3) \cdot h_{21} + \mathbf{UE2}_{in}(4) \cdot h_{22}^*}{|h_{21}|^2 + |h_{22}|^2} \right) \quad (11)$$

$$A_4 = \frac{1}{2} \left( \frac{-\mathbf{UE1}_{in}(3) \cdot h_{12} + \mathbf{UE1}_{in}(4) \cdot h_{11}^*}{|h_{11}|^2 + |h_{12}|^2} + \frac{-\mathbf{UE2}_{in}(3) \cdot h_{22} + \mathbf{UE2}_{in}(4) \cdot h_{21}^*}{|h_{21}|^2 + |h_{22}|^2} \right) \quad (12)$$

$$k = \frac{2 \operatorname{Re}\{H_1 H_4^* - H_2 H_3^*\}}{H^2} \quad (13)$$

$$H^2 = |H_1|^2 + |H_2|^2 + |H_3|^2 + |H_4|^2 \quad (14)$$

$$H_1 = h_{11}h_3 \quad (15)$$

$$H_2 = h_{12}h_3 \quad (16)$$

$$H_3 = h_{21}h_4 \quad (17)$$

$$H_4 = h_{22}h_4 \quad (18)$$

The received symbol sequence at UE4 is:

$$\mathbf{UE4}_{in} = \mathbf{UE1}_{out} \cdot h_3 + \mathbf{UE2}_{out} \cdot h_4 + \mathbf{N}_{UE4} \quad (19)$$

$\mathbf{N}_{UE4}$  is AWGN matrix at MS:

$$\mathbf{N}_{UE4} = [n_{UE4,1} \ n_{UE4,2} \ n_{UE4,3} \ n_{UE4,4}] \quad (20)$$

With this relaying virtual 4x1 MISO channel is created between UE3 and UE4, i.e.:

$$\mathbf{UE4}_{in} = \frac{1}{\sqrt{4}} \begin{bmatrix} a_1 & -a_2^* & -a_3^* & a_4 \\ a_2 & a_1^* & -a_4^* & -a_3 \\ a_3 & -a_4^* & a_1^* & -a_2 \\ a_4 & a_3^* & a_2^* & a_1 \end{bmatrix}^T \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ H_4 \end{bmatrix} + \mathbf{N} \quad (21)$$

With the proposed coding the output symbols after QOSTB decoding are given with:

$$\begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \tilde{a}_3 \\ \tilde{a}_4 \end{bmatrix} = H^2 \left( \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} k(a_4 - A_4) \\ k(-a_3 + A_3) \\ k(-a_2 + A_2) \\ k(a_1 - A_1) \end{bmatrix} \right) + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix} \quad (22)$$

As it can be seen from (22) there is still no full orthogonality between output symbols, but it is increased in comparison with virtual QOSTBC. The simulation results will confirm this improvement.

It should be noted that considered scenario is not massive MIMO, but the same principle can be applied in the scenario with more antennas, i.e. for larger antennas arrays. For these QOSTBC codes the number of received antennas can be arbitrary, and it doesn't have impact on proposed solution. Of course, the more received antennas would decrease BER. Furthermore, extension of transmit antennas also would be quite straightforward.

IV. SIMULATION RESULTS

In this section the proposed scheme and the virtual QOSTBC are compared. The simulation results are presented under assumption that the propagation conditions are defined with Rayleigh fading statistics. Also, perfect estimation of channel coefficients and no direct link between UE1 and UE4 are assumed.

Improvements can be seen from Fig. 5. For example, in the case of QPSK modulation and BER value of 10<sup>-5</sup> we can see that for the proposed scheme the required SNR is smaller for 1.8dB than in the case of virtual QOSTBC.

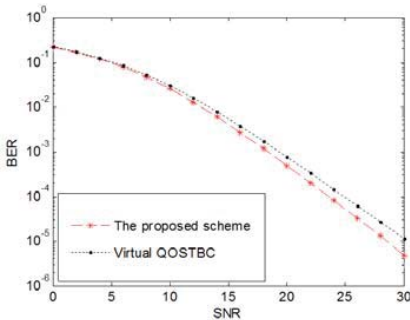


Fig. 5. BER performance of the proposed scheme and of the virtual QOSTBC, QPSK modulation

V. CONCLUSIONS

The main techniques that can meet demanding requirements of future wireless communication systems are massive MIMO, transmission on mmWave frequencies, i.e. utilization of higher frequencies, deployment of super dense, P2P communications, HetNet implementation, SDWN, etc. In this paper we present solution for improving P2P communications through distributed MIMO systems and SDWN control. New approach for improving performances of virtual MIMO communications between peers is presented, but the main contribution is that proposed solution can be extended for the scenario with large antenna arrays. The complicated spatial domain information management and coordination of massive MIMO can be facilitated with SDWN controller that will manage information and conduct coordination for future wireless networks.

REFERENCES

- [1] E. Hossain, M. Hasan, "5G Cellular: Key Enabling Technologies and Research Challenges," *IEEE Instrumentation & Measurement Magazine*, Volume 18, Issue 3, pp. 11 – 21, June 2015.
- [2] J. Guey, P. Liao, Y. Chen, A. Hsu, C. Hwang, G. Lin, "On 5G Radio Access Architecture and Technology," *IEEE Wireless Communication*, Volume 22, Issue 5, pp. 2 – 5, October 2015.
- [3] D. Soldani, A. Manzalini, "Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society," *IEEE Vehicular Technology Magazine*, Volume 10, Issue 1, pp. 32 – 42, March 2015.
- [4] K. Choi, Z. Han, "Device-to-Device Discovery for Proximity-Based Service in LTE-Advanced System," *IEEE Journal on Selected Areas in Communications*, Volume 33, Issue 1, pp. 55 – 66, January 2015.
- [5] M. Jo, T. Maksymyuk, B. Strykhalyyuk, C. Cho, "Device-to-device-based Heterogeneous Radio Access Network Architecture for Mobile Cloud Computing," *IEEE Wireless Communications*, Volume 22, Issue 3, pp. 50 – 58, June 2015.
- [6] W. Zhong, Y. Fang, S. Jin, K. Wong, "Joint Resource Allocation for Device-to-Device Communications Underlying Uplink MIMO Cellular Networks," *IEEE Journal on Selected Areas in Communications*, Volume 33, Issue 1, pp. 41 – 54, January 2015.
- [7] S. Sun, B. Rong, R. Qingyang Hu, Y. Qian, "Spatial Domain Management and Massive MIMO Coordination in 5G SDN," *IEEE Access*, Volume 3, pp. 2238 – 2251, November 2015.
- [8] M. Abolhasan, J. Lipman, W. Ni, B. Hagelstein, "Software-Defined Wireless Networking: Centralized, Distributed, or Hybrid?," *IEEE Network*, Volume 29, Issue 4, pp. 32 – 38, July-August 2015.
- [9] U. Urosecvic, Z. Veljovic, M. Pejanovic-Djurisic, "MIMO Solution for Performance Improvements of OFDM-CDMA System with Pilot Tone," *Wireless Networks*, Volume 19, Issue 8, pp. 2021-2028, Springer, November 2013.
- [10] V. Kühn, *Wireless Communications over MIMO Channels - Applications to CDMA and Multiple Antenna Systems*, pp. 283–288, Wiley, 2006.
- [11] T. Yamaoka, Y. Hara, N. Fukui, H. Kuboand T. Yamazato, "A Simple Cooperative Relaying with Alamouti Coded Transmission," *IEICE Trans. Commun.*, Vol. E95–B, No.2, pp. 643-646, February 2012.
- [12] U. Urosecvic, Z. Veljovic, M. Pejanovic-Djurisic, "A New Solution for Simple Cooperative Relaying", *Wireless Personal Communications*, Vol. 75, Issue 2, pp. 1235-1250, Springer, March 2014.
- [13] V. Tarokh, H. Jafarkhani, A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456-1467, July 1999.

# *A Survey on Live Virtual Machine Migration*

<sup>1</sup>Arsch Sharma, <sup>2</sup>Ashu Saxena, <sup>3</sup>Karthick Nanmaran  
School of Computing, SRM University, India

{<sup>1</sup>arschsharma\_vimal@srmuniv.edu.in, <sup>2</sup>ashu\_saxena@srmuniv.edu.in, <sup>3</sup>karthicknanmaran@gmail.com}

**Abstract**— Virtual Machine Migration has become a necessity for the efficient performance of computer networks. It not only assists in network load balancing, but also provides other benefits such as energy saving and efficient utilization of network bandwidth, to name a few. This survey focuses on the various techniques used for Virtual Machine Migration, and covers the traditional ‘pre-copy’ approach along with its variants as well as some of the recently developed alternatives to it.

**Keywords**— Software Defined networking, pre copy, Virtual Machine, migration, load balancing

## I. INTRODUCTION

The extensive usage and widespread of mobile devices has led to the rapid expansion of computer networks. The rise of programmable networking has also led to another level of virtualization, where an entire machine handling traffic can be virtualized upon a host node. This machine is called a Virtual Machine. The virtual Machine operates using the resources of the host system in such a way that it seems that it is an independent machine in itself, utilizing its own resources. This survey focusses on VM migration, a technique which has proven to be immensely useful when it comes to handling the modern day networks. This process has a number of advantages. Due to the migration of the VM across the network to suitable nodes, we can manage the network traffic better, as migrating the VM to nodes with higher traffic leads to better workload sharing. Also, the nodes with very less traffic can be switched off, or suspended for a while, until they are required, hence saving energy and other useful resources.

This paper has been organized into three sections. Section I is the introduction. Section II embodies the various techniques already in use for Virtual Machine Migration, Section III contains the challenges in the field, while Section IV is the concluding remark of the authors.

## II. TECHNIQUES USED FOR VM MIGRATION

Traditionally, the steps involved in Virtual Machine Migration involve the following steps:

1. Resource reservation: Resources at the destination are reserved.
2. Iterative pre-copy
3. Stop and copy
4. Commitment and Activation: After the target node indicates that it has successfully received a consistent copy of the virtual machine, resources are re-attached to the virtual machine.

Other basic steps which are required include:

- Establish layer-2 connectivity between data centres, if needed.
- If storage is not shared, transfer the application’s disk state.
- Transfer the memory state of the application to a server in the target data centre, as it continues running without interruption.

- Once disk and memory states have been transferred, pause the application briefly for final transition of memory and processor state to the target data centre

This approach can be divided into two major parts:

### 1. Disk State Migration

In the disk state migration, initial copy the disk contents of the VM is made asynchronously, which is followed by the process of synchronous replication, once the remote disk is stable. After this, the disk contents are continuously propagated across the network from the source to the target node.

### 2. Transfer of Memory State

The traditional VM migration approach makes use of the pre-copy mechanism, which iteratively copies the memory contents. After the majority of the memory contents are copied (as is the case usually after a few initial iterations), the VM at the source node is paused to copy its final memory state. The time required to do this depends upon the RAM allocation of the Virtual Machine, the Working Set Size (WSS), write rate, and the available network bandwidth.

The traditional approach to memory transfer via the pre-copy method involves sending the memory pages to the target node iteratively. These iterations can continue until either very few pages are left to be transferred, or already about 3 times of the VM’s total memory has been sent, or when a set amount of iterations has already been performed. After this, the remaining ‘dirty’ pages along with the final state of the virtual machine are sent over to the target node, and the copy of the VM running at the source node is destroyed.

### A. The Pre-Copy Mechanism

As already mentioned, pre-copy mechanism is the traditional method used in Virtual Machine migration. In this approach, the memory pages are sent before the processor state of the VM is sent to the target node. The memory pages of the VM are sent over the network in the form of numerous iterations, even as the VM continues to execute at the source node, hence the migration of the VM is live in nature. This is understandable as the processor state keeps changing until the actual point of complete transfer from the source node, at which the VM is paused at the source and the migration to the target node takes place. The iterations continue till either a small writable working set (WWS) is identified, or a set number of iterations has been reached. Any page that is ‘dirty’ in the process is retransmitted over the next iteration.

The identification the suitable WWS, or reaching the cap on number of iterations (whichever occurs first), marks the end of memory transfer, and the beginning of the service downtime. The VM at the source is then suspended, and the remaining dirty pages are sent to the target node along with these dirty pages.

The VM is restarted at the target node, and the copy of the VM at the source is destroyed. The total time required for the complete migration of the VM is commonly termed as the total migration time.

The main goal of the pre-copy mechanism is to minimize the amount of VM state to be transferred during the service downtime.

While this mechanism might seem simple enough for a single VM migration process, in case of multiple migrations, the total migration time becomes much longer than what is required. The time spent during the VM migration during the 'resource reservation' and 'commitment and activation' stages is relatively stable as compared to the other stages, and these are often referred to as the 'Pre-migration overhead' and 'Post-migration overhead'. The time taken for 'pre-copy' and 'stop and copy' are largely variable.

For the 'pre-copy' stage, the time taken for the transfer of all memory pages is roughly given by the expression:

$$t_0 = V/B, \text{ where } V = \text{the total memory size of the VM and } B = \text{available bandwidth}$$

To make allowance for the dirtied memory pages from the previous iteration, the time also to be considered is given by the expression:

$$t_1 = R^* t_0 / B, t_2 = R^* t_1 / B, t_3 = R^* t_2 / B, \dots, t_k = R^* t_{k-1} / B$$

Where  $R$  refers to the memory dirty rate. The pre-copy process is suspended when the  $k^{\text{th}}$  iteration occurs and the stop conditions are met.

After this, the VM at the source is suspended, and the remaining dirty pages, along with the processor state of the VM are transferred to the target node. This takes approximately

$$t_k = R^* t_{k-1} / B \text{ amount of time.}$$

Hence, the total time for the pre-copy and 'stop and copy' is computed as:

$$T_0 = t_0 + t_1 + \dots + t_k \\ = (V/B) * \{[1-(R/B)^{k+1}]/[1-(R/B)]\}$$

Hence, the total migration time can be computed as:

$$T_{\text{mig}} = \text{Pre-migration overhead} + T_0 + \text{Post-migration overhead}$$

#### Limitations of the pre-copy mechanism

Pre-copy mechanism suffers from a few limitations that do not make it perfect to be used in all the scenarios where VM migration is involved.

- The pre-copy mechanism introduces a limit on the number of iterations since it is not guaranteed that a small enough writable working set shall be achieved even after many iterations. This implies that if a small enough WWS is not converged to after the limit of iterations, the service downtime shall increase. This effect might not be very prominent in read-intensive workloads, but while handling even moderately write-intensive workloads can reduce the pre-copy mechanism's effectiveness.
- Also, another limitation stems from the fact that in the pre-copy mechanism, all the memory pages are transferred first, and then the pages modified during the last round are transferred iteratively. This causes a migration daemon to continually consume the network bandwidth to transfer the dirty pages in each round. In order to tackle this, an adaptive rate limiting approach has also been proposed, but it does more harm than benefit as the total migration time increases to tenfold the original time.
- Also, the maximum number of iterations must be set, as obtaining a suitably sized WWS is not guaranteed.

#### Proposals for making the pre-copy approach more efficient

- Checkpoint/Recovery and Trace/Replay (CR/TR motion)
- Another variant that has come forward is to transfer all the memory pages only once during the entire VM migration process. By doing so, a baseline migration time is achieved. This is a tasteless approach as latency rates higher than the original pre-

copy mechanism. It takes time to fetch pages from the source node.

- Smart stop and copy: This method employs a heuristic approach to make the process determine when to stop the iterations during memory transfers.
- Using page deltas.
- Adaptive Memory Compression

#### B. Pre-copy with Adaptive Memory Compression

This technique mainly focusses on Virtual Machine migration within a cluster environment. The bulk idea is to compress the memory pages before they are sent to the target node, and decompress them upon reaching the destination. Development of multi-core processors has led to copious amounts of CPU resources to be available for the virtualization process. These resources can be put to use to compress page frames by causing very little memory overhead, also, the process of decompression is very fast, and requires little or no memory.

The entire idea of Adaptive Memory Compression is to optimize the pre-copy mechanism, as due to compression of the memory pages to be sent, the VM migration time is substantially reduced, and efficiency of the process increases. For this purpose, a zero-aware *Characteristic Based Compression* (CBC) algorithm has been proposed, which is well-suited for live migration.

#### Advantages of compression

- Augments network bandwidth for migration.
- Compressed dirty pages take shorter time to fly over the network
- Network traffic drops dramatically, as there is less data to be transferred between the source and the target nodes.

#### The Characteristic Based Compression (CBC) algorithm

Some features of this algorithm are listed as:

- It is lossless: The compressed data is reconstructed at the destination node
- Overhead of compression is very small.

An overly simple algorithm with very less overhead shall be less effective. Hence, to be effective, some amount of definition and complexity is unavoidable. We must ensure that the overhead caused by the algorithm does not exceed its benefit.

#### The Explanation of the CBC algorithm

##### Notations:

$R_{\text{page}}$ : Average growth rate of dirty pages in the migrated VM  
 $R_{\text{tran}}$ : Average page transfer rate. It denotes the available network bandwidth

$R_{\text{cpr}}$ : Average memory compression rate

$\rho_{\text{cpr}}$ : Average compression ratio

$V_{\text{thd}}$ : Threshold of dirty pages where pre-copy process is stopped

$V_{\text{ms}}$ : Total memory size of virtual machine

**For MECOM (Memory Compression) approach:**

Transferred data size is represented by the vector  $V = \langle v_0, v_1, v_2, \dots, v_{n-1} \rangle$

Elapsed time in all rounds is represented by vector  $T = \langle t_0, t_1, t_2, \dots, t_{n-1} \rangle$

##### For XEN

Correspondingly, we have  $V' = \langle v'_0, v'_1, \dots, v'_{n-1} \rangle$

$T' = \langle t'_0, t'_1, \dots, t'_{n-1} \rangle$

Now, let  $\lambda = 1 - \rho_{\text{cpr}}$

Then,  $t_0 = \{(V_{\text{ms}} \cdot \lambda) / R_{\text{tran}}\} + (V_{\text{ms}} / R_{\text{cpr}})$ ,  $t'_0 = V_{\text{ms}} / R_{\text{tran}}$ ,  $t_1 = \{(R_{\text{page}} \cdot t_0) / R_{\text{tran}}\} + (R_{\text{page}} \cdot t_0 / R_{\text{cpr}})$ ,

$t'_1 = (R_{\text{page}} \cdot t'_0) / R_{\text{tran}}$  ...

$t_n = \{(R_{\text{page}} \cdot \lambda \cdot t_{n-1}) / R_{\text{tran}}\} + (R_{\text{page}} \cdot t_{n-1} / R_{\text{cpr}})$ ,

$$t_m' = (R_{page} \cdot t_{m-1}) / R_{tran} \dots\dots(I)$$

and,  
 $v_0 = \lambda V_{tms}$ ,  $v_0' = V_{tms}$ ,  
 $v_1 = R_{page} \cdot \lambda \cdot t_0$ ,  $v_1' = R_{page} \cdot t_0$

Only when  $t_0 < t_0'$ , performance can be improved.  
 If  $t_0 < t_0'$ , then  $[R_{cpr} \cdot P_{cpr} > R_{tran}] \dots\dots(II)$

This shows that product of compression time and compression rate should be larger than the network bandwidth available for migration so that VM migration shows better performance.

**Phases of Compression**

There are two phases of compression:

- Modelling : It refers to finding regularities in the data to be compressed.
- Encoding : Construction of a more concise representation of the data.

We can classify the memory pages as:

- Pages with many zero bytes and sporadic zero bytes
- Pages with high similarity
- Pages with low similarity

The way in which these pages must be dealt with is very different. For the first type of pages, it is proposed to scan the entire page to record the non-zero bytes and the offset value.

For the second type, and algorithm that embodies strong similarities is needed.

For the third type, a more general algorithm is required.

Hence, it is evident that the CBC has the following parts:

- Determine the kind of pages being compressed. For pages with very high similarity, and high number of zero bytes, a dictionary is employed, that keeps track of frequently seen words. Also, to decrease the matching overhead, the dictionary is managed as a direct mapped cache, and LRU (Least Recently Used) algorithm is used as the replacement algorithm for the dictionary.
- Making a concrete choice of compression algorithm to be used based on the kind of pages. Flags are added to the compressed pages which help in decompressing them at the destination using suitable decompression algorithm.

**Performance Evaluation**

To have an idea of the extent to which this technique is effective, the results of a comparison between the performance of VM migration in various aspects, with and without MECOM (Memory Compression) is done.

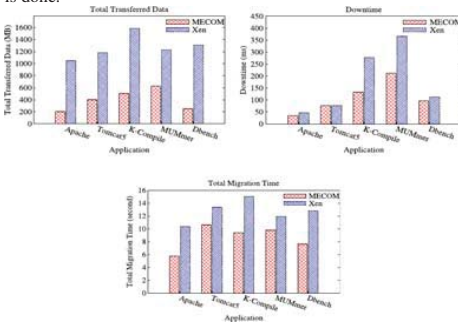


Fig 1 :Total transferred data, Downtime and total migration time of Xen and MECOM for diverse workloads

As is evident from the results, original pre-copy algorithm, which is used by Xen, is outperformed by the more efficient MECOM approach.

**C. Virtual Machine Migration using Checkpoint Recovery / Trace Replay Motion**

The Checkpoint Recovery/Trace Replay (CR/TR) Motion technique is quite a promising method, as experimental measurements have shown a reduction in migration overheads up to 72.4% on application observed downtime as compared to pre-copy.

Pre-copy is the most widely used mechanism for live VM migration. Both Vmotion and XenMotion adopt the pre-copy algorithm to reduce the downtime during the process. This has proven to be helpful, with the downtime reduced to the order of milliseconds, but there are some other issues that are not addressed by the pre-copy approach.

Some of these unsolved issues are:

- If in case of pre-copy, the rate at which the pages are dirtied is more than the replication rate, then the entire process turns out to be futile. In such a scenario, the pre-copy algorithm copies all the memory pages at once to the destination host. This is feasible only in case of high speed LANs.
- Use of some para-virtualization optimization techniques such as *stunning rogue processes* and *freeing unallocated pages* (used in XenMotion) may cause negative effect on user experience due to increased latency.
- The CPU cache is not recovered in the migration process. This causes a huge amount of degradation.

To tackle these problems, the CR/TR motion was proposed.

- The CR/TR motion implementation is done on a full system trace and replay system named ReVirt. It is ported on UMLinux, and is designed for intrusion detection. ReVirt logs only the asynchronous virtual interrupts. However, ReVirt has some overhead which is about 8% with regards to logging, and about 13 to 58% for virtualization in kernel intensive workloads. This is due to the fact that in kernel intensive workloads, mainly deterministic events occur.
- A trace daemon continuously logs non-deterministic events of the Vm while sacrificing very little performance.
- Execution trace files logged at the source host are iteratively transferred to the target host. These are used to synchronise the migrated VM's execution state.

Deterministic events refer to events like arithmetic, memory, or branch instructions. These do not need to be logged, as during the replay phase, they are executed again without any problem.

Non-deterministic events can be classified as time based and external input based. Only these events are logged during the CR/TR motion. The replay can be executed on any host with the same processor specifications as that of the source host, and no deviations occur during the replay process.

**Pre-requisites for the application of CR/TR motion**

CR/TR motion can be used only in case of certain situations. Following are the pre-requisites that must be fulfilled for the usage of this technique.

- The log transfer rate must be more than the log growth rate. This pre-requisite ensures that the source sends the log files to the target in a net beneficial way, and that the source is not

overwhelmed by a rate of log file production that exceeds the log file transfer rate.

- ii. Log replay rate must be greater than log growth rate. At the target host, the execution of replay process must be done in a way that there is net 'forward' movement.

**Phases of Live Virtual Migration using CR/TR Motion**

- **Initialisation:** Initialisation refers to choosing the appropriate target host to which the VM is to be migrated.
- **Reservation:** Reservation refers to asking the target host for the necessary resources for the VM migration
- **Checkpointing:** When the resources are reserved, and the actual migration process starts, the first thing that is done is briefly pausing the source VM, and recording its current state. This is called checkpointing.
- **Iterative log transferring:** The checkpoint file is sent to the target host, and this is followed by iterative transfer of log files. These log files contain the non-deterministic events of the source VM, and are used to carry out the replay processes at the target host.
- **Waiting and Chasing:** Waiting and chasing takes place at the target host. The log file size must converge to a small measure after some number of iterations. The final log file size must be small so that the downtime is minimized. Now, the waiting and chasing means that the source asks whether the target is ready to execute the replay process of the last log file, and if the target at that point of time has some log files to be replayed, the process has to wait, till they are done. Also, the target must execute the log files at a rate to chase the convergence of the log file size to the minimum.
- **Stop and Copy:** The last log file is transferred to the target host. Two exactly same VMs exist at source and the target. This helps in case of recovery, where the source VM is the backup in case of any failure.
- **Commitment:** This happens when the target host tells the source host that the synchronisation of both the VMs is complete. The source host acknowledges this, and the traffic is redirected to the migrated VM.
- **Service taking over:** The VM at the target host is activated and it advertises its new shifted IP address. The target host becomes the primary host, and the migration is complete.

**Implementation**

As mentioned, ReVirt is applied to implement live VM migration on top of UMLinux for x86 platform.

**Performance Evaluation**

The experiments use the following VM workloads:

- i. **Daily use:** an idle Linux OS for daily use.
- ii. **Kernel-build:** the complete Linux 2.4.18 kernel compilation is a system-call intensive workload, which is expensive to virtualization.
- iii. **Static web application:** we use the Apache 2.0.63 to measure static content web server performance.
- iv. **Dynamic web application:** a more challenging Apache workload is presented by SPECweb99, a complex application level benchmark for evaluating web servers and the systems that host them.
- v. **Unixbench:** it is a benchmark suite for Linux that integrates CPU, file I/O, process spawning and other workloads.

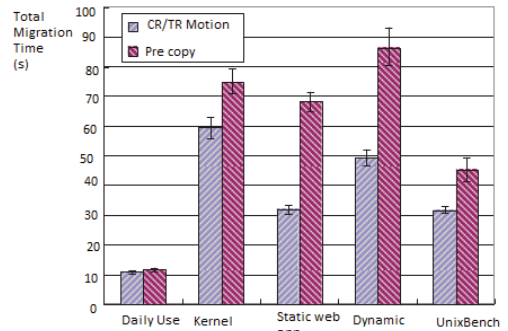
**Table 1. Time and space overhead of logging and replay**

Workloads	Log growth rate (KB/sec)	Replay Normalized rate to logging
daily use	10.3	36.8
kernel-build	2.2	1.05
static web app	247	1.63
dynamic web app	722	1.24
Unixbench	61.4	1.18

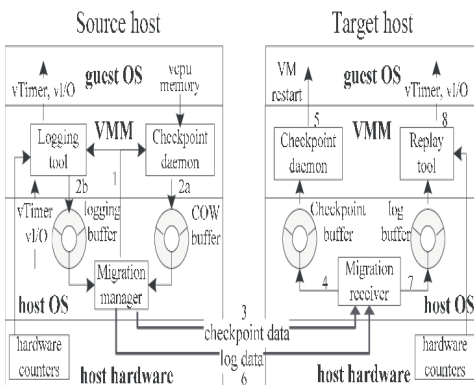
The above table shows the time and space overheads due to logging for various workloads. While the log growth rate for kernel intensive workloads (kernel-build) is very low, due to higher number of deterministic events, the same is higher for static web app workload due to logging of the number of incoming data packets.

**Comparison of Downtime and Migration Time of CR/TR motion with those of Pre-copy**

For comparing the downtime and the migration time of CR/TR motion with those of Pre-copy, the workloads are taken to a high speed LAN.



**Fig 3 :** The migration downtime of CR/TR-Motion and Pre-copy for different workloads in a LAN. The left bar labelled with "CM" and the right ones labelled with "PC" show the component parts of migration downtime for our CR/TR-Motion approach and Pre-copy approach respectively.



**Fig 2 :** CR/TR Motion System Structure



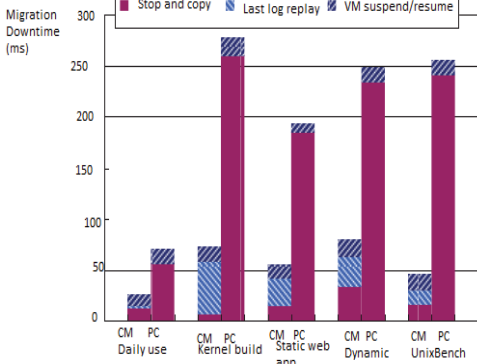


Fig 4: Total migration time of CR/TR-Motion and Pre-copy for different workloads in a LAN

**D. Post-Copy Mechanism using adaptive Pre-paging and Dynamic Self-Balloonig**

Unlike the pre-copy approach, where the memory state of the VM is transferred to the target node before the processor state, the post-copy mechanism makes use of the opposite sequence of events. The processor state is sent to the target node before the memory state. This technique uses adaptive pre-paging to eliminate duplicate page transmissions. It also prevents transfer of memory free pages using dynamic self-ballooning mechanism.

Post copy involves the VM to be started using the processor state sent to the target at the destination node, and then involves pushing the memory pages to the destination node. The advantage of this method over pre-copy is that any memory pages that do not reach the destination VM are demanded from the source, and hence each memory page is transferred only once. Pre-paging technique is used along with post-copy mechanism which involves the intelligent sequencing of the transferred pages in order to hide any effects of latency. However, both pre-copy and post copy mechanisms are not efficient enough to deal with the transmission of memory free pages. To counter this, Dynamic Self Ballooning (DSB) is employed. Ballooning is a common practice in virtualization, in which the guest OS minimizes its footprint by releasing free memory pages to the Hypervisor, or the Virtual Machine Monitor. Dynamic Self Ballooning refers to automatic self-ballooning at fixed intervals of time. As a result the migration time is reduced significantly by the elimination of free pages. Variations of post-copy include post copy with demand-paging, active push, pre-paging, and dynamic self-ballooning. Demand paging ensures that the memory pages are sent only once to the VM at the destination node unlike pre-copy where the memory pages are sent over and over again. Active push ensures that the dependency of the VM upon the source is removed as soon as possible. Pre-paging uses hints from the VM's page access patterns to reduce the network failures and the time required for the resume process. DSB reduces the number of pages transferred during the migration, hence reducing the total migration time.

**A Comparison of pre-copy and post copy VM migration**

- **Efficiency:** In ideal conditions, with no node failures, we can say that post copy performs better than pre-copy. In post copy, we could use either the active push or demand paging to ensure

that a memory page is transferred only once to the VM at the destination.

- **Reliability:** As far as failure of the source node is considered, both the mechanisms fail in the migration of the VM. However, the difference is made when the destination node fails. In this case, as post copy has the processor state of the VM residing at the destination, the entire VM is lost, whereas in case of pre-copy, the memory pages that are copied to the destination are destroyed. It is important to mention that all those sent pages are also present at the source node, and hence failure of the destination node doesn't cause the loss of VM in case of pre-copy.
- **Overhead:** The overheads in case of pre-copy are more than in the case of post copy. In pre-copy, the free memory pages are not managed, and are transferred, which increases the total migration time of the VM migration, whereas, in case of post-copy, usage of the DSB eliminates any free pages that can be transferred during the migration. Same goes for duplicate pages and active push or demand paging.

Pre copy has been the conventional method for VM migration due to its reliability, whereas post-copy is usually employed for process migration.

**E. Hybrid Virtual Machine Migration**

Hybrid Virtual Machine Migration is a cross between the pre-copy and the post copy techniques for Virtual Machine Migration. The process consists of five phases:

- Preparation phase:** The preparation phase consists of the source host reserving the necessary resources at the target host for the smooth running of the VM.
- Bounded pre-copy rounds phase:** In this part, the hybrid process has the characteristics of the pre-copy technique. As in pre-copy, memory pages are sent in iterations to the target host. However, there is a maximum number of iterations till which this iterative transfer continues.
- VM state transfer phase:** After the fixed number of memory page transfer iterations have been carried out, the Virtual Machine processor state is transferred to the target host. As usual, the VM at the source is paused, and then the processor state is transferred.
- VM resume phase:** This occurs at the target host, after the processor state of the VM is transferred.
- On demand paging phase:** Now the phase similar to post copy mechanism starts where the VM at the destination host demands the transfer of the remaining memory pages from the source. This method reduces the service downtime, and as we have seen, in case of on-demand paging, each memory page is transferred only once, and hence there is no problem of dirty pages as in pre-copy.

**Advantages**

- The major advantage of the hybrid VM migration technique against pre-copy approach is that of reduction in total migration time. Due to its partial post-copy nature, the benefits seen in post-copy technique are reaped.
- Against the post-copy algorithm, the hybrid VM migration method gives an added security in case of failures and crashes.

**III. CHALLENGES**

Virtual Machine migration techniques, though being developed at a substantial rate, face many challenges. With the increase in the number and size of networks, it has become all the more important to make networks more efficient. As mentioned earlier, the conventional physical networking devices and systems can't match up to the magnitude of the requirements in today's world. Hence, there is a need to shift to virtualization of resources as fast, and as efficiently as possible.

**IV. CONCLUSION**

surveyed the various techniques with variations used for live virtual machine migration, however, it was evident that however sophisticated, these mechanisms have faults that need to be dealt with. Hence, the need of the hour is to improve already existing mechanisms, or to develop new, better approaches for the task at hand.

A comparison of various VM migration techniques is tabulated as:

**Table 2: Comparison of various live VM migration techniques**

Pre copy	<ol style="list-style-type: none"> <li>1. Down time&lt;1 sec</li> <li>2. On aborting migration, the system doesn't crash as the machine is still running on the source host.</li> </ol>	Overhead of duplicate page transmission
Improved pre-copy approach (MECOM and delta compression)	<ol style="list-style-type: none"> <li>1. Less down time</li> <li>2. Migration throughput is increased</li> <li>3. ~34% total data transfer reduced</li> <li>4. ~32.5% total migration time is reduced</li> </ol>	Overhead of compression technique
Post copy	<ol style="list-style-type: none"> <li>1. Memory page is transferred at most once.</li> <li>2. Total migration time is achieved as baseline.</li> </ol>	Less secure in case of failure than pre copy. More downtime than pre-copy.

As we can see, each of the given techniques is good in certain scenarios, while it is inefficient in others. For single VM migration, pre copy might work good, but it might prove cumbersome in case of virtual cluster migration. Similarly, while post copy is efficient in reducing the total migration time, it increases the service downtime during the VM migration. Selection of a suitable compression algorithm is again crucial in case of improved pre-copy approach. If the pages to be transferred bear very little or no similarity at all, then the compression becomes very difficult. Hence it is important to analyse the situation and then choose the method for live VM migration wisely.

**REFERENCES**

1. Paul Hightower, "Motion Effects on GPS Receiver Time live virtual machine migration with adaptive compression-Hai Jin ,Li Deng ,Song Wu, Xanhua Shi, Xiadong Pan (Huazhong University of Science and Technology, Wuhan , China).
2. Virtual machine Migration - KM SabidurRahaman (UC Davis).
3. Virtual machine Migration-Micheal Nelson 2009.
4. Post copy based live virtual machine migration using adaptive pre-paging and dynamic self-ballooning - Micheal R. Hines and KartikGopalan(Computer Science, Binghamton University (State University of New York)).
5. Composing Software Defined Networks-ChristopherMonsato, Joshua Reich, Jennifer Rexford, David Walker(Princeton University), Nate Foster(Cornell University).
6. Coordinating Virtual Machine Migration in Enterprise Data Centers and Clouds- Haifeng Chen, Guofei Jiang, YuepingZhang(Nec laboratories

- America, Inc 4 Independence way, Princeton, NJ 08540) Hui Kang( SUNY Stony Brook University Stony Brook, NY 11794).
7. Virtual machine Migration Based on Greedy Algorithm in Cloud Computing- SomayerSoltanBaghshahi, Sam Jabbehdari, SahraAdabi( Department of Computer Engineering, Tehran north Branch, Islamic Azad University, Tehran, Iran).
8. Live VM migration via asynchronous replication and state synchronization-Hai Liu, HaiJin ,Xiaofie Liao, Chen Yu, Cheng-ZhongXu.
9. Performance and energy modelling for live migration of virtual machines- Haikun Liu, Chen-ZhongXu, Jiayu Gong, Xiaofie Liao.
10. Black box and Gray box strategies for virtual machine migration – T.Wood, PJ Shenoy,A.venkatramani, MS. Yousif.
11. Live wide-area migration of virtual machines Including Local persistent State-Robert Bradford, EvanglesKotsovinos, Arya Feldmann, Harold Schiobery [Deutsche Telekom Laboratories].
12. The efficacy of live virtual migration over the internet – SebastienGoasguen, James Westall [Clemson university].
13. Seamless live migration of virtual machine migration over the man/wan- Franco Travostino, Paul Daspit, Leon Grammans, Chetan Jog, Cees De loat, Joe Mambretti, InderMonga, Bas van Oudenoarde, SatishRaghunath, Phil Wang .
14. Composing softwarer – Defined Networks- Christopher Monsato, Joshua Reich, Nate Foster, Jennifer Rexford, David Walker.
15. Comparative study of Virtual Machine Migration Techniques and Challenges in Post Copy Live Virtual Machine Migration RajwinderSingh,Dr. K.S.Kahlon, Sarabjit Singh.
16. Evaluation of Delta Compression Techniques For Efficient Live Migration of Large Virtual Machines-PetterSvard, Benoit Hudzia, Johan Tordsson and Erick Elmroth.
17. Distributed Virtual machine Monitor for managing multiple virtual resources across multiple physical nodes-Scott H Davis, Jerry Plouffe, Benjamin J. Thomas, III, Bedford, MA (US); Steven S. Noyes, Sterling, MA (US); Tom Hazel, Andover, MA.
18. Virtual machine Monitors Scalable Multiprocessors -EdouardBugnion, Menlo Park; Scott W. Devine; Mendel Rosenblum, both of Stanford, all of Calif.
19. Packet forwardingOptimiation With virtual Machine Mobility-Ming Zhang, San Jose , CA(us) Chengelet V.Ramesh,SanJose,CA (US).
20. The Efficacy of Live Virtual Machine Migrations Over the Internet-Eric Harney, SebastienGoasguen, Jim Martin, Mike Murphy, Mike Westall.
21. Datacenter Migration and Implementation using VMware-David A. Smith.
22. Secure Virtual Network Configuration for Virtual Machine (VM) Protection-RamaswamyChandramouli.
23. Live migration of virtual machine-Christopher Clark, Keir Fraser, Steven Hand, JakobGorm Hansen†, Eric Jul†, Christian Limpach, Ian Pratt, Andrew Warfield.
24. Empirical Exploitation of Live virtual Machine Migration - Christopher Clark, Keir Fraser, Steven Hand, JakobGorm Hansen†, Eric Jul†, Christian Limpach, Ian Pratt, Andrew Warfield .
25. Fast Transparent Migration for Virtual Machines-Michael Nelson, Beng-Hong Lim, and Greg Hutchins.
26. Live Migration of Virtual Machines in Cloud -Ashima Agarwal, ShangruffRaina .
27. Virtual Machine Migration in Cloud Computing - Pankajdeep Kaur and Anita Rani .
28. A Survey on Virtual Machine Migration Techniques In Cloud Computing -Nitisichandra Vyas, Prof. Amit Chauhan .
29. A survey on virtual machine migration and server consolidation frameworks for cloud data centers- Raja Wasim Ahmad a.n , Abdullah Gani a , SitiHafizah Ab. Hamid a , Muhammad Shiraz a.
30. Survey on Virtual Machine Placement Techniques in Cloud Computing Environment -Rajeev Kumar Gupta and R. K. Pateriy

# Enhanced Energy Saving Mechanism for Multi-Radio Multi-Channel Wireless Mesh Networks

Samreen Umer

CTVR, Dept. of Computer Science,  
University College Cork, Ireland  
Email: samreen.umer@insight-centre.org

Kenneth N. Brown

CTVR, Dept. of Computer Science,  
University College Cork, Ireland  
Email: k.brown@cs.ucc.ie

Cormac J. Sreenan

CTVR, Dept. of Computer Science,  
University College Cork, Ireland  
Email: cjs@cs.ucc.ie

**Abstract**—Multi-radio multi-channel (MRMC) mesh networks improve latency and spectrum utilisation, but at the cost of increased energy consumption. The standard mesh networking power saving mechanism (802.11 PSM) does not apply to MRMC. We propose an enhanced energy saving mechanism, EESM, in which each node switches its radios between different energy states based on observed traffic. In an empirical evaluation, we investigate the tradeoff between energy savings and decreased goodput. We show that energy savings are possible without impacting goodput, and that more aggressive control can generate more significant energy savings.

## I. INTRODUCTION

With the increase in Internet of Things deployment and the move to 5G networks, the interest in multi-hop wireless mesh networking is growing (e.g. [4], [10], [15]). WMNs are also emerging as low cost access networks for developing countries (e.g. [5], [7]). A WMN consists of multiple nodes connected in a graph topology, including clients, routers and internet gateways. Nodes can communicate peer-to-peer, using multi-hop paths through the network, with acceptable latency for delay-tolerant or low volume traffic. For higher traffic demands, if the radios are restricted to a single channel, or nodes have only a single radio, then congestion can occur and latency increases.

MRMC mesh networks use multiple radios on each node, with radios able to use different channels. This reduces interference between nodes, and allows a node to handle multiple flows simultaneously, thus reducing latency and increasing throughput. However, this introduces an additional energy cost in order to support the multiple radios.

Various power saving mechanisms for WMNs have been proposed and implemented. However a power saving mechanism for a MRMC WMN is still an open problem. In this paper we have proposed an enhanced energy saving mechanism (EESM) for such networks. EESM tries to put the maximum number of radios to sleep mode while maintaining the network efficiency. The proposed technique is compared in terms of energy consumption, delay and goodput with a network using no energy saving mechanism and conventional power saving mode. Experimental evaluation shows that EESM can save significant energy over the default power saving mode (PSM),

without degrading network performance for both constant bit rate flows and HTTP traffic, for low and medium traffic loads.

## II. RELATED WORK

Different approaches have been proposed to reduce energy consumption in single radio WMNs. IEEE 802.11 provides a power-saving mechanism for distributed coordination function (DCF) based single radio mesh WLANs. The basic idea of PSM is to put radios to sleep mode whenever they are in idle state for long periods. A radio in sleep mode consume less energy as compared to when in idle state. By putting more radios into sleep mode whenever feasible, a network can save some energy. PSM divides the time into identical beacon intervals which remains fixed with each interval comprising of a small ad hoc traffic indication map (ATIM) beacon exchange interval and larger data exchange interval. In each ATIM beacon exchange interval, radio on each node switches to a default channel to exchange ATIM beacons. Nodes which want to communicate with another exchange these beacons and then exchange data in data interval. Rest of the nodes which are not actively communicating switch their radios to sleep mode and stay asleep for the interval. The duration of this ATIM beacon exchange interval has great importance as a very small period would not be sufficient enough to receive all ATIM beacons causing delays and a very large period would not be very energy efficient. PSM is the standard approach for saving power but does not cater for MRMC. Basic implementation of PSM over MRMC WMN only considers state of a node putting all its radios in same mode or controls each radio individually as a single entity. Both approaches fail to take full advantage of having multiple radios per node for energy saving. Nodes end up keeping more radios awake even in a case where less radios could do the same job.

[1] proposed an adaptive ATIM beacon exchange interval for PSM but nodes need to propagate these changes among each other and still require tight time synchronization. Energy aware routing assigns routes on the basis of total energy consumption per route or remaining energy in battery operated nodes (e.g. [11], [12]). They have discussed energy efficient routing but did not consider the capability of nodes or radios

to sleep for saving overall energy. One proposal is to adapt the transmission power of the radios while satisfying throughput needs [3]. Having multiple radios operating on different power levels on same node are prone to self interference and shadowing etc. ([2], [6]) have reviewed common energy efficient protocols for network and data link layer respectively, and [8] has discussed energy saving protocols at routing layer for WMNs.

The available literature mainly focuses on single radio WMNs and only a few papers considered MRMC WMNs. [14] proposed an adapted PSM technique for MRMC networks. In their approach, nodes should maintain a table for their neighbouring nodes and respective common channels. One of the radios per each node switches to a default common channel in every ATIM exchange period. In ATIM exchange period, nodes which want to communicate with each other negotiate the channel and radio over common channel and then exchange data in data interval over negotiated link. Radios which are not involved in communication over this period are put into sleep mode. This scheme is applicable to MRMC networks where each radio is assigned a fixed channel and each radio acts as independent transceiver. In this method, a node can communicate with at most  $R$  number of neighbours in one interval where  $R$  is the number of radios on that node. Nodes will have less chances to put radios into sleep mode with higher number of traffic flows in the network.

### III. ENHANCED ENERGY SAVING MECHANISM (EESM)

EESM aims to take advantage of having multiple radios per node for energy saving. Nodes divide radios into receivers and transmitters. Receivers are assigned channels in fixed manner, the transmitters switch dynamically to the channels over which they can communicate to respective nodes. A single radio can transmit up to  $N$  number of neighbours where  $N$  is the number of available channels. In general more number of channels are available than the number of radios per node  $N \geq R$ . This allows less number of radios to communicate with more number of nodes thus permitting more radios to go to sleep in contrast to standard PSM [14].

Lets consider a mesh topology with  $M$  nodes and  $N$  orthogonal channels, where each node is equipped with  $R$  number of radios. Each node classifies one radio as a receiver  $R_r$  and others as transmitters  $R_t$ . Channels are assigned to receiver radios using game theoretic approach given in [9]. Each node stores a table for neighbouring nodes with their respective receiving channel. When there are multiple flows  $J$  going through a node, a virtual queue  $Q_j$  for each flow is maintained and assigned to a corresponding neighbour/channel. Node assigns active channels to the transmitter radios uniformly.

Buffer capacity of a node is  $L$  and queue occupancy of a node is given by

$$Q_{occupancy} = \sum_{j=1}^J Q_j \quad (1)$$

$$Q_{occupancy} < L \quad (2)$$

Time is divided into equal intervals  $T_{int}$  and within each interval a transmitter radio switches to respective channels for communicating with assigned neighbours accordingly. Different scheduling mechanisms can be used for switching transmitter channels. A simple fixed round robin scheme is explained below.

#### A. Fixed Round Robin Scheduling

From  $J$  flows, node assigns a list of  $n$  neighbours to a transmitter. Transmitter radio remains on receiving channel of first neighbour in  $n$  for a fixed time of  $T_{ch}$ , then switches to the receiving channel for the next neighbour in list  $n$  and so on. If there is only one neighbour assigned to a transmitter then during each  $T_{int}$  the transmitter will stay on the same channel and will not require switching. If more than one neighbour is being assigned to the transmitter then it will switch between  $n$  channels in each  $T_{int}$  for equal time  $T_{ch}$ .

$$T_{ch} = \frac{T_{int}}{n} \quad (3)$$

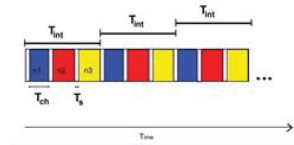


Fig. 1. Fixed Switching Schedule for a Transmitter Radio Over 3 Channels

Figure 1. shows an example where a transmitter radio has been assigned three active queues which have packets to transmit. Transmitter keeps switching between three channels and each interval is distributed evenly between queues. For more than one active queues per transmitter, each queue has to wait for time  $T_{del}$  before its served again.

$$T_{del} = (n - 1)T_{ch} + nT_s \quad (4)$$

$T_s$  is a very small switching delay which incurs for every channel switch and can be calculated as

$$T_s = t_{cs}(f - f') \quad (5)$$

$t_{cs}$  is a fixed switching unit and depends on the hardware of radio card,  $f$  and  $f'$  are frequencies of previous and new channels respectively.  $T_s$  is very small and negligible.  $T_{del}$  should be kept under a threshold level to avoid long delays at radios. A threshold value depends on type of traffic and its delay sensitivity. A smaller value allows radio switching only if resulting additional delays are very small for current flows. Larger value on the other hand put radios into sleep mode more often and results in higher queuing delays.

$$T_{del} \leq T_{th} \quad (6)$$

TABLE I  
PARAMETER NOTATIONS AND VALUES FOR EXPERIMENTS

Parameter	Notation	Value for Exp.
Number of nodes	M	16
Number of channels	N	11
Radios per node	R	3
Buffer size of a node	L	255 pkts
Transmitter radios per node	$R_t$	2
Receiver radios per node	$R_r$	1
Active flows per node	J	1-10
Active queues per transmitter	n	1-11
Time interval for one cycle	$T_{int}$	0.1s
Delay threshold	$T_{th}$	65,75,80 ms

### B. Objective

A radio can be turned ON or OFF at a node and a node can further switch the states of ON radio between awake or asleep mode. ON radio which is taking part in transmission or reception is in active state and consumes highest energy. ON radio which is not participating in any active communication is in idle state. When a radio is idle it is consuming less energy than active state and can switch to active state by itself whenever there is a demand to transmit or receive. In sleep mode, radio consumes the least energy but can not take part in any communication.

Each node defines a variable  $X_i$  to maintain the state of radio  $i$  on it.

$$X_i = \left\{ \begin{array}{l} 0 \quad \text{awake} \\ 1 \quad \text{asleep} \end{array} \right\} \quad (7)$$

The objective, as per below equation, is to save energy, while satisfying Eq(6).

$$Max \sum_{i=1}^R X_i \quad (8)$$

### C. State Transition

The state switching decision for a radio is made at each node in a distributed manner.

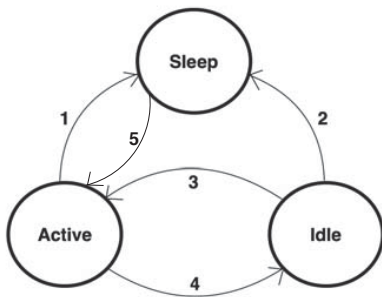


Fig. 2. State Diagram of a Radio Interface

- 1) Active  $\rightarrow$  Sleep: If by putting radio into sleep mode, constraint Eq(6) is not violated, the node will assign the active queues on that radio to another awake radio and put it into sleeping mode.
- 2) Idle  $\rightarrow$  Sleep: If a radio has been idle for a time period  $t$ , then the node will switch it to sleep mode. No active queues should be assigned to it as long as it is sleeping.
- 3) Idle  $\rightarrow$  Active: If there are packets in any assigned queue for a radio it will start transmitting.
- 4) Active  $\rightarrow$  Idle: If all assigned queues are empty for a radio, it will go to idle state.
- 5) Sleep  $\rightarrow$  Active: If Buffer is full ( $Q_{occupancy} = L$ ) causing incoming packets to drop or  $T_{del}$  is exceeding the ( $T_{th}$ ) with sleeping radios available then one or more sleeping radios are put back into active mode and assigned some active queues. This will make queueing delay smaller for each flow emptying buffer in a faster manner avoiding packet drops.

Each node maintains a table for its neighbours and their respective receiver channels. It also stores a transmitter table keeping information about channels assigned and current states of each transmitter. Each node can independently make a decision of switching states of transmitters and then transmitter table is updated accordingly. It is to be noted that this information is required on the node locally and does not effect the route of the flow. Each node tries to use least number of interfaces while satisfying the traffic demand and keeps other interfaces in sleep mode hence minimizing energy consumption.

## IV. RESULTS AND ANALYSIS

A number of experiments were carried out to evaluate EESM. A grid topology mesh network comprising of 16 nodes each equipped with three radios is simulated using network simulator ns3. A set of 11 orthogonal channels is available and channel assignment scheme [9] is used to minimize co-channel interference. Performance of EESM is compared with PSM and a network without any energy saving (WES) mechanism. The values and parameters used in experiments are given in Table I and II [13]. An extensive search was carried out to find suitable values for  $T_{th}$ .

TABLE II  
ENERGY PARAMETER VALUES

Energy Parameters	Value
Supply Voltage $V$	3 V
Idle Current of a radio $I_{idle}$	0.313 A
Transmitting Current of a radio $I_{tr}$	0.79 A
Receiving Current of a radio $I_r$	0.367 A
Switching Current of a radio $I_s$	0.0167 A
Sleeping Current of a radio $I_{sleep}$	0.096 A

Total running time for simulation is hundred seconds. Two different types of traffic flows are used. In first case ten constant bit rates (CBR) flows for peer-to-peer traffic are used. In second case ten HTTP flows are generated with mixed

peer-to-peer and gateway oriented traffic. CBR (UDP) flows are benchmark to evaluate the simulations and HTTP (TCP) flows reflect real traffic. Each experiment was run ten times with different seeds for generating random flows while keeping duration and load constant per flow. Performance is measured in terms of energy, delay and goodput. Energy efficiency is measured by dividing total energy consumed by sum of successful transmissions in the network. This gives energy used per successful bit transmission. A network having lower value for this metric is more energy efficient. Packet Delay is end-to-end delay a packet faced from source to destination over each flow. Goodput is calculated using successfully received packets in transmission duration for each flow.

**A. CBR Traffic using UDP**

In first set of experiments, ten CBR flows are generated using udp client and server in peer-to-peer fashion. Packet size is 1460 bytes, packet interval is 0.01s and, bandwidth is set to 1Mbps.

Figure 3 shows the mean values of energy consumed per node and energy efficiency of network for each scheme. Figure 4 gives the comparison for network performance parameters which are mean packet delay for a flow from source to destination and mean goodput across a flow.

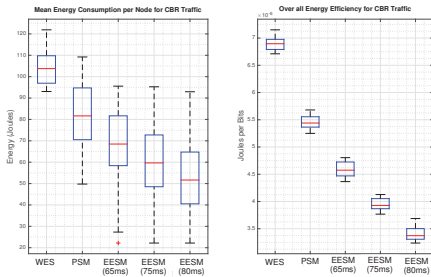


Fig. 3. Comparison of Mean Energy Consumption per Node and Over all Energy Efficiency of CBR Traffic

WES is a mesh topology without using any energy saving mechanism and as expected, consumes the highest energy per node. PSM uses less energy compared to WES while keeping mean packet delay and goodput for each flow equivalent with WES. For EESM, three different sets of simulations were conducted to analyze the impact of different values of  $T_{th}$ . With higher values of  $T_{th}$ , nodes tend to put radios into sleep mode more aggressively which results in sharing most of the traffic load among few active radios. This may increase end-to-end packet delay and decrease in throughput for flows going through more congested nodes. However results show that EESM outperforms WES and PSM in terms of energy conservation and efficiency without significant impact on delays and throughput. EESM with  $T_{th} = 65ms$  saves 15% energy per node compared to PSM. EESM with  $T_{th} = 75ms$

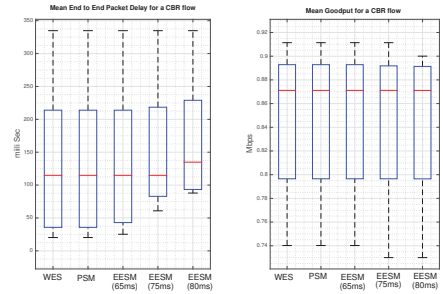


Fig. 4. Mean Packet delay and Goodput for a CBR flow

and  $T_{th} = 80ms$  saves 26.83% and 36.59% more energy per node respectively. Figure 5 shows there is slight increase in packet delay for aggressive EESM in order of milliseconds but goodput remains approximately similar. As long as there is low to medium traffic load in the network, energy saving mechanisms have better opportunities to conserve energy by putting more radios into sleep. With higher traffic density, more radios are required to keep the traffic flowing without unwanted delays hence giving less chances for switching states of radios. In case of fully saturated network, PSM and EESM will eventually converge with WES by keeping all radios ON. This behaviour is evident where energy efficiency is degrading in PSM and EESM as the number of traffic flows increases in the network.

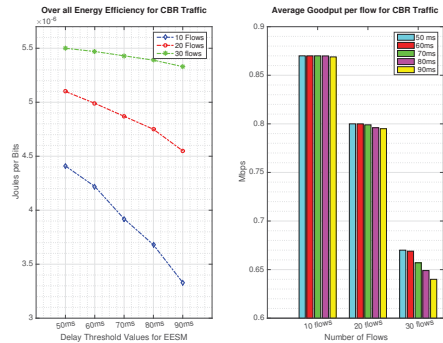


Fig. 5. Comparison of Energy Efficiency and Goodput for different number of CBR flows

**B. HTTP Traffic using TCP**

In second case, ten HTTP flows were generated randomly. Out of ten, four flows are directed to two gateways and rest are peer-to-peer. Having TCP as its underlying protocol, HTTP communication is connection based and slower than UDP traffic. TCP controls the data flow based on link quality

and when congestion on a node tries to adjust the flow rate to avoid packet losses. For low to medium traffic in the network, this behaviour of TCP does not interfere with energy saving mechanism. Figure (6&7) show similar tendency for HTTP flows as for CBR traffic in terms of energy efficiency and network performance. The decrease in goodput is more obvious because of slower data rates of HTTP flows.

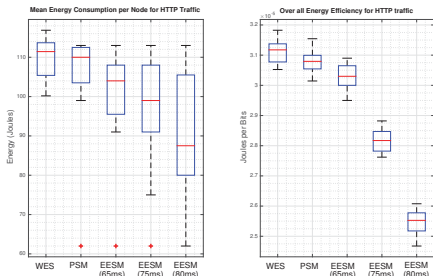


Fig. 6. Comparison of Mean Energy Consumption per node and Overall Energy efficiency of Network for HTTP Traffic

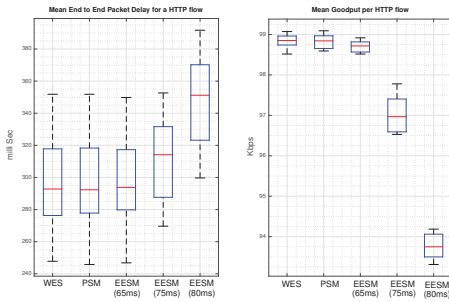


Fig. 7. Mean Packet delay and Goodput for a HTTP flow

With  $T_{th} = 65ms$ , at the cost of a small delay of few milli seconds and 0.3% drop in goodput, EESM achieves 2.5% better energy efficiency than WES or PSM and saves 5% energy per each node. Using  $T_{th} = 75ms$  which incurs 10% increase in packet delay and 2.02% drop in throughput, EESM saves 10.9% energy at each node and improves energy efficiency by 8.9%.  $T_{th} = 80ms$  saves even more energy per node (20%) but at the cost of approximately 7% drop in throughput and 15% increase in end to end packet delay.

In case of high traffic density if nodes put their radios to sleep more aggressively, TCP detects the increase in processing delay over active radios and slows down the flow rates as shown in Figure 8. Switching delay  $T_{del}$  does not change with this decrease in flow rate but over all end to end delay

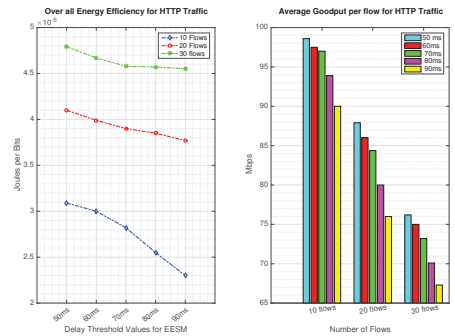


Fig. 8. Comparison of Energy Efficiency and Goodput for different number of HTTP flows

increases and throughput drops. Adjusting the threshold value according to traffic density in the network can mitigate this problem.

### V. CONCLUSION

To improve the energy efficiency of multi-radio multi-channel wireless mesh networks, we have proposed EESM, the enhanced energy savings mechanism. We assume that each node in a network has a dedicated receiver radio with a known channel, and multiple transmitter radios. Based on observed flow rates and delays, a node may put individual transmitters to sleep; nodes wake up transmitters again once sufficient congestion is observed. An empirical evaluation demonstrates that EESM can save significant energy over the default power saving mode (PSM), without degrading network performance for both constant bit rate flows and HTTP traffic, for low and medium traffic loads. As the traffic load increases, energy efficiency begins to decrease, and EESM eventually converges to PSM. Different choices of a threshold parameter for EESM gives different tradeoffs between energy consumption and network performance. For delay tolerant traffic, a higher value of the threshold is recommended to conserve energy. In future work, we will investigate dynamic selection of the threshold values in response to changing flows rates in TCP, and we will investigate flow-specific thresholds, in order to achieve different QoS performance as required by different applications.

### ACKNOWLEDGMENT

This work is supported by the CTVR Grant (SFI 10/CE/1 1853) from Science Foundation Ireland.

### REFERENCES

[1] Shengyang Chen and Gabriel-Miro Muntean. E-mesh: An energy-efficient cross-layer solution for video delivery in wireless mesh networks. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–7, 2012.

- [2] Soumya Dey, Himanshu Sarmah, Subhadeep Samantray, Devraj Divakar, and SS Pathak. Energy efficiency in wireless mesh networks. In *Computational Intelligence and Computing Research (ICCC), 2010 IEEE International Conference on*, pages 1–4. IEEE, 2010.
- [3] Lúcio Studer Ferreira and Luís M Correia. Energy-efficient radio resource management in self-organised multi-radio wireless mesh networks. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pages 232–236. IEEE, 2011.
- [4] Vaibhav Garg, Kotaro Kataoka, and Siva Subramanya Rohith Talluri. Performance evaluation of wireless ad-hoc network for post-disaster recovery using linux live usb nodes. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2015 IEEE 11th International Conference on*, pages 125–131, 2015.
- [5] Christian Köbel, Walter Baluja Garcia, and Joachim Habermann. A survey on wireless mesh network applications in rural areas and emerging countries. In *Global Humanitarian Technology Conference (GHTC), 2013 IEEE*, pages 389–394, 2013.
- [6] Sarra Mamechaoui, Fedoua Didi, and Guy Pujolle. A survey on energy efficiency for wireless mesh network. *Computer Networks & Communications*, 5(2), 2013.
- [7] Christian Mannweiler, Pratip Chakraborty, Andreas Klein, and Hans D Schotten. Solarmesh-deployment aspects for wireless mesh networks in developing countries. In *International Conference on e-Infrastructure and e-Services for Developing Countries*, pages 105–114. Springer, 2012.
- [8] TN Nagabhushan, SP Shiva Prakash, and Kirill Krinkin. Power saving routing algorithms in wireless mesh networks: A survey. *Convergence of FRUCT association*, 2013.
- [9] M. A. Nezhad, L.ç Cerdà-Alabern, B. Bellalta, and M.G Zapata. A semi-dynamic, game based and interference aware channel assignment for multi-radio multi-channel wireless mesh networks. *International Journal of Ad Hoc and Ubiquitous Computing*, 14(3):200–213, 2013.
- [10] M. Portmann and A. A. Pirzada. Wireless mesh networks for public safety and crisis management applications. *IEEE Internet Computing*, 12(1):18–25, 2008.
- [11] SP Prakash, TN Nagabhushan, and K. Krinkin. Energy aware power save mode management in wireless mesh networks. In *Open Innovations Association (FRUCT), 2013 14th Conference of*, pages 122–131. IEEE, 2013.
- [12] SP Shiva Prakash, TN Nagabhushan, and Kirill Krinkin. Energy aware power save mode based aodv routing for ieee 802.11 s wireless mesh networks. *Energy*, 137(4), 2016.
- [13] P. Serrano, A. Garcia-Saavedra, G. Bianchi, and A. Banchs, A. and Azcorra. Per-frame energy consumption in 802.11 devices and its implication on modeling and design. *IEEE/ACM Transactions on networking*, 23(4):1243–1256, 2015.
- [14] Jianfeng Wang, Yuguang Fang, and Dapeng Oliver Wu. A power-saving multi-radio multi-channel mac protocol for wireless local area networks. In *Proceedings INFOCOM 2006, 25TH IEEE International Conference on Computer Communications*, pages 1–12, 2006.
- [15] A. Yarali, B. Ahsant, and S. Rahman. Wireless mesh networking: A key solution for emergency & rural applications. In *Advances in Mesh Networks, 2009. MESH 2009. Second International Conference on*, pages 143–149. IEEE, 2009.



# A Topological Map Merging Method Based on Localization of Multi-Robot SLAM

Rui Qu<sup>1</sup>, Yi Zhao<sup>2</sup>, Zhuqing Jiang<sup>1</sup>, Beihang Zhang<sup>1</sup>, Chan Wang<sup>1</sup>, Aidong Men<sup>1</sup>, Chonghua Liu<sup>2</sup>

<sup>1</sup>Beijing University of Posts and Telecommunications, Beijing, China 100876

<sup>2</sup>Beijing Institute of Spacecraft System Engineering, Beijing, China, 100094

**Abstract**—In recent years, the research on single-robot simultaneous localization and mapping (SLAM) has made a great success. It is required for mobile robots to be able to explore a prior unknown space without a global positioning reference. However, multi-robot SLAM faces many challenging problems, including unknown robot poses, unshared map, complexity, and unstable communication. In this article, we operate the construct called probabilistic generalized Voronoi diagram directly on occupancy grid maps to determine the relative transformation. In addition, the Unscented Transformation is used in this article, accounting for the uncertainty in the fusion process.

**Keywords:** SLAM, Voronoi Diagram, Map Merging

## I. INTRODUCTION

Recent advances in mobile robotics have allowed autonomous systems to be involved in many successful applications, including planetary exploration, search and rescue, surveillance, and other service scenarios[1]. For the purpose of accomplishing a generic task, a main prerequisite for a mobile robot which is deployed in an unknown area is the ability to autonomously navigate, while exploiting the information acquired through the joint estimation of its positions and a model of the surrounding environment. The problem of estimating the robot pose and the environment representation is usually defined as simultaneous localization and mapping (SLAM)[2, 3].

When multiple robots explore an unknown environment, they share raw data[5] or processed maps[4]. In contrast, sharing maps uses less bandwidth and reduces the need to process raw data; however, the performance is dependent on the map quality. Each robot has its own task, such as building a map of local position. Moreover, they have to generate shared maps based on their data. Using shared maps, robots coordinate their exploration strategies to maximize the efficiency of exploration. Multiple autonomous mobile robots can complete the task through cooperation and give a more accurate map by data fusion. Fenwick et al. extended single robot SLAM algorithm based on EKF to the multiple robots[6,7]. Howard extended single-robot SLAM algorithm based on particle filter to multi-robot applications[5]. Thrun put forward multi-robot hybrid map building method that combines fast maximum likelihood map growing with a Monte Carlo localizer[8]. Thrun points out that the key to multi-robot map building is to determine the initial position and orientation of each single robot[7].

Multiple-robot SLAM can also be categorized based on the method used to process measurement data. There are mainly two kinds of multi-robot SLAM:

### A. Feature-based SLAM

It is often performed by cameras, unique objects could be extracted from measurements and used for fusion, such as features and landmarks. Feature-based multiple-robot SLAM has been implemented using an information filter[9], an extended Kalman filter[10], and a particle filter[11].

### B. View-based SLAM

It is performed by entire laser scans, which are matched using scan-matching algorithms. Our approach is based on view-based SLAM. Thrun proposes a probabilistic multiple-robot view-based SLAM algorithm in [8]. This method is robust; however, the approximate initial poses of the robots are assumed to be known before the start of the mission.

In this article, we employ the generalized Voronoi diagram(GVD) to extract the skeleton. Then the uncertainty in the maps is applied to build a probabilistic generalized Voronoi diagram(PGVD), after the PGVDs are built, the edges are matched using a two-dimensional (2-D) cross-correlation that will preferentially match the areas of the maps that have higher confidences. There is still nonlinear uncertainty existing in the relative transformation, to solve this problem, we operate the Unscented Transformation(UT) on one PGVD.

The rest of the paper is organized as follows: Section II presents the method to solve the uncertainty propagation problem employed in this paper, Section III introduces the method for topological multi-robot SLAM, Section IV presents some experimental results, and Section V makes some general conclusions and discusses future work.

## II. TRANSFORMATION UNCERTAINTY

In this article, uncertainty in map fusion are divided into two types:

1) *Uncertainty due to pose and sensor measurements*: It is mainly caused by the uncertain pose of each robot. This uncertainty could be considered at the individual robot level. It could be solved by increasing the ability of measurements. But it does not happen in map fusion process, so we call it map uncertainty in this article.

2) *Uncertainty due to transformation*: This uncertainty happens mainly in the relative transformation between two maps. It is the main problem we will solve in this article, the transformation uncertainty, rotation uncertainty, and translation uncertainty are all this type of uncertainty.

The uncertainty associated with the relative transformation matrix is represented as a covariance matrix. In general, the nonlinearity and uncertainty of the transformation will cause a non-Gaussian distribution. However, it is possible to approximate the non-Gaussian distribution with a Gaussian distribution.

Assume the average of  $x$  is  $\bar{x}$ , the covariance matrix is  $P_x$ , if  $g$  is linear mapping:  $y = Lx$ , then the average and covariance are  $\bar{y} = L\bar{x}$  and  $P_y = LP_xL^T$ . But if  $g$  is nonlinear mapping, to obtain the optimal solution in Minimum Mean Square Error, we must use infinite number of parameters to describe the infinite dimensional distribution of  $y$ . The general practice is to linearize  $g$  to get the linear mapping of  $g$ ,  $T: R^{N_x} \rightarrow R^{N_y}$ . For example, the classic method often uses Taylor expansion, ignoring the effect of series more than second-order. This method is called linear method, but it has several defects:

(1) When mapping  $g$  has serious linear problem in the neighborhood of  $\bar{x} + \delta x$ , ignoring the high order component could cause great error. Although there is a second order linear method, it requires a large amount of calculation and complexity, so it is difficult to apply.

(2) Because the calculation involves the Jacobian matrix, so in some certain situations where Jacobian matrix does not exist, this method is invalid.

(3) For several complex nonlinear mapping function, calculation of the Jacobian matrix is very complicated.

To solve the nonlinear estimation problem mentioned above, traditional linear method is to do linear approximation and estimation. Actually, using finite parameters to approximate the probabilistic statistical characteristics of random variables is easier than approximate any nonlinear mapping function[13-15]. Based on this, S.J.Julier proposed Unscented Transformation[12,13,16]. The basic steps of UT are: Production of the sigma-point set of  $x$ , nonlinear transformation of uncertainty, the calculation of statistical properties on the  $Y$ :

(1) Choosing a set of sigma-points  $\Gamma_x = [x^{(i)}, w^{(i)}]_{i=0}^p$ , where the weights satisfy the normalization condition  $\sum_{i=0}^p w^{(i)} = 1$ , and

$$\sum_{i=0}^p w^{(i)} x^{(i)} - \bar{x} = 0 \quad (1)$$

$$\sum_{i=0}^p w^{(i)} (x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T - P_x = 0 \quad (2)$$

(2) Using nonlinear mapping  $g: R^{N_x} \rightarrow R^{N_y}$  to transfer uncertainty of random quantity, getting new sigma-point set  $\Gamma_y = [y^{(i)}, w^{(i)}]_{i=0}^p$ ,  $y^{(i)} = g(x^{(i)})$ , weights are the same.

(3) Calculating the average and variance of  $y$  after nonlinear mapping based on sigma-point set.

$$\hat{y} = \sum_{i=0}^p w^{(i)} y^{(i)} \quad (3)$$

$$P_y = \sum_{i=0}^p w^{(i)} (y^{(i)} - \hat{y})(y^{(i)} - \hat{y})^T \quad (4)$$

$$P_{xy} = \sum_{i=0}^N w^{(i)} (x^{(i)} - \bar{x})(y^{(i)} - \hat{y})^T \quad (5)$$

Then we repeat the example based on the articles of E.Wan and R.Merwe[17-21], this image shows the difference between UT method and linear method on statistical characteristics of the random quantity after nonlinear transformation. The UT method could reach much more estimation accuracy. The usage of UT in this algorithm will be explained in next section.

### III. MAP MERGING WITH PGVD

In this section, we will discuss the procedure of the map merging process, including three main challenges that need to be overcome:

- 1) The relative transformation from map1 to map2 needs to be found.
- 2) Uncertainty of the transformation should be taken into account in the transformation.
- 3) The occupancy grid map(OGM) probabilities from map2 need to be incorporated with the OGM probabilities of map1. Here is the procedure:

#### A. Rotation

1. Using the Radon transform, we could resolve the relative rotation by looking for peaks in the Radon images of both maps.

$$\psi = \arg \min_{\theta} R_{\theta=0:180}(map_1) - \arg \min_{\theta} R_{\theta=0:180}(map_2) \quad (6)$$

where  $R_{\theta=0:180}(m_i)$  is the Radon image of the input map over the specified domain of angles  $\theta$ . The two aligned maps now are defined as  $m1 = map_1$  and  $m2 = T_{0,0,\psi}(map_2)$ , where  $T_{x,y,\alpha}(m)$  is a transformation matrix representing a translation by  $(x,y)$  and a rotation by  $\alpha$ .

2. After finding the rotation, the uncertainty on the map is propagated by Gaussian blur filter, assuming zero translation and  $\sigma_{xx}^2 = \sigma_{yy}^2 = \sigma_{xy}^2 = 0$ . Then is the process of finding the relative translation.

#### B. Translation

1. To build the PGVD of the two maps, there are two steps:

1) To find the GVD efficiently using mathematical morphological operations on the binary OGM.

The GVD of the binary map is generated by eight D-type hit-and-miss transform masks [18]. Each mask is designed for a particular situation to guarantee the connectedness (using the connectivity-eight model). The GVD is represented as a

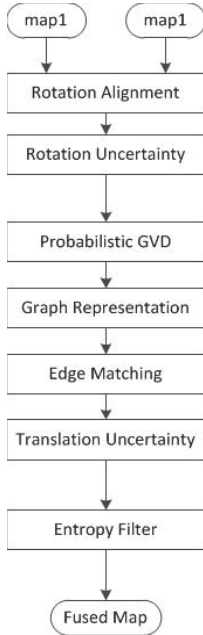


Fig. 1: Procedure of PGVD method

matrix,  $S = [s_{i,j}]$ , with the same size as the map,  $M \times N$ , defined as

$$S = [s_{i,j}]_{i=1..M, j=1..N} = \begin{cases} 1, & \text{if } m(i, j) \in GVD \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

2) To compute the associated probabilities of each cell in the GVD based on the actual probabilities in the OGM.

To build PGVD of the map, two definitions will be introduced below:

**Definition1:** Contact Points. The contact points of a cell in the GVD are all the occupied cells in  $m$  that have the same distance from the cell as the closest occupied cell. The probabilities of the contact points in the OGM are the foundation to build the PGVD.

**Definition2:** Probabilistic GVD. Each cell in the GVD is represented by a binary random variable (RV) representing the probability that it has two or more occupied contact points based on their probabilities of occupancy in the OGMs.

The PGVD  $S_p$  has the same structure as the GVD, except that each of the  $s$  cells is represented as a binary RV  $G_i^p$ ,  $i = 1..s$ , where  $p(G_i^p = 1) = p(G_i^p)$  is the probability that cell has been correctly placed in the GVD. Each cell in the PGVD has an associated set of contact points, A cell belongs in the GVD if at least two of the contact points are occupied. For each cell in the PGVD  $G_i^p$ ,  $i = 1..s$ , we must

determine the probability that it has at least two occupied contact points based on their probabilities of occupancy. The function  $f(n_i, k)$  can be defined recursively as the following:

$$f(n_i, k) = p(c_{n_i})f(n_i - 1, k - 1) + (1 - p(c_{n_i}))f(n_i - 1, k) \quad (8)$$

$k$  means the number of contact points that are actually occupied.

2. To identify edges in the graph, a dilation mask is applied to each vertex in  $V$ , which is the set of all vertices. The result of the dilation operation is a new map  $D$ , that contains the dilated vertices. Then we could build the edge matrix, probabilistic edge matrices (PEMs), preparing for the edge matching.

3. After building the probabilistic edge matrices (PEMs), we use them to find the transform matrix between the two maps. Edges with short lengths are removed to avoid processing short edges which have a higher chance of producing false positives. The edges of each probabilistic edge matrix are matched using a 2D cross correlation:

$$\varepsilon^{(i_1, i_2)} = e_1^{i_1} * e_2^{i_2} \quad (9)$$

where  $\varepsilon^{(i_1, i_2)}$  is the cross-correlation operation of edge  $e_1^{i_1}$  and edge  $e_2^{i_2}$  which has size  $(m_1 + m_2 - 1) \times (n_1 + n_2 - 1)$ . We can define the similarity matrix  $\Gamma_1^2 \dots$

between  $E_1^P$  and  $E_2^P$  as

$$\Gamma_1^2 = [\gamma_{i_1 i_2}]_{i_1=1..N_1, i_2=1..N_2} \quad (10)$$

where  $\gamma_{i_1 i_2} = \max(\varepsilon^{(i_1, i_2)})$

Then the best candidate for a match corresponds to the maximum value is chosen in the similarity matrix  $\Gamma_1^2$ :

$$[i_1^*, i_2^*] = \arg \max(\Gamma_1^2) \quad (11)$$

where  $i_1^*$  and  $i_2^*$  are the indices of the most likely similar edges from the two maps. Then we could calculate the relative translation base on the following relation[17]:

$$[\delta_x, \delta_y] = [\mu_x(e_1^{i_1^*}) - \mu_x(e_2^{i_2^*}), \mu_y(e_1^{i_1^*}) - \mu_y(e_2^{i_2^*})] \quad (12)$$

The final transformed map is defined as:

$$m_2^T = T_{x,y,0}(T_{0,0,\psi}(map_2)) = T_{x,y,\psi}(map_2) \quad (13)$$

To avoid the false matches, the average confidence is introduced for the matched edge as:

$$\alpha_k^{i_k} = \frac{1}{L} \sum_{l=1..L} p(\varepsilon_k^{i_k}) \quad (14)$$

Only when the maximum value in the correlation matrix satisfy the following could the best matching be considered as a candidate match:

$$\max(\varepsilon^{(i_1, i_2)}) > \rho \alpha_1^{i_1} \alpha_2^{i_2} \min(l_1, l_2) \quad (15)$$

where  $\rho$  is a desired matching percentage.

4. The translation uncertainty could be solved using Unscented Transform (UT), which is introduced in section II.

### C. Fusion

After finding the relative transformation between the two maps, the entropy filter is applied to the fused map, the entropy of each cell can be described as:

$$H(m(i, j)) = -p_{ij} \log p_{ij} - (1 - p_{ij}) \log(1 - p_{ij}) \quad (16)$$

Mutual information could be described as:

$$I_{ij} = H(m_1(i, j)) - H(m_{fused}(i, j)) \quad (17)$$

The final fused map is defined as:

$$m_{final}(i, j) = \begin{cases} m_{fused}(i, j), & I_{ij} \geq 0 \\ m_1(i, j), & I_{ij} < 0 \end{cases} \quad (18)$$

where only values from  $m_{fused}(i, j)$  which result in positive information are kept.

## IV. SIMULATIONS AND RESULTS

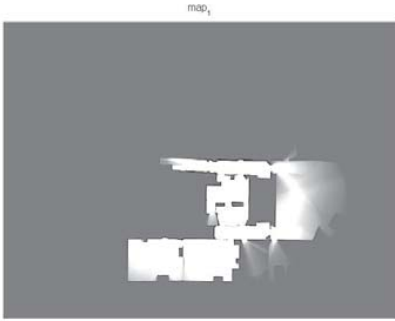


Fig. 2: Map1 before Alignment

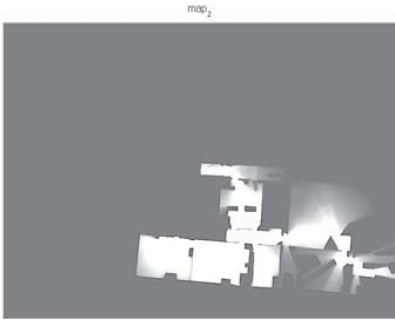
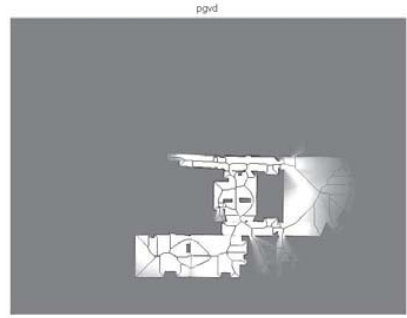


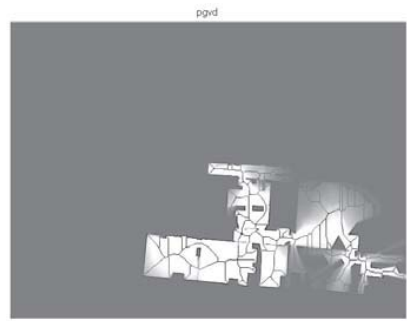
Fig. 3: Map2 before Alignment

The experiment is performed on RADISH Fort AP Hill data set. Fig.2 and Fig.3 show two input maps before alignment.



oavd

Fig. 4: PGVDs of Map1



navd

Fig. 5: PGVDs of Map2

Fig.4 and Fig.5 show the PGVDs of the maps after being aligned. The alignment is  $5.5^\circ$ . The marked edges are the filtered edges for cross-matching. The  $\Gamma_1^2$  matrix is shown in equation(20). From the matrix, we could find that the fourth edge of the first map has the highest similarity to the third edge of the second map. After calculating the translation vector  $T = [-25, -1]^T$  and rotation  $5.5^\circ$  using these two edges, we should find whether the match is valid according to equation(15). The matching percentage is chosen to be 95%, the lengths of these two edges are 44 and 45.

$$0.95 * 0.86 * 0.89 * \min(44, 45) = 31.99 < 32.9 \quad (19)$$

$$\Gamma_1^2 = \begin{bmatrix} 31.6 & 2.0 & 6.4 & 6.1 & 1.9 & 0.8 & 5.4 \\ 12.9 & 1.9 & 2.0 & 11.7 & 0.9 & 0.5 & 1.6 \\ 11.9 & 2.8 & 5.4 & 3.0 & 1.4 & 0.5 & 2.5 \\ 10.7 & 6.7 & 32.9 & 1.0 & 5.3 & 1.8 & 8.1 \\ 9.1 & 1.9 & 8.9 & 2.3 & 0.9 & 0.5 & 1.6 \\ 5.9 & 0.4 & 2.1 & 0.4 & 0.4 & 0.2 & 1.9 \\ 3.2 & 9.9 & 3.9 & 0.5 & 11.6 & 4.0 & 6.9 \\ 0.9 & 7.2 & 2.9 & 0.3 & 10.2 & 3.5 & 4.1 \end{bmatrix} \quad (20)$$

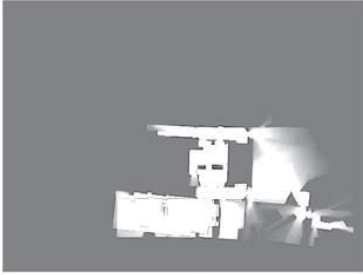


Fig. 6: Fusion Map after Entropy Filter and UT

Method	Processing(s)	Verification(%)
PGVD based	15	95
Map segmentation[22]	105	95
ARW map merging[5]	168	93

TABLE I: Comparison of Three Algorithms

Fig.6 shows the final fusion map after UT and entropy filter. The major benefit of edge matching for map fusion is the low processing time requirement. Table I shows the comparison of the processing times and verification for the data set among edge matching method, map segmentation method and ARW map merging. Map segmentation and ARW map merging are two typical map merging methods, which perform well in accuracy but lack the speed. As the results show, the proposed method operates at least eight times faster, and the verification index[5] shows the accuracy of the results. The comparison results for the simulation are not provided since the other two methods are not effective for such large and complex environments.

The results obtained in the experiments showed that SLAM based on PGVD using UT generates better solutions than the other two map merging algorithms, because the method used in this article accounts for the transformation uncertainties to improve the accuracy.

## V. CONCLUSIONS AND FUTURE WORK

In this article, a topological map merging method based on PGVD is applied to realize the multi robot SLAM, and the uncertainty of relative transform is solved by UT. In Section IV, some experiment results show the feasibility of the method. From this article, we could find that this method has more time efficiency and more accuracy than the method mentioned before. In future work, we will think about the reasons which influence the value in the variance matrix, to solve the problem when there are several values which meet the conditions.

## ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (61671264 and 61671077).

## REFERENCES

- [1] Sajad Saeedi,Liam Paull, Efficient Map Merging Using a Probabilistic Generalized Voronoi Diagram,*International Conference on Intelligent Robots and Systems* pp.4419-4424 Harlow, England: Addison-Wesley, 1999.
- [2] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*, Springer, NewYork, NY,USA, 2008.
- [3] S.Thrun, Simultaneous localization and mapping, in *Springer Tracts in Advanced Robotics*, M. Jefferies andW. Yeap, Eds., vol.38, pp. 13C41, Springer, Berlin, Germany, 2008.
- [4] H. Zhang, X.-P. Fan, and Y.-L. Liu, Simultaneous localization and mapping for mobile robots, *Journal of Data Acquisition and Processing*,vol. 20, no. 4, pp. 458C465, 2005.
- [5] A. Birk and S. Carpin, Merging occupancy grid maps from multiple robots,Proc. *IEEE: Special Issue Multi-Robot Syst.*, vol. 94, no. 7, pp. 1384C1387, 2006.
- [6] A. Howard, Multi-robot simultaneous localization and mapping using particle filters, *Int. J. Robot. Res.*, vol. 25, no. 12, pp. 1243C1256, 2006.
- [7] J. W. Fenwick, P. M. Newman, and J. J. Leonard, Cooperative concurrent mapping and localization, in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp.1810C1817,Washington, DC, USA, May 2002.
- [8] S. Thrun and Y. Liu, Multi-robot SLAM with sparse extended information filters, *Robotics Research*, vol. 15,no. 5, pp. 254C266,2005.
- [9] S. Thrun, A probabilistic on-line mapping algorithm for teams of mobile robots, *International Journal of Robotics Research*, vol.20, no. 5, pp. 335C363, 2001.
- [10] P. Dario and R. Chatila, Multi-Robot SLAM with Sparse Extended Information Filters,(*Springer Tracts in Advanced Robotics vol. 15*). Berlin, Heidelberg, Germany: Springer, pp. 254C266, 2005.
- [11] X. S. Zhou and S. I. Roumeliotis, Multi-robot SLAM with unknown initial correspondence: The robot rendezvous case, in *Proc. IEEE/RSJ Int. Conf.Intelligent Robots Systems*, 2006, pp. 1785C1792.
- [12] A. Gil, O. Reinoso, M. Ballesta, and J. Miguuel, Multi-robot visual SLAM using a Rao-Blackwellized particle filter, *Robot. Auton. Syst.*, vol. 58, no. 1, pp. 68C80, 2010.
- [13] Julier S J,Uhlmann J K,Unscented filtering and nonlinear estimation[J],*Proc IEEE*, 2004, 92( 3) : 401- 422.
- [14] Julier S J, Uhlmann J, Durrant-Whyte H F.A new method for the nonlinear transformation of means and covariances in filters and estimators[J],*IEEE Trans Automatic Control*, 2000, 45( 3) : 477- 482.
- [15] Julier S.The spherical simplex unscented transformation[C].*Proc Amer Control Conf*, 2003, 3: 2430- 2434.
- [16] Chen Zhe.Bayesian filtering: from Kalman filters to particle filters, and beyond[R].Hamilton: McMaster University, 2003. Julier S J, Uhlmann J K.A new extension of the Kalman filter to nonlinear systems[C].*Proc of the 11th Int Symp Aerospace/Defense Sensing, Simulation and Controls*, 1997: 182- 193.
- [17] Wan E, van der Merwe R, Nelson A T.Dual estimation and the unscented transformation[C].*Neural Information Processing Systems*, 2000, 12: 666- 672.
- [18] Wan E A, van der Merwe R.The unscented Kalman filter for non-linear estimation[C].*The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000 ( AS-SPCC )*, Lake Louise, Oct 2000: 153- 158.
- [19] van der Merwe R, Wan E.The square- root unscented Kalman filter for state and parameter estimation[C].*Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ( ICASSP )*, Salt Lake City, May 2001: 3461- 3464.
- [20] van der Merwe R, Wan E.Efficient derivative- free Kalman filters for online learning[C].*2001 Proc European Symposium on Artificial Neural Networks( ESANN )*, Bruges, April 2001: 205- 210.
- [21] van der Merwe R, Wan E.Sigma- point Kalman filters for probabilistic inference in dynamic state- space models[EB/OL].*Workshop on Advances in Machine Learning*, Montreal, June 2003. http: //www.iro.umontreal.ca/ kegl/CRMWorkshop/program.html.

- [22] S. Saeedi, L. Paull, M. Trentini, and H. Li, Multiple robot simultaneous localization and mapping, in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2011, pp. 853C858.

# DDoS Attack Detection in SDN-based VANET Architectures

Martina Todorova, Stamelina Todorova, Anton Katov,  
 Albena Mihovska and Ramjee Prasad  
 Center for TeleInfrastruktur (CTIF)  
 Aalborg University (AAU)  
 Aalborg, Denmark  
 mtodor15@student.aau.dk, stodor15@student.aau.dk,  
 annka@es.aau.dk, albena@es.aau.dk, prasad@es.aau.dk

Vladimir Pulkov  
 Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 fktk-dekan@tu-sofia.bg

**Abstract** — Software-defined networking (SDN) is an emerging technology, which is based on decoupling the network control plane from the data forwarding plane and implementing the control functions as software modules on a centralized controller. This paper investigates an SDN-based Vehicular Ad Hoc Networks (VANET) and how a Distributed Denial of Service (DDoS) attack can be detected and prevented. We estimate how DDoS can influence the traffic characteristics in an SDN-based VANET scenario and we design and test a DDoS detection algorithm. Based on a predetermined set of traffic features, entropy is used to measure the degree of randomness of occurrence of the destination IP address of the transmitted packets. In the current paper we investigate the fluctuations of the entropy values in the presence and the absence of an attack and determine the appropriate threshold values, which can be used in the attack detection process.

**Keywords**— SDN, VANET, security, DDoS, attack detection, entropy, botnet

## I. INTRODUCTION

Software-defined networking (SDN) is an emerging technology, with the ability to decouple the network control and data forwarding planes, thus focusing the control functionalities in a centralized controller and reducing the required networking equipment to simple programmable data forwarding devices [1]. Due to this centralized control the network becomes more dynamic, and the network resources are managed in a more efficient and cost-effective manner

Vehicular Ad Hoc Networks (VANET) is another rapidly growing technology that has gained a lot of popularity as a key component of Intelligent Transport Systems (ITSs). Its exclusive characteristics include predictable mobility, no power limitations, variable network density, high computational ability, rapid changes in network topology, etc.

Security can be seen as one of the most challenging concern for both technologies, especially in a scenario of joint implementation. The centralized control of SDN can be considered as one of the main benefits, but the entire network could be compromised, if the controller is under attack. One of the most common types of attack is the Distributed Denial of Service (DDoS) attack.

In the literature, a lot of DDoS attack detection methods based on different traffic features have been proposed for conventional networks, but there is no research done for SDN-based VANET networks. Existing DDoS attack methods for the detection and mitigation of malicious traffic in SDN and VANET separately were reported in [2], [3], [4], [5], [6], [7], [8]. This paper proposes the use of entropy for implementing a DDoS detection algorithm for an SDN-based VANET scenario. We adjusted the parameters of the entropy method proposed for a static SDN [8], after taking into account the dynamic nature of the VANET networks. In order to achieve realistic results, we tested the algorithm using dynamic topologies to correctly reflect the node mobility in a VANET scenario. It should be noted that VANET nodes implement the functionalities of both hosts and switches at the same time [9]. We examined how the network is affected by a DDoS attack coming from one host, followed by a case study of the influence of launching malicious traffic via botnet. The paper focuses on the impact of the DDoS attack on the control plane layer. The proposed algorithm is implemented as a software module on top of the controller network operating system, by means of additional functions for attack traffic detection. The obtained results investigate the effectiveness of the proposed DDoS attack detection solution.

This paper is further organized as follows. Section II explores security challenges in SDN-based VANET networks with a focus on DDoS attacks, Section III presents the proposed solution for DDoS attack detection, Section IV contains the simulations and the results and Section V is devoted to the conclusion.

## II. SECURITY CHALLENGES IN SDN-BASED VANET NETWORKS

Some security problems and vulnerabilities in OpenFlow-based networks, as well as various threat vectors in an SDN architecture were reported in [10]. There are some threat vectors specifically directed to SDN, with the most critical ones appearing when the attacker is launching the attack on the control plane. OpenFlow networks are subject to a variety of security and dependability problems such as spoofing, repudiation, tampering, information disclosure, denial of service, elevation of privileges, and the assumption that all

applications are benign and will not affect the SDN operation [11].

Security and privacy in VANETs face many challenges. In [12] the authors explored the sources of attacks and malicious activities in VANET. They claim that for Dedicated Short-range Communications (DSRCs) range, the vehicular network is open and reachable from everywhere, and an easy target for malicious intentions. The authors identified jamming, forgery, impersonation and privacy as the main threat vectors. In a vehicular environment any incorrect or malicious message could directly affect the humans' life, particularly in the light of the public acceptance of the technology and therefore achieving high level security is of primary concern. In [13] the types of attacks in VANET were elaborated.

#### A. DDoS Attack in SDN-based VANET Environment

There is a vulnerable relation between SDN-based networks and DDoS attacks. Network capabilities, such as the global view of the network and the dynamic updating of forwarding rules, can facilitate the DDoS attacks detection, but the separation of the control plane from the data plane in an SDN leads to emerging new types of attacks [2]. In SDN networks the switches belong to the data plane and they request to obtain flow rules from the control plane when the new packets arrive and the switches do not know how to handle them [14]. The attacker can take advantage of that functionality of the network and if she sends a large volume of traffic it will exhaust the resources of the controller.

The impact of the DDoS attacks can grow greatly if the attackers use a botnet to overwhelm their victim's network. Botnet is one of the most serious network security threats which the home user, organizations and government can meet. A botnet is a group of many infected machines communicating with each other, called zombies, managed by a malicious entity called the botmaster [15]. In SDN-based VANET scenario the attacker can build a botnet, if she/he infects several vehicles and launches the attack through them. In this paper we investigate the impact of UDP flooding DDoS attack, including a simulation test case, focused on how the botnets can overwhelm the network. SDN-based VANET networks, where the vehicles can be seen as mobile devices, bring a perfect platform for attackers to launch DDoS attacks. In an attempt to offer a possible solution, we propose an entropy-based mechanism for early detection of DDoS attacks.

### III. DDoS ATTACK DETECTION SOLUTION

#### A. Theory of the Solution

The common aspect in all kinds of denial of service attacks is pushing big amount of traffic into the network to exhaust its resources. In the general case, anomalies are related to the type of data in the network [4].

In this paper Used Datagram Protocol (UDP) packets are used to launch the normal and attack traffic into SDN-based VANET network. UDP is a datagram protocol which offers a minimal transport service. UDP flood is a type of DDoS attack associated with inundating random ports on the targeted host with IP packets containing the UDP datagrams. The fact that UDP is a connectionless networking protocol can make UDP

communication even more vulnerable to abuse, as source IP address may be spoofed. The fact that a huge amount of UDP packets can be used to initiate a DDoS attack in an SDN-based VANET network motivated us to use entropy as a detection tool that can measure the randomness of the occurrence of the destination IP addresses. Entropy is an essential concept of information theory, which is a measure of the uncertainty or randomness associated with a random variable or in our case data coming over the network (destination IP address).

The in-built functionality of SDN networks to operate on a overlay network abstractions, can provide the necessary network status overview and help the controller to decide if the attack has occurred or not. The reason why packets come to the controller is that the source IP address is new. For every new incoming connection, the controller is installing a flow into the switch, so that the rest of the incoming packets will be directed to the destination without further processing. Furthermore, it can be deduced that the destination host is in the network of the controller [8].

The reference scenario for this work is based on SDN architectures for VANET. The network consists of switches (RSUs and Base Stations in our case) and hosts (vehicles and RSUs) that are connected to it. As it was stated above, it can be safely assumed that the packets in the controller are always new and the destination is in the network. Based on this we propose a mechanism that quantifies the level of randomness by calculating the entropy based on a window size. The window size is the number of the incoming new packets that are used for calculating entropy. In the general case, maximum entropy occurs when each packet is destined to exactly one host. Minimum entropy occurs when all the packets in a window are destined for a single host. If a set of data  $W$  exists and it consists of  $n$  distinct elements and  $x$  is an event in the set, then the probability of  $x$  happening in  $W$  is shown in (1):

$$p_i = \frac{x_i}{n} \quad (1)$$

To measure the entropy (labelled by  $H$ ), the sum of the probability of occurrence of all elements in the set is calculated as shown in (2):

$$H = -\sum_{i=1}^n p_i \log p_i \quad (2)$$

where  $W$  denotes the window size. In our case we group the packet headers into equal sets (windows) and inspect the relevant header fields. To detect a DDoS attack the calculated instantaneous values of entropy has to be compared to a predefined threshold. In the current project we analyze the obtained entropy values and determine appropriate threshold for each test scenario. If the entropy is less than this threshold and it lasts for a minimum of 3 entropy periods in a row, it will be considered as an attack.

#### IV. SIMULATION SETUP AND RESULTS

The mechanism proposed in this paper aims to offer a solution for early detection of DDoS attacks and therefore we examine a set of instant traffic characteristics, that are not dependable on the node mobility. In order to do this we consider a very small window size of 30 packets. Furthermore, we consider topologies with 30 or more hosts connected to it. Every 30 Packet\_In messages will be parsed for their



destination IP address and the entropy of that list of packets will be computed. For the needs of the current experiment we test two different topologies in order to cover various scenarios where the RSUs and the vehicle act as either hosts or switches.

**A. Simulation Environment**

The experiments include two big test cases, each of which consists of three sub-cases of attacks and a normal traffic. One additional setup examines how the entropy of the destination IP address will change if the attack is launched by a Botnet of 3 hosts. Both the normal, and the attack traffic generation is realized by running two or more parallel Scapy programs. Scapy [16] is a very powerful packet manipulation tool that is used to generate the UDP packets and to spoof their source IP addresses. The function that is generating the attack sends the packets faster than the one, which generates normal traffic. Three different intensity attacks are run on one host and one intensity attack on 6 hosts, connected to the same switch and subnet. Normal traffic is sent to all the switches with randomly generated packets going to all hosts. The codes that generated normal and attack traffic were started manually. Mininet [17] is the network emulator used for running the experiments. For test cases with attack traffic, one host is randomly chosen, to launch the attack to one or to a subnet of 6 hosts. The destination port is set to be 80, the type of the attack is DDoS and all the traffic packet will be UDP. The packet header is only sent to the controller by default in OpenFlow. Several popular controllers for SDN networks are available. The one that is used in the current project is POX [18]. The L3\_learning module of the POX controller was modified by adding two main functions for statistic collection and entropy computation.

**B. Experiments and Related Topologies**

The first experiment, presented on fig. 1, is based on a tree topology of depth two with 7 switches and 36 hosts.

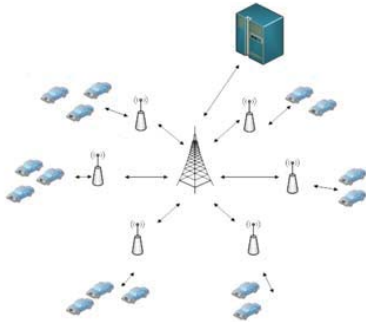


Fig. 1. Network topology for test case 1

This setup represents a real VANET scenario, where the RSUs and the base stations are static, whereas the hosts move from one place to another, resulting in periodic topology changes. Here, the RSUs can be regarded as entry points to the underlying SDN network infrastructure and therefore are

simulated as SDN switches. In order to design a reliable mechanism for early detection of DDoS attacks, we calculate the entropy values for a set of destination IP addresses under normal and attack conditions in order to determine appropriate threshold values for the entropy.

In the second experiment (see, Fig. 2) we assume that the RSUs are part of the data plane communication and therefore are represented as hosts. The vehicles, therefore, communicate directly with the base station, which is simulated as an SDN switch. The number of the vehicles is chosen to be 40, the number of the RSUs to be 4 and the base station is connected directly to the controller. For this test case a central control mode was chosen [19].

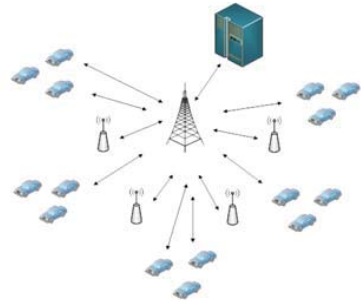


Fig. 2. Topology for the second experiment

The third experiment involves both of the above mentioned topologies by examining the entropy values of the destination IP addresses, when the corresponding networks are under botnet attacks.

**C. Results**

Three sub-test cases for each experiment were defined – one estimating the entropy under normal conditions, one calculating the entropy fluctuations when one network host is under attack and one when 6 hosts are being attacked. For each of the sub-test cases five runs were performed with 750 packets (25 windows with 30 packets). The normal traffic interval was set to 0.1 seconds. The source IP addresses were spoofed. In the sub-cases, where the network was under attack on one host, the used attack traffic intervals were 0.03 and 0.05 seconds. In the experiment for the attack on 6 hosts we used an attack traffic interval of 0.03 seconds. For experiment 3 we used an intensity of the attack traffic of 0.05 seconds.

To detect DDoS attack the proposed detection solution requires a threshold. The implemented a solution is based on which if the entropy is lower than the threshold, and it persists for 3 windows in a row, an attack is in progress. The threshold for each experiment is defined based on the results of the entropy. We decided these windows to be 3, because this work is interested in really small time interval, due to the dynamic nature of the VANET network.

Fig. 3 shows the fluctuations of entropy of the destination IP addresses under normal conditions for experiment 1 (sub-test case 1).



Fig. 3. Entropy in normal circumstances for experiment 1

The second sub-test case examines the entropy in network under attack on one host with two different rates of packet generation. The third sub-test case evaluates the entropy value for network under attack on a subnet of 6 hosts. Information about the entropy in test case 1 is collected, after performing the experiment and the results are summarized in Fig. 4.

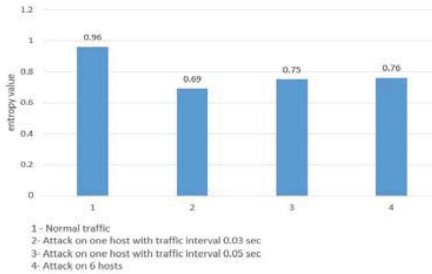


Fig. 4. Average entropies obtained from experiment 1

According to Fig. 4 we can notice that the difference between the entropy of normal traffic and the highest entropy of attack traffic is 0.2. For the purpose of our work, we are choosing the threshold to be 0.76. Only then we can be sure that our solution will detect the attack if it occurs and the entropy is lower than 0.76 and it lasts for at least 3 windows.

Experiment 2 follows the same pattern, covering the topology, depicted in Fig 2. The obtained information about the entropy is presented in Fig. 5 and 6.



Fig. 5. Entropy in normal circumstances for experiment 2.

Comparing to the first experiment here the entropy of destination IP addresses for normal traffic is higher – 1.29. The reason for that is that the window size was kept to be 30, while as we included more hosts to our network, the diversity between the IP addresses increased. The code for packet generation can choose to send the traffic between more nodes in the same window size. According to Fig. 4 and Fig. 6 we can notice that the maximum value of entropy for network under attack is really close to the value of the entropy for normal traffic in the first test case. Therefore, we can conclude, it is very important to find and adjust the threshold according to the current network topology. That was one of the goals of our experiments, to show how in dynamic network topology the parameters have to be carefully considered, because constant values are not applicable for VANET scenario. According to that we are choosing the threshold for the second experiment to be 0.94. That corresponds to drop in the entropy with 0.35.

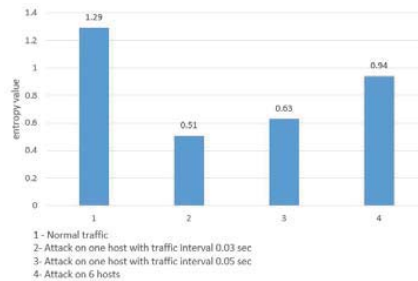


Fig. 6. Average entropies obtained from experiment 2

The impact of DDoS attacks can grow greatly, if the attackers take advantage of botnet to overwhelm their victim's network. Experiment 3 is focused on determining how the network will be influenced by attack launched via a botnet. We divided the experiment into 2 sub-test cases; the first sub-test case is performed using the topology of Fig.1 and the second using the topology from Fig. 2. The tests included an examination of the entropy, when the network is under attack on one and a subnet of 6 hosts with a traffic interval of 0.05 seconds for each topology. All the attacks were launched by a botnet of 3 infected nodes. The results of the experiments on the first topology are presented in Fig. 7, 8 and 9. Fig. 7 and Fig. 8 show the fluctuations of the entropy when an attack occurred in the third window.

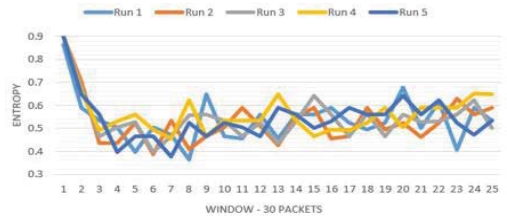


Fig. 7. Entropy fluctuations in network under botnet attack on one host – topology 1

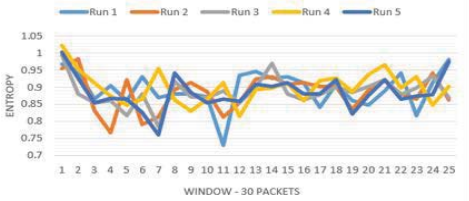


Fig. 8. Entropy fluctuations in network under botnet attack on 6 hosts – topology 1.

Examining the results, obtained in experiment 3 for the first topology, it can be noticed that when the attack is deployed on one host the value of the entropy decreased almost with  $\frac{1}{2}$  compared to the entropy value of normal traffic. In case of attack on a subnet of 6 hosts, as expected, the results have higher value, because of the higher diversity of destination IP addresses. Subsequently, a really small drop in the entropy was observed comparing to the value in normal circumstances. The observed drop was only 0.07. As a result, we can conclude that the use of entropy-based detection, considering 3 consecutive windows of size 30 packets cannot provide reliable detection under botnet attack conditions on topology 1.

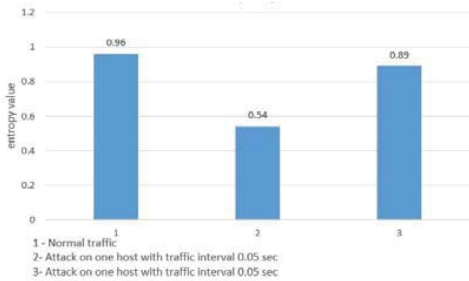


Fig.9. Average entropies obtained by the tests on the first topology under botnet attack

The same experiments were performed for the second topology and the results are summarized in Fig. 10, 11, 12.



Fig. 10. Entropy fluctuations in network under botnet attack on one host – topology 2

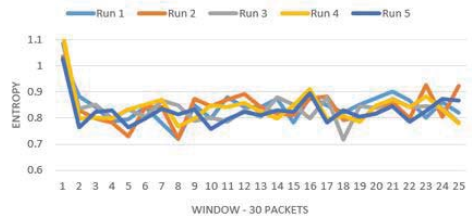


Fig. 11. Entropy fluctuations in the network under botnet attack on 6 hosts- topology 2.

In order to improve this result a bigger windows size with a larger number of required consecutive windows under the threshold should be used. Anyway, we can adjust the threshold to be 0.89, if we are aware of the nature of the attack, i.e. if we know that attempts of DDoS attacks launched by a botnet can occur in the network. In our test case that is not very reliable, because if we assume the threshold to be 0.89, some incorrect detections can occur, due to the fact the average entropy for attack on one host, that we obtained was 0.54.

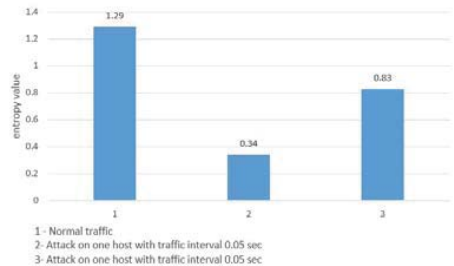


Fig. 12. Average entropies obtained by tests on the second topology under botnet attack

The observed drop of entropy for the second topology, however, was much greater and therefore would allow reliable detection. This leads to the conclusion that the entropy values can vary greatly depending on the current topology and therefore dynamic selection of the detection algorithm parameters is required. We can choose the threshold to be 0.83 by taking into account the same assumptions that we made for the first topology.

## V. CONCLUSION

Since the security in the networking, can be consider as a major concern, protecting the operating system of SDN-based VANET architecture by detecting a DDoS attack was the center of this work. Early detection of the occurrence of attack was crucial point in our architecture, due to the fact that if the SDN controller was compromised, the whole network could become unreachable. Furthermore, for our VANET test scenario, which includes applications that are working with life critical information, the importance of security is even more significant.

In this paper, we directed our efforts to a solution, which works extremely good for SDN scenario, based on its

specifications, standards, points of strength and limitations. We wanted to implement it in SDN-based VANET architecture and to check if it will work properly with different topologies in order to provide a basis of a solution for DDoS attack detection in dynamic environments. We made a precise definition of each parameter that we used. We decided to test the solution in three different test cases, using two SDN-based VANET topologies. Based on traffic features, such as destination IP addresses and following a sequence of 3 windows of 30 packet, the entropy was used to measure the degree of randomness of the packets that are received by the SDN controller. By computing the entropy and after appropriate definition of the threshold for each case, we were capable to detect an attack on one host or a subnet of 6 hosts and when the attack is launched by botnet of 3 hosts. Definition of different thresholds was needed for each test case, since we examined two different topologies that simulate dynamic VANET scenario. We believe that based on our test results, a dynamic model, which adjust its parameters instantly after obtaining current status from the network, can be developed. The SDN controller can take into account the number of switches and hosts that are connected to it and according to that current topology, to change its threshold.

#### VI. ACKNOWLEDGMENTS

We would like acknowledge the valuable advice of Dr Neeli R. Prasad in relation to the open VANET challenges.

#### REFERENCES

- [1] "Software-defined networking: The new norm for networks," Palo Alto, CA, USA, White Paper, Apr 2012
- [2] Q. Yan; F. R. Yu, "Distributed denial of service attacks in software-defined networking with cloud computing," *IEEE Communications Magazine*, 2015, Volume: 53, Issue: 4, Pages: 52 - 59,
- [3] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, "SDN Security: A Survey," *Proc. IEEE SDN for Future Networks and Services (SDN4FNS)*, 2013
- [4] C. Ji M. Thottan, "Anomaly Detection in IP Networks," *IEEE Transaction on Signal Processing*, vol. 51, no. 8, pp. 2291-2204, Aug 2003.
- [5] D. Schnackenberg, et al, "Statistical Approaches to DDoS Attack Detection and Response," in *DARPA Information Survivability Conference and Expedition*, vol. 2003, Apr.
- [6] T. Nakashima, T. Sueyoshi S. Oshima, "Early DoS/DDoS Detection Method using Short-term Statistics," in *International Conference on Complex, Intelligent and Software Intensive Systems*, 2010, pp. 168-173.
- [7] M. Nidhal MEJRI, J. BEN-OTHTMAN, "Entropy as a New Metric for Denial of Service Attack Detection in Vehicular Ad-hoc Network/ data plane communication", C.2.0 [COMPUTER-COMMUNICATION NETWORKS]: General, Security and Protection; C.2.1 [COMPUTER-COMMUNICATION NETWORKS]: Network Architecture and Design, Network Communications
- [8] S. M. Mousavi, "Early Detection of DDoS Attacks in Software Defined Networks Controller", Carleton University Ottawa, Ontario, 2014
- [9] Ian Ku, You Lu, Mario Gerla, Francesco Ongaro, Rafael L. Gomes, Eduardo Cerqueira, *Towards Software-Defined VANET: Architecture and Services*, 2014 13th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)
- [10] D. Kreutz, F. M. Ramos, and P. Verissimo, "Towards secure and dependable software defined networks," in *Proc. 2nd ACM SIGCOMM Workshop Hot Topics Softw. Defined Netw.*, 2013, pp. 55-60.
- [11] R. Kloti, "OpenFlow: A security analysis," M.S. thesis, Dept. Inf. Tech. Elec. Eng., Swiss Fed. Inst. Technol. Zurich (ETH), Zurich, Switzerland, 2013.
- [12] P. Tyagi and D. Dembla, "Investigating the security threats in Vehicular ad hoc Networks (VANETs): Towards security engineering for safer on-road transportation," *Advances in Computing, Communications and Informatics (ICACCI)*, 2014 International Conference on, New Delhi, 2014, pp. 2084-2090
- [13] M. Todorova, S. Todorova, "DDoS attack detection in SDN-based VANET architectures", *ICTE Master Thesis*, CTIF, Aalborg University, 2016
- [14] S. Sezer et al., "Are We Ready for SDN? Implementation Challenges for Software-Defined Networks," *IEEE Commun. Mag.*, vol. 51, no. 7, 2013.
- [15] N. Hoque, D. K. Bhattacharyya and J. K. Kalita, "Botnet in DDoS Attacks: Trends and Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2242-2270, Fourthquarter 2015.
- [16] Available: <http://www.secdev.org/projects/scapy/>
- [17] Available: <http://mininet.org/>
- [18] M. McCauley, Available: <http://www.noxrepo.org/>, November, 2013
- [19] I. Ku, et al., "Towards Software-Defined VANET: Architecture and Services," 2014 13th Ann. Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET).

# Coordinated Hybrid Precoding for 3D Millimeter Wave Massive MIMO in 5G H-CRAN

Jialin Cai\*, Songlin Sun<sup>†</sup>, Bo Rong<sup>‡</sup>,

\*<sup>†</sup>Key Laboratory of Trustworthy Distributed Computing and Service (Beijing University of Posts and Telecommunications), Ministry of Education,

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China

<sup>‡</sup>Communications Research Centre, Canada

**Abstract**—In this paper, we study the hybrid digital-analog precoding designs for 3D Millimeter Wave Massive multiple-input multiple-output (MIMO) Systems in 5G heterogeneous cloud radio access networks (H-CRAN). To overcome the great radio resource management (RRM) problem caused by massive antenna coordination, we utilize H-CRAN to deal with spatial domain information centrally and arrange flexible coordination for the whole system. In the proposed algorithm, the baseband unit (BBU) pool of H-CRAN processes the user channel state information (CSI) centrally and form the hybrid precoding matrix based on null-space of victim user channels. Simulation results show that our design can effectively alleviate interference between network tiers and support a high quality of service (QoS) accommodation.

**Index Terms**—5G H-CRAN, massive MIMO, hybrid precoding, mmWave

## I. INTRODUCTION

The fifth-generation (5G) network is expected to provide diversified terminals with enhanced data rate and support intelligent RRM [1]. To improve the system capacity, operators promote the heterogeneous networks (HetNets) architecture using densified low power nodes (LPNs) to provide the nearby mobile users with high data rate by frequency reuse [2] [3]. However, too dense LPNs incur severe interference to Hetnets and impact the network performance [4]. To deal with this problem, H-CRAN is proposed to provide central coordination in HetNets, where the central BBU pools take over most base station functionalities and the remote radio heads (RRHs) just perform the remaining radio frequency (RF) functionality [5]. In this way, the RRM for LPNs and the baseband data processing are moved to BBU pools taking advantage of cloud computing.

Another two promising technologies, Massive MIMO and Millimeter Wave (mmWave), are also regarded as solutions for the future 5G mobile system [6]. Recently, to make better use of the space at base stations, academia suggest packing two-dimensional (2D) antenna plane into three-dimensional (3D) massive MIMO systems. Working on small wavelength, mmWave communication technology enables 3D massive MIMO systems owing to its ability to integrate large amounts of antenna elements into a small scale.

Project 61471066 supported by NSFC.

Generally, traditional precoding schemes are carried out by a completely digital precoder which can theoretically achieve optimal performance. However, it brings enormous hardware complexity and energy consumption as it requires an expensive RF chain for every antenna. Obviously, it will become a serious problem in mmWave systems due to the huge number of antennas at the base station (BS). Fortunately, academia has proposed a hybrid precoding technique recently, which combines digital and analog precoding to significantly reduce the number of RF chains. This technique is considered as a potential solution for the practical application of mmWave massive MIMO systems [7] [8].

In this paper, we focus on novel signal processing methods for 3D mmWave Massive MIMO system with the aid of H-CRAN architecture to enable future 5G networks. We further design the H-CRAN structure to manage huge spatial domain resource brought by the Massive MIMO system and provide coordinated information processing for the whole network. Based on the designed H-CRAN architecture, we proposed the null-space based hybrid (NS-Hybrid) precoding scheme for the 3D mmWave Massive MIMO system, which aims to mitigate the inter-tier interference. In the proposed architecture, BBU pool takes the work of integrating huge spacial domain information and performing null space calculation concerning massive MIMO coordination. Therefore, the null space information can be integrated to enable the whole network coordination and thus mitigate the interference between each RRH.

The rest of this paper is organized as follows. Section II presents the 5G H-CRAN architecture supporting RRM and Section III introduces the 3D mmWave Massive MIMO channel model. Section IV proposes our two tier interference free mmWave hybrid precoding scheme. Section V presents the performance evaluation of the proposed scheme. In the end, Section VI concludes the paper.

## II. H-CRAN AIDED 5G ARCHITECTURE

Hetnets and 3D mmWave massive MIMO improve the network capacity as well as bringing a huge amount of RRM and coordination problem [9]. In order to utilize these

techniques technology well, we introduce H-CRAN to aid the 5G network.

H-CRANs combines HetNets with C-RANs. H-CRAN simplifies the LPNs in Hetnets to just RRH and establishes a central unit to perform data processing and coordination for each of them. In 5G H-CRAN, BBUs of LPNs are centralized together to form a BBU pool collecting and managing spatial information for the whole network. The BBU pools and RRHs are commonly connected by optical fibers in order to maintain low latency between them. The BBU pool performs most processing and computing procedures, such as data processing and RRM. While RRH is only responsible for radio functionalities like power amplification.

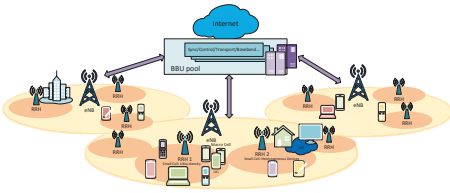


Fig. 1. A typical H-CRAN system

To process the huge amounts of spatial domain information brought by 3D massive MIMO in HetNets, we propose an H-CRAN architecture to handle the RRM and massive MIMO coordination for 5G network. In the proposed architecture, eNBs and RRHs collect the user information and report it to the BBU pool through backhaul links and fibers respectively. Then BBU pool will carry out information processing such as analyzing CSI and calculating the null space. After that, it returns the processed information and instructions to eNB/RRHs to get better coordination by concerning the global knowledge of network spatial information. The BBU pool takes over most of the information processing work from eNB and LPNs to achieve a more effective network.

Compared to traditional LTE systems, by employing H-CRAN technology, the 5G network can achieve high scalability and flexibility, as the BBU pool gathers whole network information.

### III. CHANNEL MODEL AND ANTENNA ARRAY CONFIGURATION

In this section, we introduce the 3D mmWave Massive MIMO channel model used in our work.

#### A. 3D Massive MIMO Channel Model

Assume a downlink mmWave Massive MIMO system where the access point is equipped with  $N$  antennas serving  $K$  single-antenna users,  $N, K \in \mathbb{N}_+$ ,  $N \geq K$ . Assuming non-line-of-sight (NLOS) scenario, the channel model can be described as [10]

$$\mathbf{H}^H = \mathbf{G}^H \mathbf{D}, \quad (1)$$

where  $\mathbf{H}^H = [\mathbf{H}_1^H, \mathbf{H}_2^H, \dots, \mathbf{H}_K^H]$  denotes the channel matrix, where  $\mathbf{H}_k \in \mathbb{C}^{1 \times N}$  ( $k = 1, 2, \dots, K$ ) is the channel between the access point and the  $k$ th user equipment (UE).  $\mathbf{D} = \text{diag}(\sqrt{\beta_1}, \sqrt{\beta_2}, \dots, \sqrt{\beta_K})$  is the large-scale propagation matrix, in which  $\beta_k = \kappa d_k^{-\gamma} \xi_k$ .  $\kappa$  is a constant value determined by the carrier frequency and antenna characteristic,  $d_k$  is the distance between the access point and the  $k$ th UE,  $\xi_k$  denotes the log-normal shadow fading factor, and  $\gamma$  is the path loss exponent.

Compared to tradition MIMO antenna array, the number of antennas rises to a larger scale in 3D mmWave Massive MIMO channel model. Thus we should consider the mutual impedance of adjacent antennas. In the mutual coupling channel model,

$\mathbf{G}^H = [\mathbf{G}_1^H, \mathbf{G}_2^H, \dots, \mathbf{G}_K^H]$  denotes the fast fading matrix, and the vector  $\mathbf{G}_k \in \mathbb{C}^{1 \times N}$  ( $k = 1, 2, \dots, K$ ) is described by the equation follows [11]:

$$\mathbf{G}_k^H = \mathbf{Z} \mathbf{R}_k \mathbf{v}_k. \quad (2)$$

where  $\mathbf{R}_k$  is steering matrix and  $\mathbf{v}_k$  denotes the Gaussian stochastic factor.  $\mathbf{Z} \in \mathbb{C}^{N \times N}$  is the constant mutual coupling matrix concerned with antenna configuration and has the following form:

$$\mathbf{Z} = (\mathbf{Z}_A + \mathbf{Z}_L)(\mathbf{\Gamma} + \mathbf{Z}_L \mathbf{I})^{-1}, \quad (3)$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} Z_A & Z_M & 0 & \dots & 0 \\ Z_M & Z_A & Z_M & \dots & 0 \\ 0 & Z_M & Z_A & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & Z_M & Z_A \end{bmatrix}. \quad (4)$$

The complex value of  $Z_{L,Z_A}$  and  $Z_M$  denote the load impedance, antenna impedance, and mutual impedance respectively.

#### B. Antenna Array Configuration

The form of  $\mathbf{R}_k$  and  $\mathbf{v}_k$  depend on the antenna array configuration [10]. In the linear antenna array scenario, suppose the  $k$ th user has  $A_k$  different angles of arrivals (AoAs) expressed by  $\theta_{k,l}$  ( $l = 1, 2, \dots, A_k$ ). Here  $\mathbf{R}_k$  has  $A_k$  column of steering vectors, i.e.  $\mathbf{R}_k = [\mathbf{r}_{k,1}, \mathbf{r}_{k,2}, \dots, \mathbf{r}_{k,A_k}]$ , where

$$\mathbf{r}_{k,l} = \mathbf{a}(\theta_{k,l}) = \frac{1}{A_k} \left[ 1, e^{\frac{j2\pi d}{\lambda} \sin \theta_{k,l}}, e^{\frac{j2\pi d}{\lambda} 2 \sin \theta_{k,l}}, \dots, e^{\frac{j2\pi d}{\lambda} (N-1) \sin \theta_{k,l}} \right]^T. \quad (5)$$

$d$  is the adjacent antenna distance, and  $\lambda$  stands for the wave length. We define  $\mathbf{a}(\theta)$  as the steering vector function for later convenience.

While in the rectangular antenna array scenario, the  $k$ th user has  $A_k$  different azimuths of arrival (AoAs) and elevations of arrival (EoAs) written as  $\theta_{k,l}$  and  $\phi_{k,l}$  respectively. Then the column steering vectors in  $\mathbf{R}_k = [\mathbf{r}_{k,1}, \mathbf{r}_{k,2}, \dots, \mathbf{r}_{k,A_k}]$  is described by:

$$\begin{aligned} \mathbf{r}_{k,l} &= A_k^2 \times \text{vec} \left[ \mathbf{a}(\theta_{k,l}) \otimes \mathbf{a}(\phi_{k,l})^T \right] \\ &= A_k^2 \times \text{vec} \left[ 1, e^{\frac{j2\pi d}{\lambda} \sin\theta_{k,l}}, e^{\frac{j2\pi d}{\lambda} 2\sin\theta_{k,l}}, \dots, e^{\frac{j2\pi d}{\lambda} (N-1)\sin\theta_{k,l}} \right] \quad (6) \\ &\otimes \left[ 1, e^{\frac{j2\pi d}{\lambda} \sin\phi_{k,l}}, e^{\frac{j2\pi d}{\lambda} 2\sin\phi_{k,l}}, \dots, e^{\frac{j2\pi d}{\lambda} (N-1)\sin\phi_{k,l}} \right], \end{aligned}$$

where  $\otimes$  is the Kronecker product of matrices, and the  $\text{vec}(\cdot)$  function represents vectorization of matrix.

With the 3D channel model above, we present our proposed precoding scheme in the next section.

#### IV. TWO TIER INTERFERENCE FREE MMWAVE HYBRID PRECODING SCHEME

This section introduces our proposed precoding scheme to implement effective RRM in mmWave 5G network. In this design, we use NS-Hybrid precoding to overcome the hardware constraints and mitigate inter-tier interference in mmWave H-CRAN.

##### A. System Model

We assume a downlink 3D mmWave Massive MIMO system with  $K$  single-antenna users, where the eNB and RRHs all contain  $N$  transmit antennas, as shown in Fig.2.

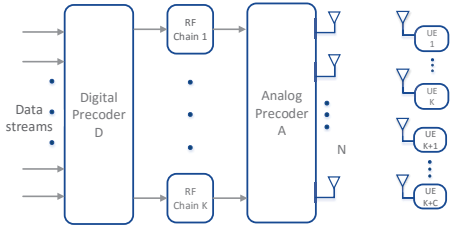


Fig. 2. Hybrid mmWave Precoding Structure

$K_j$  stands for the number of users of the  $j$ -th RRH and  $C_j$  is the number of victim users interfered by the  $j$ -th RRH and served by other access points.  $K_j$  and  $C_j$  should meet the constraint  $K_j + C_j \leq N$ .

The channel matrix  $\mathbf{H}_j \in \mathbb{C}^{(K_j+C_j) \times N}$  between RRH  $j$  and the users it serves is

$$\mathbf{H}_j = [\mathbf{H}_{j,1}, \dots, \mathbf{H}_{j,K_j}, \mathbf{H}_{j,K_j+1}, \dots, \mathbf{H}_{j,K_j+C_j}]^H \quad (7)$$

where  $\mathbf{H}_{j,k} \in \mathbb{C}^{1 \times N}$  is the channel between the  $j$ -th RRH and the  $k$ -th user. particularly,  $\mathbf{H}_0$  denotes the channel matrix between eNB and the users it serves. Correspondingly,  $\mathbf{H}_{0,k}$  is the channel between the eNB and the  $k$ -th user.

As shown in Fig. 3, in our proposed CSI acquisition method, BBU pool collects the CSI information of all the RRHs, i.e. receiving every  $[\mathbf{H}_{j,K_j+1}^H, \dots, \mathbf{H}_{j,K_j+C_j}^H]^H$  corresponding to the  $j$ th

RRH. The BBU pool then performs singular value decomposition (SVD) algorithm and generates null-space vector  $\tilde{\mathbf{V}}_{j,v}^{(0)}$  for RRH  $j$ . After that, it will use the null-space vector  $\tilde{\mathbf{V}}_{j,v}^{(0)}$  to design precoding matrix based on the hybrid precoding algorithm in the next subsection. At last, the BBU pool sends each RRH the corresponding precoding matrix through downlink information. Thus the BBU pool carries out all the matrix decomposition calculating and precoding work.

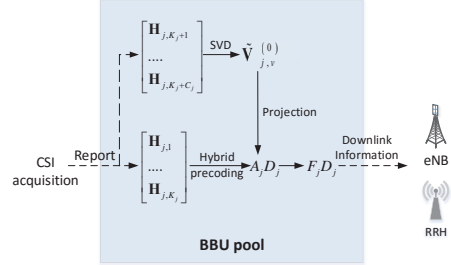


Fig. 3. Proposed CSI acquisition methods

##### B. NS-Hybrid Precoding

The hybrid precoding matrix is consisted of digital precoding  $\mathbf{D} \in \mathbb{C}^{K \times K}$  and analog precoding  $\mathbf{A} \in \mathbb{C}^{N \times K}$ . Each entry of  $\mathbf{A}$  is normalized to satisfy  $|\mathbf{A}_{a,b}| = \frac{1}{\sqrt{N}}$ , where  $|\mathbf{A}_{a,b}|$  denotes the magnitude of the  $(a, b)$ th element of  $\mathbf{A}$ . We perform the analog precoding according to  $\mathbf{A}_{a,b} = \frac{1}{\sqrt{N}} e^{j\varphi_{a,b}}$ , where  $\varphi_{a,b}$  is the phase of the  $(a, b)$ th element of  $\mathbf{H}^H$ , the conjugate transpose of the channel matrix. After the analog precoding, we can observe an equivalent channel  $\mathbf{H}_{eq} = \mathbf{H}\mathbf{A}$  of a low dimension  $K \times K$  at the baseband. Then ZF precoding is used in digital domain, thus the digital precoding matrix is designed as  $\mathbf{D} = \mathbf{H}_{eq}^H (\mathbf{H}_{eq} \mathbf{H}_{eq}^H)^{-1}$ .

For the  $k$ -th user in RRH  $j$ , the received signal can be written as

$$y_{jk} = \mathbf{H}_{j,k} \mathbf{A}_j \mathbf{D}_j \mathbf{s}_j + \mathbf{H}_{0,k} \mathbf{A}_0 \mathbf{D}_0 \mathbf{s}_0 + n_{j,k}, \quad (8)$$

where  $\mathbf{s}_j \in \mathbb{C}^{K_j \times 1}$  is the transmitted signal vector for the overall  $K_j$  users of RRH  $j$ .  $\mathbf{s}_0$  is the transmitted signal vector at eNB. Here  $\mathbf{D}_j \in \mathbb{C}^{K_j \times K_j}$  and  $\mathbf{A}_j \in \mathbb{C}^{N \times K_j}$ .  $n_{j,k}$  denotes the additive white Gaussian noise of zero mean and variance  $\sigma^2$ .

The complementary space concerning the victim users for RRH  $j$  is given as

$$\tilde{\mathbf{H}}_{j,v} = [\mathbf{H}_{j,K_j+1}^H \dots \mathbf{H}_{j,K_j+C_j}^H]^H, \quad (9)$$

where  $\tilde{\mathbf{H}}_{j,v}$  includes these victim users' channel matrixes. To eliminate the interference to victim users, the precoding matrix should satisfy the conditions below

$$\tilde{\mathbf{H}}_{j,v}^H \mathbf{A}_j \mathbf{D}_j = \mathbf{0}^{C_j \times K_j}. \quad (10)$$

After performing SVD method,  $\tilde{\mathbf{H}}_{j,v}$  can be further expressed as

$$\begin{aligned}\tilde{\mathbf{H}}_{j,v} &= \tilde{\mathbf{U}}_{j,v} \tilde{\mathbf{\Lambda}}_{j,v} \tilde{\mathbf{V}}_{j,v}^H \\ &= \tilde{\mathbf{U}}_{j,v} \left[ \sum_{j,v} \mathbf{0} \right]_{C_j \times N} \left[ \tilde{\mathbf{v}}_{j,v,1} \tilde{\mathbf{v}}_{j,v,2} \cdots \tilde{\mathbf{v}}_{j,v,N} \right]^H.\end{aligned}\quad (11)$$

Define the null-space vector as  $\tilde{\mathbf{V}}_{j,v}^{(0)} = [\tilde{\mathbf{v}}_{j,v,C_j+1} \cdots \tilde{\mathbf{v}}_{j,v,N}]$ . We will have  $\tilde{\mathbf{H}}_{j,v} \tilde{\mathbf{V}}_{j,v}^{(0)} = \mathbf{0}$ , as the column vectors belonging to  $\tilde{\mathbf{V}}_{j,v}^{(0)}$  locate in the null-space of all victim users.

For RRH  $j$ , we adapt a projection matrix  $\mathbf{M}_j$  constituted by the null-space as

$$\mathbf{M}_j = \tilde{\mathbf{V}}_{j,v}^{(0)} (\tilde{\mathbf{V}}_{j,v}^{(0)})^H, \quad (12)$$

After projection, the new analog precoding matrix becomes

$$\mathbf{F}_j = \mathbf{M}_j \mathbf{A}_j, \quad (13)$$

Define  $P_j$  as the transmit power at the  $j$ -th RRH which satisfies  $E[\mathbf{s}_j \mathbf{s}_j^H] = \frac{P_j}{K_j} \mathbf{I}_{K_j}$ , further, we normalize  $\mathbf{F}_j$  to satisfy  $\|\mathbf{F}_j \mathbf{d}_j\|_F^2 = K_j$  for the total transmit power constraint.

The SINR of the  $k$ -th user served by RRH  $j$  is calculated as

$$SINR_{j,k} = \frac{\frac{P_j}{K_j} |\mathbf{H}_{j,k} \mathbf{F}_j \mathbf{d}_{j,k}|^2}{\sigma_{j,k}^2 + \sum_{r \neq k} \frac{P_r}{K_r} |\mathbf{H}_{j,k} \mathbf{F}_r \mathbf{d}_{j,r}|^2} \quad (14)$$

where  $\mathbf{d}_{j,r}$  stands for the  $r$ th column of  $\mathbf{D}_j$ .

To meet each user's QoS requirement, the SINR should satisfy

$$B \log(1 + SINR_{j,k}) \geq R_{j,k}, \quad (15)$$

for every  $j = 1, 2, \dots, J$  and  $k = 1, 2, \dots, K_j$ , where  $R_{j,k}$  is the data rate demanded by the  $k$ -th user in RRH  $j$  and  $B$  is the total bandwidth.

## V. SIMULATION RESULTS

This section presents the numerical results of NS-Hybrid precoding scheme based on the simulations in 3D mmWave Massive MIMO system in H-CRAN scenario.

The network topology combines mmWave system and H-CRAN architecture and our proposed algorithm aims to achieve target QoS with low complexity. We use the NS-Hybrid precoding to reduce hardware complexity in mmWave systems and mitigate the interference to the victim users in Hetnets.

Fig. 4 shows the simulation scenario where the intended and victim users are in red and green respectively. We will analyze the system performance over different simulation parameters. Table I presents main simulation parameters which characterize the communication network.

In the following text, we will introduce the simulation results in the scenario described in Fig. 4, where each user has QoS demand of 20Mbps. In the proposed scheme, BBU pool performs the null-space construction and transmits the

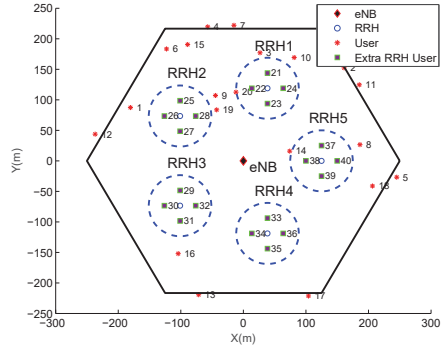


Fig. 4. The topology of the H-CRAN

TABLE I  
SIMULATION PARAMETERS

Parameters	Values
Antennas of the eNB and RRHs	$N \in \{30, 40, 50, \dots, 90\}$
Antenna Interval for Linear Antenna Array	0.5 wave length
Antenna Interval for Rectangular Antenna Array	0.5 wave length
$Z_A, Z_L, Z_M$	50 $\Omega$
SNR	-10dB to 30dB
Frequency	38GHz
Tx Rx Antenna Height	36m, 1.5m
Tx Rx Antenna Gain	25/13.3
Scenario	NLOS

corresponding precoding vectors to the RRHs as described in section V. For comparison, we also simulated two null-space based linear precoding methods, null-space based ZF (NS-ZF) precoding and null-space based max ratio transmission (NS-MRT) precoding. Theoretically, NS-ZF precoding will have the optimal performance.

### A. System throughput under different number of antennas

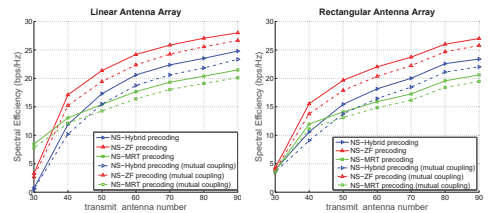


Fig. 5. System throughput under different number of antennas



Fig. 5 illustrates the system throughput under different number of antennas. Here we compare our proposed NS-Hybrid precoding scheme to NS-ZF and NS-MRT precoding schemes in 3D mmWave systems with linear antenna array and rectangular antenna array respectively. All the precoding schemes apply the null-space method to mitigate inter-tier interference. As the antennas increase, Massive MIMO performs better in system throughput and the increased antennas can eliminate the interference to the victim users.

Simulation results show that the NS-Hybrid precoding scheme can gain considerable throughput comparing to the optimal NS-ZF precoding and overwhelms NS-MRT as antennas increase. When considering mutual coupling, the performance of all the schemes decrease, but the relative position of the three precoding schemes remain the same. It means that our proposed NS-Hybrid precoding scheme can support comparable performance at certain complexity reduction.

B. Spectral efficiency under different SNRs

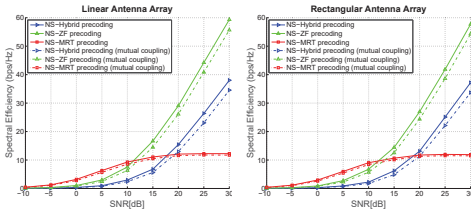


Fig. 6. Spectral efficiency under different SNRs of users

In Fig. 6, we compare the spectral efficiency of NS-Hybrid precoding scheme with NS-ZF and NS-MRT precoding schemes in different SNR conditions. We also do the simulation in 3D mmWave systems with linear antenna array and rectangular antenna array respectively, as well as considering the mutual coupling. Simulation results show that the NS-Hybrid precoding scheme gets higher spectral efficiency than the NS-MRT precoding. This indicates the proposed NS-Hybrid precoding scheme will be suitable for realistic 3D mmWave system due to its practicality and comparable spectral efficiency.

C. SINR distribution under different number of extra antennas

In Fig. 7, we evaluate the SINR distribution of the proposed NS-Hybrid precoding scheme and the conventional hybrid precoding. It can be seen that the proposed scheme gets better SINRs indicating an improved system performance. Such difference between the two schemes is due to the fact that the victim users suffer strong interference from other nodes in the conventional scheme. However, our proposed hybrid precoding scheme utilizes the extra antennas to indicate victim users and project its own users to the corresponding null space. In this way, the strong interference to the victim users are eliminated and the SINRs of the victim users will

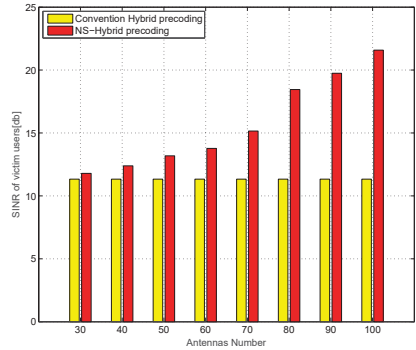


Fig. 7. Effect of the extra antennas

increase. More extra antennas can indicate more victim users and therefore eliminate more interference in the system. It means that deploying more extra antennas contributes to higher SINRs significantly.

VI. CONCLUSION

In this paper, we have studied the precoding issues in 3D mmWave Massive MIMO Systems in H-CRAN, concerning the spatial domain resource management and interference mitigation. We have further designed the H-CRAN architecture to provide coordinated RRM for the whole network and proposed the NS-Hybrid precoding scheme with lightweight CSI acquisition method. Simulation results show that our work can significantly decrease the impact on neighboring victim users in RRH-covered small cells. The proposed design can support sufficient network capacity and improve user SINRs. It also has an advantage in practice as it overcomes the physical constraints in the mmWave system.

REFERENCES

- [1] Li, Bin, et al. "Deep sensing for future spectrum and location awareness 5G communications." *IEEE Journal on Selected Areas in Communications* 33.7 (2015): 1331-1344.
- [2] R. Q. Hu, Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems." *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94-101, May. 2014.
- [3] R. Q. Hu, Y. Qian, "Heterogeneous cellular networks." *John Wiley and Sons*, 2013.
- [4] Sun, Songlin, et al. "An intelligent SDN framework for 5G heterogeneous networks." *IEEE Communications Magazine* 53.11(2015):142-147.
- [5] Peng, Mugen, et al. "Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies." *IEEE Wireless Communications* 21.6(2014):126-135.
- [6] E. Larsson, O. Edfors, F. Tufvesson and T. Marzetta, "Massive MIMO for next generation wireless systems.", *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

- [7] A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems." *Selected Topics in Signal Processing*, vol.8, no.5, pp.831-846, Oct. 2014.
- [8] A. Alkhateeb, M. Jianhua, N. Gonzalez-Prelcic and R.W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems." *Communications Magazine*, vol.52, no.12, pp.122-131, Dec. 2014.
- [9] Sun, Songlin, et al. "Integrating network function virtualization with SDR and SDN for 4G/5G networks." *Network IEEE* 29.3(2015):54-59.
- [10] K. Zheng, S. Ou, and X. Yin, "Massive MIMO Channel Models: A Survey," in *International Journal of Antennas and Propagation*, vol. 2014, Article ID 848071, 2014.
- [11] C. Masouros, M. Sellathurai and T. Ratnarajah, "Large-Scale MIMO Transmitters in Fixed Physical Spaces: The Effect of Transmit Correlation and Mutual Coupling." in *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2794-2804, July 2013.

# Precoding Design for Massive MIMO Time Division Duplex System

Gao Wei, Wu Muqing, Zhao Min, Cheng Yanqing

Beijing Laboratory of Advanced Information Network

Beijing University of Posts and Telecommunications

Beijing 100876, P. R. China

Email: {iopjklm, wumuqing, zhaomin, chengyq}@bupt.edu.cn

**Abstract**—In massive MIMO (Multiple-Input Multiple-Output) time division duplex (TDD) system, the performance of downlink transmission is limited by channel reciprocity errors called antenna reciprocity errors, which is hard to be eliminated completely in practical systems. In order to reduce the performance degradation of the downlink transmission caused by antenna reciprocity errors, a precoding algorithm is put forward based on the statistical characteristics of the antenna reciprocity error. We derive the Signal to Leakage and Noise Ratio (SLNR) in downlink of massive MIMO system, in the presence of antenna reciprocity error. The target function is to maximize the average SLNR of users. We find that massive MIMO system channel capacity is improved approximate linearly with the increasing of the number of users (UEs) and Signal to Noise Ratio (SNR), when reciprocity antenna error is eliminated completely. The simulation results prove that proposed algorithm can outperform the traditional precoding algorithms, and that the larger MIMO system is, the better performance of the proposed precoding algorithm is.

**Index Terms**—massive MIMO; large scale MIMO; channel reciprocity error; precoding algorithm.

## I. INTRODUCTION

With the exponential growth in wireless data traffic, number of active terminal and variety of new services, it puts forward higher demands on reliability of network, transmission rates and channel capacity. In order to meet these requirements, Massive MIMO technology was firstly proposed by Bell Labs in 2010 [1], it was considered the key technology of the 5-th Generation (5G) cellular system.

Massive MIMO is also known as large-scale MIMO, and is based on existing mature MIMO technology. In current communication standards, the maximum number of antennas that can be configured in MIMO single-end is 8 [2]. In massive MIMO, the number of base station (BS) antenna configuration will have increased significantly, typically 100-1000 [3]. Massive MIMO is now being fierly studied as the promising technology in next generation communication system. Massive MIMO can greatly improve the cell throughput and spectral efficiency [4]. And massive MIMO can reduce electromagnetic interference and interference to other users. Moreover, massive MIMO can improve the reliability of communication system [5]. This feature is very useful for interference limited cellular network. In addition, massive MIMO radio frequency (RF) module has strong portability and low cost [6], [7].

TDD system can make use of reciprocity of the uplink and downlink channel, and estimate downlink channel according

to the uplink channel state information (CSI). So TDD is considered more suitable for the development of large-scale MIMO technology [8]. So the system mentioned in the rest part of the presentation is refer to the TDD system.

However, in realistic systems, uplink and downlink propagation channel is not only an ideal reciprocal channel, but also includes RF transceiver module of UE antenna and BS antenna. Because of differences in the structure and manufacturing process on the RF transceiver module which includes internal clock, amplifier, phase locked loop (PLL) and other devices, different RF module has different RF gain, which means that each antenna has a random phase and amplitude. Especially for large-scale MIMO system adopted low-cost RF unit, the phase noise, antenna coupling and other factors of the RF module make this phenomenon particularly serious. Therefore, the random gain of UE and BS antennas will cause reciprocal errors of uplink and downlink channel, which often referred to as the antenna reciprocity error, and lead to the performance degradation of the downlink transmission.

Precoding is a practical solution under this circumstance. The precoding technology is a kind of pretreatment strategy. It obtains the symbols to be transmitted from multiple antennas through weighting original data symbols. There are several precoding schemes that have been proposed to mitigate interference among different users in [9], [10], [11], [12]. And other precoding algorithms are intend to maximizing system capacity. In Ref. [13], the authors compared the system capacity under the Match Filter (MF) precoding and the Zero Forcing (ZF) linear precoding, and found that the performance of precoding algorithms relies on the reciprocity of the channel. In Ref. [14], channel reciprocity error was modeled as a diagonal matrix, whose diagonal elements change slowly over time, temperature and power. In Ref. [15], the authors analyzed the impact of antenna reciprocity errors on the some of the precoding algorithm. In Ref. [16], the authors pointed out that the channel reciprocity system error calibration algorithm consumes lots of time-frequency resources, and that it is difficult to completely eliminate errors.

The rest part of the paper is organized as follows. Section II describes the system model of massive MIMO, and deduces the relationship between channel matrices of uplink and downlink. In Section III, it is discussed to reduce the effects of the antenna reciprocity error on the downlink channel

from the perspective of precoding algorithm designing. In the presence of channel reciprocity error, we derive the average SLNR of large-scale MIMO, and find a suitable precoding algorithm aiming to maximize average SLNR of users. Section IV gives the simulation setup, and compares the performance of proposed algorithm with the typical pre-coding algorithms from both bit error rate (BER) and sum rate of system. In the last section, we make conclusion about the proposed precoding and discuss possible directions that may be studied in future work.

## II. SYSTEM MODEL

### A. System Description

In this article, the massive MIMO system is working in TDD mode. We assume that the system consists of a BS with  $N$  antennas and  $M$  UEs with one antenna, as it is shown in Fig. 1:

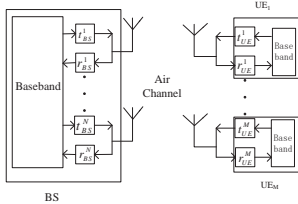


Fig. 1. System model of massive MIMO.

### B. Antenna Reciprocity Error Model

Uplink and downlink channels between the BS and the UE can be represented by Eq. (1) and Eq. (2).

$$H_{DL} = T_{BS} H R_{UE} \in C^{N \times M} \quad (1)$$

$$H_{UL} = R_{BS} H T_{UE} \in C^{N \times M} \quad (2)$$

where the matrix  $H$  represents Rayleigh slow flat fading air channel between the BS and the UE, which means the time when transmission channel state remains constant is longer than a single symbol period. Matrices  $T_{BS}$  represents receive gain of BS, and  $R_{BS}$  represent send gain of BS,  $T_{UE}$  and  $R_{UE}$  represent transmit gain and receive gain of UEs' RF respectively. The four matrices are modeled as diagonal matrices where diagonal elements are slowly varying function of time, temperature and power [14]. This paper argues that their diagonal elements are complex random variables whose amplitude and phase are independent and identically distributed (i.i.d).

Relationship between channel matrix of downlink and uplink can be derived from Eq. (1) and Eq. (2):

$$H_{DL} = T_{BS} R_{BS}^{-1} H_{UL} T_{UE}^{-1} R_{UE} \quad (3)$$

where  $(\cdot)^{-1}$  represents the inverse matrix. As  $T_{UE}$ ,  $R_{UE}$ ,  $T_{BS}$  and  $R_{BS}$  are diagonal matrices, so that  $E_{BS} = T_{BS} R_{BS}^{-1} = \text{diag}\{e_{1,BS} \cdots e_{1,BS} \cdots e_{N,BS}\}$  represents BS antenna reciprocity error, equally  $E_{UE} = T_{UE} R_{UE}^{-1} = \text{diag}\{a_{1,UE} \cdots a_{M,UE}\}$

represents the UE antenna reciprocity errors. Then Eq. (3) becomes  $H_{DL} = E_{BS} H_{UL} E_{UE}^{-1}$ .

## III. PRECODING DESIGN

### A. SLNR Analysis of Downstream Transmission When Antenna Reciprocity Error Exists

In downlink transmission.

$$y = 1/\sqrt{\delta} H_{DL} W S + n = 1/\sqrt{\delta} E_{BS} H_{UL} E_{UE}^{-1} W S + n \quad (4)$$

where  $y$  represents the received signal.  $\delta = \|W S\|^2 / P$  represents the power control factor,  $P$  is the transmitting power of the BS.  $W = [w_1, w_2 \cdots w_N] \in C^{M \times N}$  is precoding matrix.  $S = [s_1, s_2 \cdots s_N]^T \in C^{N \times 1}$  is the BS transmitting signal vector.  $n$  is the additive white Gaussian noise,  $\sigma_n^2$  is the noise variance. Precoding matrix  $W$  depend on the precoding algorithms and baseband uplink channel matrix estimation  $H_{UL}$ . If the antenna reciprocity error of UE and BS does not exist, which means  $E_{BS}$  and  $E_{UE}$  are unit matrices.

Since the coupling characteristics of user's Signal to Interference plus Noise Ratio (SINR), it is difficult to obtain precoding explicit expression, if we maximize SINR. So this paper set the average SLNR of each user as the target function. Downlink SLNR of user  $k$  is described as:

$$S L N R_k = \frac{s_k^H w_k^H h_{k,DL}^H h_{k,DL} w_k s_k}{\sigma_n^2 \delta + \sum_{i \neq k} s_k^H w_k^H h_{i,DL}^H h_{i,DL} w_k s_k} \quad (5)$$

$$= \frac{s_k^H w_k^H e_{k,UE}^H h_{k,UL}^H e_{k,BS} e_{k,BS} h_{k,UL} e_{k,UE} w_k s_k}{\sigma_n^2 \delta + \sum_{i \neq k} s_k^H w_k^H e_{i,UE}^H h_{i,UL}^H e_{i,BS} h_{i,UL} e_{i,UE} w_k s_k} \quad (6)$$

where  $(\cdot)^H$  represents the conjugate transpose,  $h_{k,DL}$  represents the downlink channel matrix of the  $k$ -th row,  $e_{k,BS}$  represents the  $k$ -th diagonal element of  $E_{BS}$ ,  $e_{k,UE}$  represents the  $k$ -th diagonal element of  $E_{UE}^{-1}$ . It was proved that in LTE-Advanced standards, users can obtain equivalent channel through demodulation reference signal (DM-RS), and eliminate the influence of  $E_{UE}$  on receiving signal through the proper channel equalization [17]. So we consider that  $E_{UE}$  had no effect on the downlink channel. However, the BS antenna correlation is unavoidable. By Eq. (6), we find that  $E_{BS}$  make a precoding vector rotate, generating additional inter-user interference, significantly reduce the performance of downlink channel. The BS antenna reciprocity error does exist, which means  $E_{BS} \neq I$ .

### B. Design of Precoding

Since the influence of  $E_{UE}$  on receiving signal can be eliminated, the paper considers that  $E_{UE} = I$ , therefore, the Eq. (6) becomes:

$$S L N R_k = \frac{s_k^H w_k^H h_{k,UL}^H e_{kk,BS}^H e_{kk,BS} h_{k,UL} w_k s_k}{\sigma_n^2 \delta + \sum_{i \neq k} s_k^H w_k^H h_{i,UL}^H e_{ii,BS} h_{i,UL} w_k s_k} \quad (7)$$

We derive the suitable precoding algorithm by maximizing  $E\{S L N R_k\}$ , where  $E\{\cdot\}$  means to take expectations. the target function is:

$$\max_w E\{S L N R_k\} \quad (8)$$

Assuming the transmission power of each base station antenna is equal. Typically, the precoding matrix and modulation symbols are independent. So in this paper, we normalize the power of modulation symbol, which means that  $E(SS^H) = I$ . According to the Jensen inequality, the lower bound of target function can be expressed as:

$$E\{SLNR_k\} \geq \frac{w_k^H h_{k,UL}^H E\{e_{kk,BS}^H e_{kk,BS}\} h_{k,UL} w_k}{\sigma_n^2 \delta + \sum_{i \neq k} w_k^H h_{i,UL}^H E\{e_{ii,BS}^H e_{ii,BS}\} h_{i,UL} w_k} \quad (9)$$

Making  $Q_{BS} = E\{E_{BS}^H E_{BS}\}$ ,  $Q_{BS}$  represents the autocorrelation matrix of the BS antenna reciprocity error, it includes variance and mean of BS antenna reciprocity error. Statistical characteristics of the antenna reciprocity error can be obtained by a large number of experimental measurements and statistical parameters modeling. As mentioned above, elements of  $E_{BS}$  were modeled as random variables whose amplitude and phase are uniform distribution. This article assumes that  $Q_{BS}$  is known and symmetric positive definite, so  $Q_{BS}$  can be Cholesky decomposed. At last, we obtain  $D_{BS}$  by decomposing  $Q_{BS}$  ( $Q_{BS} = D_{BS}^H D_{BS}$ ).

Making the average user SLNR lower bound as the objective function, the Eq. (8) of the optimization problem is rewritten as

$$\max_{w_k} \frac{s_k^H w_k^H h_{k,UL}^H Q_{i,BS} h_{i,UL} w_k s_k}{\sigma_n^2 \delta + \sum_{i \neq k} s_k^H w_k^H h_{i,UL}^H Q_{i,BS} h_{i,UL} w_k s_k} \quad (10)$$

where  $Q_{i,BS}$  represents the  $i$ -th element of  $Q_{BS}$ . When  $W = H_{UL}^H D_{BS}^H (D_{BS} H_{UL} H_{UL}^H D_{BS}^H)^{-1}$ , lower bound of  $E\{SLNR_k\}$  has a maximum value. Because at this time,  $w_k^H h_{i,UL}^H E\{e_{ii,BS}^H e_{ii,BS}\} h_{i,UL} w_k = 0$  ( $i \neq k$ ), which means that expectation of energy antenna  $K$  leaked to other antenna is 0, and expectation of interference among the antennas is also 0. If we don't take the antenna reciprocity error into consideration, traditional SLNR precoding [18] performs pretty well. However, the antenna reciprocity error does exist, our precoding matrix can make some compensation to antenna reciprocity errors based on the statistical characteristics of the antenna reciprocity errors. The algorithm can reduce the effect of the errors on massive MIMO downlink channel performance.

#### IV. SIMULATION ANALYSIS

In this section, the performance of the proposed algorithm on massive MIMO is evaluated by simulation. Because the algorithm is for large scale MIMO system, this paper simulation scenario is BS with 256 antennas, which means  $N=256$ . Air channel matrix  $H$  between BS and UE is the Rayleigh fading channel. Assume that uplink channel matrix  $H_{UL} = H$ . Like the above assumptions, we ignore the UE antenna reciprocity error ( $E_{UE} = I$ ), and make the BS reciprocity error  $E_{BS}$  a diagonal matrix, the amplitude of its diagonal elements satisfies the uniform distribution  $U(-A, A)$ , the unit is dB, and the phase satisfies uniform distribution  $U(-\theta, \theta)$ , the unit is  $^\circ$ . The parameters  $A$  and  $\theta$  are represent antenna reciprocity error. Statistical properties matrix  $Q_{BS}$  of the BS antenna reciprocity error is known. According to the above conditions,  $H_{DL} =$

$E_{BS} H$ . In the simulation, the carrier frequency is  $6 \times 10^9$  Hz, the arrangement of transmitting and receiving antennas is linear uniform array, interval of antennas is  $\lambda/2 = 0.025$  m,  $\lambda$  means wavelength. The simulation environment is urban microcell, using the 2PSK modulation.

Suppose uplink channel correlation matrix  $H_{UL}$  and auto-correlation matrix  $Q_{BS}$  of the BS antenna reciprocity error are known, according to the algorithm proposed above, we calculate the corresponding pre-coding matrix  $W$ . As Eq. (7) shows, after signal propagating over downlink channel  $H_{DL}$  and superimposing noise  $n$ , we can calculate the SLNR of  $k$ -th user. The sum rate  $Ra$  that downlink channel of massive MIMO can meet can be expressed as below, in the presence of antenna reciprocity error.

$$Ra = \sum_{k=1}^M \log_2(1 + SLNR_k) \quad (11)$$

$SNR = P/\sigma_n^2$ , representing the ratio of the BS transmitting power and the receiving noise power, we usually normalize BS transmitting power, which means  $P = 1$ , and  $SNR = 1/\sigma_n^2$ . Next, we will simulate the performance of precoding algorithms, under the circumstances and assumption mentioned above. We get the following simulation figures.

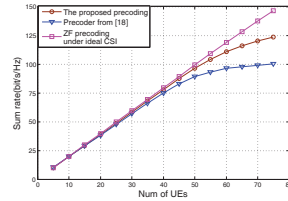


Fig. 2. The average sum rate with a system employed 256 BS antennas and different UE antennas,  $A = 2$  dB,  $\theta = 20^\circ$ ,  $SNR = 5$  dB.

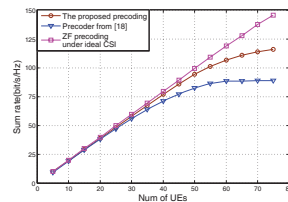


Fig. 3. The average sum rate with a system employed 256 BS antennas and different UE antennas,  $A = 2$  dB,  $\theta = 20^\circ$ ,  $SNR = 5$  dB.

Fig. 2 and Fig. 3 show the average sum rate curves with different UEs for the following antenna configuration:  $N = 256$  BS antennas and  $SNR = 5$  dB for  $A = 2$ ,  $\theta = 20^\circ$  and  $A = 3$ ,  $\theta = 30^\circ$  respectively. The figures compare three schemes: the proposed precoding; the ZF precoding under the ideal CSI which means the absence of BS antennas reciprocity error; and the precoder from [18] which is the typical SLNR-based

precoding algorithm. Seen from the figures, the BS antenna reciprocity errors significantly reduce sum rate. The figures also show that with the increase of the UEs, the average sum rate gain that the proposed algorithm compares with the precoder from [18] becomes larger and larger, which means the larger the size of the MIMO system is, the better proposed algorithm perform. Compared Fig. 2 with Fig. 3, it is suggested that the proposed precoding can well adapt to large-scale MIMO system.

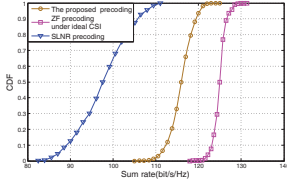


Fig. 4. CDF of Sum rate under different algorithms, with a multiuser MIMO system employed  $N = 265$  BS antennas and  $M = 64$  UEs, each equipped with one antenna, and  $A = 2dB$ ,  $\theta = 20^\circ$ ,  $SNR = 5dB$

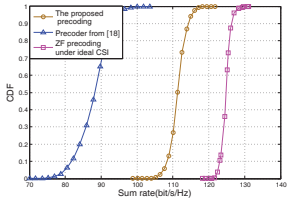


Fig. 5. CDF of Sum rate under different algorithms, with a multiuser MIMO system employed  $N = 256$  BS antennas and  $M = 64$  UEs, each equipped with one antenna, and  $A = 3dB$ ,  $\theta = 30^\circ$ ,  $SNR = 5dB$

Fig. 4 and Fig. 5 are the simulation of cumulative distribution function (CDF) curve of sum rate, the antenna configuration is same with Fig. 2, except the number of UEs is 64. In Fig. 4, when  $A = 2$ ,  $\theta = 20^\circ$ , the sum rate that the proposed algorithm is higher than the precoder from [18] is 16.42 bit/s/Hz, in Fig. 5,  $A = 3$ ,  $\theta = 30^\circ$ , which means the BS antennas reciprocity error is increased, the sum rate that the proposed algorithm is higher than the precoder from [18] is 23.96 bit/s/Hz. It shows that though sum rate of every schemes is reduced due to the existence of BS antennas reciprocity error, the proposed precoding has better performance in resistance antenna reciprocity errors, achieve greater sum rate, because the proposed precoding can make some compensation to antenna reciprocity errors to reduce the equivalent noise.

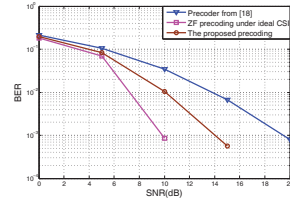


Fig. 6. BER with different SNR, with  $N = 256$  BS antennas and  $M = 64$  UEs,  $A = 2$  and  $\theta = 20^\circ$ .

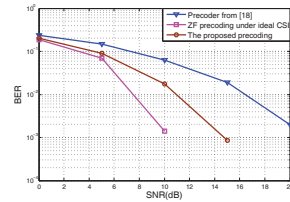


Fig. 7. BER with different SNR, with  $N = 256$  BS antennas and  $M = 64$  UEs,  $A = 3$  and  $\theta = 30^\circ$ .

Fig. 6 and Fig. 7 show the BER curves of each precoding algorithm under different SNR when  $A = 2dB$ ,  $\theta = 20^\circ$  and  $A = 3dB$ ,  $\theta = 30^\circ$  respectively. As we can see from figures, with the increase of SNR, the BER of the proposed precoding drop faster, even the antennas reciprocity error increases.

In addition, as it is seen from Fig. 2 and Fig. 3 that relationship between sum rate and the SNR (dB) basically is a linear relationship when the number of UEs is much less than BS antennas, the following part will explain the reasons.

According to information theory, in lossless transmission, The upper bound of sum rate  $Ra$  is the channel capacity. Since the number of BS antenna is more than the number of UEs, channel capacity of massive MIMO system downstream transmission is calculated by Eq. (12):

$$C = \log_2 \left[ \det \left( I_M + SNR \frac{R}{N} * H * H^H \right) \right] \quad (12)$$

where

$$\frac{H * H^H}{N} = \begin{bmatrix} \frac{h_{11}h_{11}^H}{N} & \frac{h_{11}h_{12}^H}{N} & \dots & \frac{h_{11}h_{1M}^H}{N} \\ \frac{h_{21}h_{11}^H}{N} & \frac{h_{21}h_{12}^H}{N} & \dots & \frac{h_{21}h_{1M}^H}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{h_{M1}h_{11}^H}{N} & \frac{h_{M1}h_{12}^H}{N} & \dots & \frac{h_{M1}h_{1M}^H}{N} \end{bmatrix} \quad (13)$$

$N$  is the number of transmitting antennas,  $M$  is the number of UEs,  $H$  is a  $M \times N$  two-dimensional,  $h_1, h_2 \dots h_M$  are row vectors of matrix  $H$ , we assume that elements of  $H$  are i.i.d with mean being 0 and variance being 1.

$H$  is the channel Gaussian channel matrix, its elements are Gaussian i.i.d. According to Strong law of large numbers, when  $N$  is large enough:

$$\frac{h_i^2}{N} = \frac{|h_1^2| + |h_2^2| + \dots + |h_N^2|}{N} \approx \text{Var}[h_i] + (E[h_i])^2 = 1 \quad (14)$$

$$\begin{aligned} \frac{h_i h_j^H}{N} &= \frac{h_1^i h_1^{j*} + h_2^i h_2^{j*} + \dots + h_N^i h_N^{j*}}{N} \\ &= \frac{g_1 + g_2 + \dots + g_N}{N} \approx (E[g_i])^2 = 0 \quad (i \neq j) \end{aligned} \quad (15)$$

where  $()^*$  represents a conjugated,  $h_j^i$  means the  $j$ -th element of  $h_i$ ,  $g_k$  represents the Gaussian variables obtained by multiplying two Gaussian variables  $h_k^i$  and  $h_k^{j*}$ , because the expectation of  $h_k^i$  and  $h_k^{j*}$  are 0,  $E[g_i] = 0$ .

$$\frac{H * H^H}{N} = \begin{bmatrix} \frac{h_1^2}{N} & \frac{h_1 h_2^H}{N} & \dots & \frac{h_1 h_M^H}{N} \\ \frac{h_2 h_1^H}{N} & \frac{h_2^2}{N} & \dots & \frac{h_2 h_M^H}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{h_M h_1^H}{N} & \frac{h_M h_2^H}{N} & \dots & \frac{h_M^2}{N} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{M \times M} \quad (16)$$

Therefore, the Eq. (12) becomes

$$\begin{aligned} C &= \log_2 \det \left( \begin{bmatrix} 1 + SNR & 0 & \dots & 0 \\ 0 & 1 + SNR & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 + SNR \end{bmatrix}_{M \times M} \right) \\ &= M \log_2 (1 + SNR) \end{aligned} \quad (17)$$

As can be seen from Eq. (17), when the transmitter antenna is large and far greater than receiver antenna, channel capacity not only has a linear relation with SNR (dB), but also proportional to the number of the UEs in the ideal condition. It is impossible to eliminate antenna reciprocity error completely, and in real system, there are many factors that restrain sum rate, it is no doubt that sum rate can not reach its upper bound-channel capacity. Anyway the proposed precoding can effectively suppress the antenna reciprocity error.

## V. CONCLUSIONS

In this paper, we develop a precoding algorithm to reduce the effect of the antenna reciprocity error on TDD system's downlink channel adopted massive MIMO technology. The proposed algorithm targets to maximize the average SLNR of UE based on the statistical characteristics of the antenna reciprocity error. We conduct a simulation mainly from sum rate and BER. The simulation prove that proposed precoding algorithm can outperform existing popular algorithm, e.g. traditional SLNR algorithm. The simulation suggests that the proposed algorithm performance becomes better when the antenna array of massive MIMO becomes larger. When the transmitter antenna is large and far greater than receiver antenna, we also find that massive MIMO system channel capacity is improved approximate linearly with the increasing of the number of UEs and SNR in absence of antenna reciprocity error. As its good performance, the proposed algorithm is worthy for massive MIMO system. In future work, we intend to investigate further system aspects such as precoding design in multiple cell.

## ACKNOWLEDGMENTS

This work is supported by the Director Funds of Laboratory of Network System Architecture and Convergence (2015BKL-NSAC-ZJ-06).

## REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo: How many antennas do we need?" in *Communication, Control, and Computing (Allerton)*, 2011 49th Annual Allerton Conference on, Sept 2011, pp. 545–550.
- [4] J. Lee and S. H. Lee, "A compressed analog feedback strategy for spatially correlated massive mimo systems," in *Vehicular Technology Conference (VTC Fall)*, 2012 IEEE, Sept 2012, pp. 1–6.
- [5] W. Li, L. Liu, C. Tao, Y. Lu, J. Xiao, and P. Liu, "Channel measurements and angle estimation for massive mimo systems in a stadium," in *Advanced Communication Technology (ICACT)*, 2015 17th International Conference on, July 2015, pp. 105–108.
- [6] Z. Yu, J. Zhou, R. Zhang, and W. Yang, "A digital rf transmitter prototype for the future massive mimo system," in *Wireless Symposium (IWS)*, 2015 IEEE International, March 2015, pp. 1–3.
- [7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo in the ul/dl of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, February 2013.
- [8] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [9] R. Chen, J. G. Andrews, and R. W. Heath, "Multiuser space-time block coded mimo with downlink precoding," in *Communications, 2004 IEEE International Conference on*, vol. 5, June 2004, pp. 2689–2693 Vol.5.
- [10] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user mimo channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, May 2007.
- [11] V. Stankovic and M. Haardt, "Generalized design of multi-user mimo precoding matrices," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 953–961, March 2008.
- [12] C. Masouros, "Correlation rotation linear precoding for mimo broadcast communications," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 252–262, Jan 2011.
- [13] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, February 2013.
- [14] Z. Gu, N. Wei, and Z. Zhang, "Analysis and modeling of channel reciprocity errors based on an experimental ofdm/tdd coordinated multi-point transmission system," in *Communications in China (ICCC)*, 2013 IEEE/CIC International Conference on, Aug 2013, pp. 292–296.
- [15] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive mimo systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, Nov 2014.
- [16] R. Rogalin, O. Y. Bursalioğlu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale mimo," in *Information Theory and Applications Workshop (ITA)*, 2013, Feb 2013, pp. 1–10.
- [17] H. Wang, Y. Huang, S. Jin, F. Yu, and L. Yang, "Performance analysis on precoding and pilot scheduling in very large mimo multi-cell systems," in *Wireless Communications and Networking Conference (WCNC)*, 2013 IEEE, April 2013, pp. 2722–2726.
- [18] M. Sadek and S. Aissa, "Leakage based precoding for multi-user mimo-ofdm systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2428–2433, August 2011.

# Performance Improvement of Relay System Using Polarization Diversity

Maja Delibasic  
Podgorica, Montenegro  
Email: majai@ac.me

Milica Pejanovic-Djurisic  
Research Centre for ICT  
Faculty of Electrical Engineering, University of Montenegro  
Podgorica, Montenegro  
Email: milica@ac.me

**Abstract**—Wireless relaying has been proved as an efficient transmission technique which improves wireless system reliability and durability, through increased network range, capacity and energy efficiency. However, the performances of wireless relay systems are strongly influenced by the worst communication link between the transmitter and the receiver. Since realistic communication scenario usually assumes non-identical channels parameters, it is necessary to consider some additional techniques to improve the characteristics of this bottleneck link, and performance of relay system as whole. In this paper we assume amplify-and-forward relay systems with variable gain amplification, operating in Rayleigh fading environment. In order to improve system performance we propose implementation of polarization diversity. We derive two statistical functions describing total received SNR (signal-to-noise ratio), probability density function and moment generating function, which enabled the derivation of the system error rate performance. The obtained results show that, using polarization diversity, improvements for the considered system performance can be realized with significantly lower SNR, despite a certain level of correlation and power unbalance between the diversity branches.

## I. INTRODUCTION

In order to support great expansion of wireless communication systems over the last decade was appropriate technical solutions are proposed and adopted, which have succeeded to meet customers' demands. One of the solutions that is proved to provide greater network range with higher quality of service is relay (R) based wireless communication system, [1]. One of the most commonly used relay based solution assumes dual-hop scenario, where the source terminal (S), that cannot achieve direct communication with the destination terminal (D), sends data over the intermediate relay node.

R stations can be classified as regenerative and non-regenerative. Regenerative R stations decode the received signal and again re-encode it before forwarding toward destination (D) terminal, while non-regenerative R stations just amplify the received signal prior forwarding to D. Non-regenerative relay systems are less complex than regenerative, and they introduce shorter latency. Depending on the possibility to estimate the channel between S and R, non-regenerative R stations may amplify the received signal with the fixed gain (FG), or with the variable gain (VG). Analyses presented in this paper consider the two hop non-regenerative (AF - Amplify and Forward) relay system

applying VG at the R station. Although such systems have been widely explored in the literature, there is still ongoing intensive research activity with the goal to further improve their performance, energy efficiency or to optimize their functions.

One of the main drawbacks of relay systems is the existence of bottleneck link, which limits the performance of the overall communication system. In order to improve the characteristics of the bottleneck link, several methods are proposed, such as selective relaying, [2], opportunistic relaying, [3], multi-antenna relaying, [4], cooperative relaying, [5], or some combination of these methods, [6]–[8].

In the selective relaying one (or several) best relays among the  $L$  potential relays is selected to relay the transmitted signal. Opportunistic relaying assumes that the relay with the strongest transmitter-relay-receiver path is selected among several candidates in a distributed fashion using instantaneous channel measurement. In multi-antenna relaying communicating nodes are equipped with multiple antennas, which are used to provide diversity gain. In the cooperative relaying D combines signals received from multiple sources (usually S and R) improving the overall system performance.

In this paper we propose a type of multi-antenna relaying. Instead of implementing several antennas at the receive end of a bottleneck link, we propose the implementation of one compact, dual-polarized antenna. It has been proved that polarization diversity systems offer increased signal to noise ratio (SNR) and reliability, comparable to the space diversity. However, due to the limited size of network nodes, the implementation of multiple adequately separated antennas can be rather difficult. That is why the polarization diversity is very attractive for implementation, despite the fact that this type of diversity is limited to the use of only two antennas using a single aperture, [9]. However, due to the inevitable correlation between the received diversity signals, analytical modelling of such systems becomes more complex.

In order to examine the influence of diversity implementation in the relay system, as a solution for system performance improvements, in this paper we derive full analytical model for BER (Bit Error Rate) analyses of the two-hop AF VG relay system implementing dual polarized antenna on the receive end of a worse relay channel. We



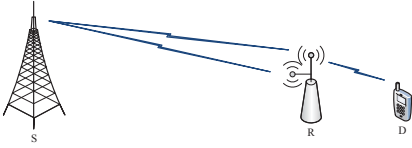


Fig. 1. System model

assume MRC (Maximal Ratio Combining) of diversity signals, as the optimal diversity combining technique. Then we compare the obtained BER results for the assumed DPSK (Differentially Phase Shift Keying) mapping of the information bits, with BER performance of ordinary AF VG relay system without diversity.

The paper is organized as follows. Section II describes the analyzed system and channel model, with the most important statistical functions describing received SNR. In Section III we derive the BER expression for the received signal at destination. The obtained results are presented in Section IV. Concluding remarks are given in Section V.

## II. SYSTEM AND CHANNEL MODEL

In this paper we consider on two hop relay system consisting of one source (S), destination (D) and intermediate node that assists in signal transmission, called relay (R). Amplify-and-forward (AF) relaying strategy with variable gain (VG) amplification is adopted. Knowing that the performance of relay system is limited with the worst link between the S and D, in order provide better channel conditions on that bottleneck link we propose the implementation of receive diversity (Fig. 1). Diversity can be implemented at R or D, depending on the channel conditions on the two links. Received diversity signals are combined applying MRC (Maximal Ratio Combining), since it is the optimal combining technique, [10]. Although in this paper we conduct the analysis for the case when diversity is implemented at R, the presented model is general, and thus applicable also for the scenario when diversity combining is implemented at D.

In order to keep the diversity implementation as simple as possible, we propose the use of polarization diversity. This type of diversity system has been proven to provide certain wireless system performance improvements, with no additional space needed for its implementation. It uses only one compact, dual-polarized antenna, [11], which makes this type of diversity very attractive for implementation, especially in dense urban environments, or when the size of network node limits the possibility to mount several antennas.

The received diversity signals are partially correlated. Several measurements reported that typical values for correlation coefficient vary from 0.2-0.4, [11], while studies have shown that significant performance improvement can be achieved as long as the correlation coefficient is less than 0.7, [12].

Another parameter typical for polarization diversity is cross-polar discrimination (XPD). It denotes power difference between the average powers of co-polarized and cross-polarized signals. It is strongly dependent on the environment, and has larger influence on overall system performance than correlation coefficient. Typical values of XPD vary from 1-10 dB in urban and suburban environment, and 10-18 dB in rural environment [12]. Still, it has been shown, [9], [11]–[13] that despite a certain level of the correlation between the signals from different diversity branches and power imbalance between them, polarization diversity provides significant performance improvements of wireless systems.

In order to describe the performance of such communication system, it is necessary to have a good knowledge of wireless channels characteristics. One of the most adopted models to describe the behaviour of the wireless propagation channel is Rayleigh fading model, which is suitable for non-line-of-sight (NLoS) communications.

In fading channels, signal fading amplitude is a random variable (RV), and the faded signal is perturbed by additive white Gaussian noise (AWGN) at the receiver. AWGN is assumed to be statistically independent of the fading signal. Denoting the instantaneous signal-to-noise ratio (SNR) as  $\gamma$ , and the average SNR as  $\bar{\gamma}$ , probability density function (PDF) of RV  $\gamma$  in the case of Rayleigh fading is defined as, [14]:

$$p(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}}, \quad (1)$$

while the PDF of the total SNR at the output of diversity combiner, when the two diversity signal are partially correlated, with non-identical average SNRs per diversity branch is given as, [14]:

$$p(\gamma) = \frac{e^{-\frac{\gamma v + \gamma_h \sqrt{(\bar{\gamma}_v + \bar{\gamma}_h)^2 - 4\bar{\gamma}_v \bar{\gamma}_h (1-\rho)}}{2\bar{\gamma}_v \bar{\gamma}_h (1-\rho)}} \gamma - e^{-\frac{\gamma v + \gamma_h + \sqrt{(\bar{\gamma}_v + \bar{\gamma}_h)^2 - 4\bar{\gamma}_v \bar{\gamma}_h (1-\rho)}}{2\bar{\gamma}_v \bar{\gamma}_h (1-\rho)}} \gamma}{\sqrt{(\bar{\gamma}_v + \bar{\gamma}_h)^2 - 4\bar{\gamma}_v \bar{\gamma}_h (1-\rho)}}. \quad (2)$$

In (2) we use the indexes  $v$  and  $h$  to denote parameters of vertically and horizontally polarized diversity signals. Furthermore, we put  $a = \bar{\gamma}_1 = \bar{\gamma}_v + \bar{\gamma}_h$  and  $b^2 = 2\bar{\gamma}_v \bar{\gamma}_h (1-\rho)$ .

For AF VG relay systems, the total received SNR at the D is obtained as:

$$\gamma = \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2 + c}, \quad (3)$$

with  $\gamma_1$  and  $\gamma_2$  representing SNRs of the S-R and R-D link, respectively. The exact total SNR is given by substituting  $c=1$ , while it is well approximated (at higher SNR) for  $c=0$ . Without loss of generality, in further analyses we will assume that  $c=0$ , [15]. Applying the MRC of the received signals at R, instantaneous SNR of S-R link is  $\gamma_1 = \gamma_v + \gamma_h$ .

### A. Probability Density Function

In order to derive system's performance, it is necessary to find statistical properties of total received SNR. The total SNR

at D is given in (3), while its probability density function can be derived as:

$$p(\gamma) = \int_{-\infty}^{\infty} p_1(\gamma_1) p_2(\gamma_2) \gamma_{1,\gamma} d\gamma_2. \quad (4)$$

Putting  $\gamma_1 = \frac{\gamma\gamma_2}{\gamma_2 - \gamma}$ , thus  $\gamma_{1,\gamma} = \frac{d\gamma_1}{d\gamma} = \frac{\gamma_2^2}{(\gamma_2 - \gamma)^2}$ , and introducing the change of variable  $\gamma_2 - \gamma = w$ , PDF of the total SNR can be rewritten:

$$p(\gamma) = \frac{1}{\gamma_2 \sqrt{a^2 - 2b^2}} e^{-\frac{\gamma}{\gamma_2}} (\mathfrak{S}_1 - \mathfrak{S}_2), \quad (5)$$

with  $\mathfrak{S}_1$  and  $\mathfrak{S}_2$  defined as:

$$\begin{aligned} \mathfrak{S}_1 &= e^{-\frac{a-\sqrt{a^2-2b^2}}{b^2}\gamma} \int_0^{\infty} \frac{(\gamma+w)^2}{w^2} e^{-\frac{w}{\gamma_2}} e^{-\frac{a-\sqrt{a^2-2b^2}}{b^2}\frac{\gamma^2}{w}} dw \\ \mathfrak{S}_2 &= e^{-\frac{a+\sqrt{a^2-2b^2}}{b^2}\gamma} \int_0^{\infty} \frac{(\gamma+w)^2}{w^2} e^{-\frac{w}{\gamma_2}} e^{-\frac{a+\sqrt{a^2-2b^2}}{b^2}\frac{\gamma^2}{w}} dw. \end{aligned} \quad (6)$$

With the help of [16, Eq. (3.471-9)]:

$$\int_0^{\infty} x^{\nu-1} e^{-\frac{\beta}{x} - \alpha x} dx = 2 \left(\frac{\beta}{\alpha}\right)^{\frac{\nu}{2}} K_{\nu}(2\sqrt{\beta\alpha}), \quad (7)$$

and substituting  $\alpha = \frac{1}{\gamma_2}$ , and  $\beta = \frac{a-\sqrt{a^2-2b^2}}{b^2}\gamma^2$  for  $\mathfrak{S}_1$ , i.e.  $\beta = \frac{a+\sqrt{a^2-2b^2}}{b^2}\gamma^2$  for  $\mathfrak{S}_2$ , PDF is obtained in a closed form suitable for further use as in (8).  $K_{\nu}(\cdot)$  denotes modified Bessel function of the second kind.

### B. Moment Generating Function

In order to determine the average bit error rate (BER) as one of the most important performance measures of a wireless communication systems, we can use one of the two standard approaches: MGF (Moment Generating Function) based approach [14, Eq. (5.3)], or averaging the conditional bit error probability over the SNR distribution [14, Eq. (1.8)]. In further analysis we use the MGF-based approach.

The MGF of a random variable is defined as the Laplace transform of its PDF. From (8) and using [16, Eq. (6.621-3)], the MGF is obtained as in (9).  $F(\cdot)$  is hypergeometric function.

### III. AVERAGE BIT ERROR RATE ANALYSIS

In the following part we use the results obtained in previous section on probability density function (PDF) and moment generating function (MGF) of total SNR at D, which are necessary tools for BER (Bit Error Rate) performance analysis. We derive BER performance analysis for the dual-hop DPSK (Differential Phase Shift Keying) modulated AF VG relay system with polarization diversity implemented at the receiving end of a worse link, in the case of Rayleigh fading channels on both hops.

For the considered DPSK modulation of the information bearing signal, it is convenient to use MGF based approach for BER analysis, as in that case BER at D is obtained as [14]:

$$P_b = 0.5M_{\gamma}(1), \quad (10)$$

with  $M_{\gamma}$  denoting MGF of total SNR, given in (9).

## IV. RESULTS

In this Section we present some of the results obtained by the previously derived mathematical model. The presented analytical results are confirmed by comparing them to the results obtained through Monte Carlo simulations.

Simulation model assumes dual-hop AF VG relay system implementing polarization diversity at R. That means that two partially correlated diversity signals having unbalanced SNRs are generated on the S-R communication channel. R performs MRC of the received signals, and amplifies the signal prior forwarding to D. It is assumed that R uses knowledge on channel state information (CSI) about the S-R link to calculate the amplification gain  $G$ , [17]:

$$G(t) = \sqrt{\frac{E_R}{E_S|h_1(t)|^2 + N_{01}}}, \quad (11)$$

where  $E_R$  and  $E_S$  denote the average symbol energies transmitted by S and by R, while  $N_{01}$  is the noise variance at R. The channel transfer functions on both hops are generated as independent Gaussian complex random variables with zero mean.

BER performance of DPSK modulated AF VG relay systems with polarization diversity, as a function of the average SNR per hop, is given in Fig. 2. In order to get an insight in performance improvements realized through the implementation of diversity, BER curve of two hop AF VG relay system is also provided. We assume that average SNR at the output of MRC combiner  $\gamma_1 = \gamma_w + \gamma_h$  is equal to the SNR of the second hop,  $\gamma_2$ . Different values of the two parameters typical for polarization diversity, XPD and correlation coefficient between diversity signals' envelopes are assumed, in order to examine their influence on the overall system BER.

Markers represent the results obtained by Monte Carlo simulation, while lines represent analytically obtained values.

From the results presented in Fig. 2 it can be seen that the SNR gain realized through the implementation of polarization diversity is about 2.5-3 dB, for BER values  $10^{-3}$ .

Furthermore, it is clearly shown that XPD and correlation coefficient do not have a significant influence on the BER performance of the overall relay system, even if they have a major impact on performance of polarization diversity systems. Thus, in further analyses we assume the following typical values of these two parameters: correlation coefficient of diversity signal envelopes is 0.3, and XPD is 6 dB.

In our further analysis we choose non-identical average SNRs per hop. In fact, most of the practical implementation scenarios of relay systems assume non-identical fading channels parameters, due to the different distances between the network terminals, i.e. different attenuation of the transmitting signal. Since it is well known that the performance of relay system is dominated by the worst link between S and D, it is crucial to analyze and determine the bottleneck link in the particular scenario. By finding the worst link, techniques for

$$p(\gamma) = \frac{2\gamma e^{-\frac{\gamma}{\bar{\gamma}_2}}}{\bar{\gamma}_2 \sqrt{a^2 - 2b^2}} \left[ e^{-\frac{a - \sqrt{a^2 - 2b^2}}{b^2} \gamma} \sum_{k=0}^2 \binom{2}{k} \left( \frac{\bar{\gamma}_2 (a - \sqrt{a^2 - 2b^2})}{b^2} \right)^{\frac{1-k}{2}} K_{1-k} \left( \frac{2\gamma \sqrt{a - \sqrt{a^2 - 2b^2}}}{b \bar{\gamma}_2} \right) - e^{-\frac{a + \sqrt{a^2 - 2b^2}}{b^2} \gamma} \sum_{k=0}^2 \binom{2}{k} \left( \frac{\bar{\gamma}_2 (a + \sqrt{a^2 - 2b^2})}{b^2} \right)^{\frac{1-k}{2}} K_{1-k} \left( \frac{2\gamma \sqrt{a + \sqrt{a^2 - 2b^2}}}{b \bar{\gamma}_2} \right) \right]. \quad (8)$$

$$M_\gamma(s) = \frac{8}{3\bar{\gamma}_2 \sqrt{a^2 - 2b^2}} \sum_{k=0}^2 \binom{2}{k} \left[ \frac{(4(a - \sqrt{a^2 - 2b^2})/b^2)^{1-k}}{\left( \frac{1}{\bar{\gamma}_2} + \frac{a - \sqrt{a^2 - 2b^2}}{b^2} + \frac{2}{b} \sqrt{\frac{a - \sqrt{a^2 - 2b^2}}{\bar{\gamma}_2}} + s \right)^{3-k}} F\left(3-k, \frac{3}{2} - k; \frac{5}{2}, \frac{1}{\bar{\gamma}_2} + \frac{a - \sqrt{a^2 - 2b^2}}{b^2} - \frac{2}{b} \sqrt{\frac{a - \sqrt{a^2 - 2b^2}}{\bar{\gamma}_2}} + s\right) - \frac{(4(a + \sqrt{a^2 - 2b^2})/b^2)^{1-k}}{\left( \frac{1}{\bar{\gamma}_2} + \frac{a + \sqrt{a^2 - 2b^2}}{b^2} + \frac{2}{b} \sqrt{\frac{a + \sqrt{a^2 - 2b^2}}{\bar{\gamma}_2}} + s \right)^{3-k}} F\left(3-k, \frac{3}{2} - k; \frac{5}{2}, \frac{1}{\bar{\gamma}_2} + \frac{a + \sqrt{a^2 - 2b^2}}{b^2} - \frac{2}{b} \sqrt{\frac{a + \sqrt{a^2 - 2b^2}}{\bar{\gamma}_2}} + s\right) \right] \Gamma(3-k) \Gamma(1+k). \quad (9)$$

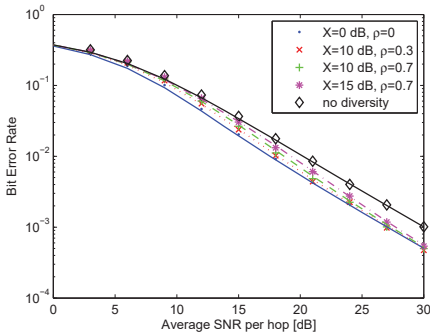


Fig. 2. Influence of XPD and correlation coefficient on the system BER

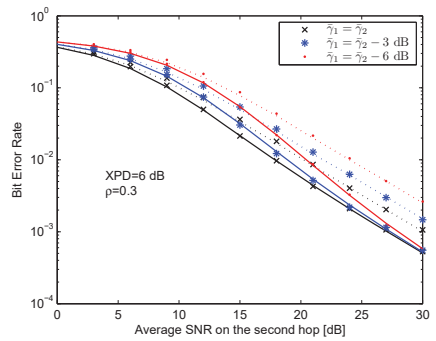


Fig. 3. BER of two-hop AF VG relay system under non-identical channel parameters. Solid line - system with polarization diversity; dotted line - relay system with no diversity

system performance improvements, such as proposed diversity, can be efficiently applied.

The obtained analytical results for BER performance of the analyzed two hop AF VG relay system, under such realistic assumptions (non-identical fading channels parameters) are presented in Fig. 3. For the sake of comparison, BER performance of the ordinary AF VG relay system, with no diversity implementation, is also presented (dotted line).

In this asymmetric fading channels scenario, higher SNR gain is achieved by implementing polarization diversity at the receive end of a bottleneck link. The level of system performance improvements is presented in Fig. 4.

All the presented results confirm the accuracy of the undertaken analytical approach, as we have excellent matching between the analytical results and simulation results for all

SNR values.

As expected, in the communication scenario where one relaying link has lower average SNR than the other, implementation of diversity can lead to significant system performance improvements. Polarization diversity can be considered as very attractive practical solution, since this type of diversity does not require additional space for mounting several antennas, which is very important due to limited size of the communicating network nodes.

## V. CONCLUSION

Wireless relay systems gain significant research attention in last decade, as an efficient network topology suitable for implementation in the infrastructure cellular systems with

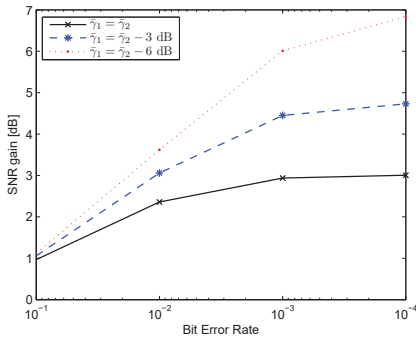


Fig. 4. SNR gain

low implementation costs, as well as in the wireless sensor networks that becomes more popular for implementation in wide range of applications.

Motivated by the constantly growing demands for wireless system capacity and the quality of the established communication, in this paper we analyze the possibility of wireless relay systems design improvement by implementing polarization diversity. This type of diversity has already proven as space and cost efficient solution, with the possibility to implement two collocated antennas with sufficiently low correlation between the diversity signals' envelopes.

Up to authors knowledge the presented mathematical model for describing total SNR and average BER of the considered two-hop relay system with polarization diversity implemented at the receive end of a bottleneck link can not be found in literature. The knowledge of the statistical functions, PDF and MGF, describing the total SNR of the communication system is of the considerable practical importance, since they are required for further system performance evaluation.

The obtained results show that the overall system performance does not significantly depend on the parameters describing polarization diversity, XPD and  $\rho$ . The obtained SNR gain can not be neglected for most of the typical values of correlation coefficient and power unbalance between the two diversity branches.

REFERENCES

[1] E. C. Van Der Meulen, "Three-terminal communication channels," *Advances in applied Probability*, vol. 3, no. 1, pp. 120-154, Spring 1971.

[2] I. Krikidis, H. A. Suraweera, P. J. Smith, and C. Yuen, "Full-duplex relay selection for amplify-and-forward cooperative networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4381-4393, December 2012.

[3] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659-672, 2006.

[4] S. N. Datta, S. Chakrabarti, and R. Roy, "Comprehensive error analysis of multi-antenna decode-and-forward relay in fading channels," *IEEE Communications Letters*, vol. 16, no. 1, pp. 47-49, January 2012.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927-1938, 2003.

[6] V. N. Q. Bao, D. H. Bac, T. D. Thuan *et al.*, "Performance analysis of partial relay selection with multi-antenna destination cooperation," in *International Conference on ICT Convergence (ICTC)*. IEEE, 2011, pp. 101-105.

[7] A. Adinoyi and H. Yanikomeroglu, "Cooperative relaying in multi-antenna fixed relay networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 533-544, February 2007.

[8] S. S. Ikki and M. H. Ahmed, "Performance analysis of cooperative diversity with incremental-best-relay technique over Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2152-2161, August 2011.

[9] M. Ilic-Delibasic, M. Pejanovic-Djurisic, and R. Prasad, "A novel method for performance analysis of OFDM polarization diversity system in Ricean fading environment," *Wireless Personal Communications*, vol. 63, no. 3, pp. 751-764, April 2012.

[10] X. Dong and N. C. Beaulieu, "Optimal maximal ratio combining with correlated diversity branches," *IEEE Communications Letters*, vol. 6, no. 1, pp. 22-24, January 2002.

[11] W. C. Lee and Y. Yeh, "Polarization diversity system for mobile radio," *IEEE Transactions on Communications*, vol. 20, no. 5, pp. 912-923, October 1972.

[12] J. Jootar and J. R. Zeidler, "Performance analysis of polarization receive diversity in correlated Rayleigh fading channels," in *IEEE Global Telecommunications Conference (GLOBECOM'03)*, vol. 2. IEEE, November 2003, pp. 774-778.

[13] L. C. Lukama and D. J. Edwards, "Performance of spatial and polarization diversity," in *Wireless Personal Multimedia Communications (WPMC01)*, September 2001.

[14] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels (2nd ed.)*. John Wiley & Sons, 2005.

[15] M. O. Hasna and M.-S. Alouini, "End-to-end performance of transmission systems with relays over Rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1126-1131, November 2003.

[16] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products (7th ed.)*. San Diego, CA: AcademicPress, 2000.

[17] M. O. Hasna and M.-S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1963-1968, November 2004.

# Estimation of Radio Signal Spatial Local Mean

Pamela Njemcevic

Faculty of Electrical Engineering, University of Sarajevo  
Sarajevo, Bosnia and Herzegovina

Vlatko Lipovac

Faculty of Electrical Engineering, University of Dubrovnik  
Dubrovnik, Croatia

**Abstract**—The accurate estimation of RF signal spatial local mean has a crucial importance in wireless communication systems planning, design and operation: within coverage planning tools, for channel access and power control, in handoff algorithms, etc. However, to date, no universally applicable and verified approach for calculation of relevant averaging parameters is developed. In this paper, detail and comprehensive literature overview of spatial averaging problem is presented, with a critical review of all overviewed methods for its solving. As a result, closed-form expressions for averaging parameters calculation are delivered based on the approaches which are evaluated as the most accurate ones. Finally, the validity domains of these expressions are tested, indicating that even the most accurate and comprehensive ones have limited applicability for averaging parameters calculation just under the specific propagation conditions.

## I. INTRODUCTION

The overall spatial variations of the received RF signal are caused by shadowing and multipath propagation effects. The shadowing-caused variations correspond to slow variations of the signal local mean  $m(x)$  that is noticeable only over the large-scale receiver location changes (of the order of tens of wavelengths), implying the constant spatial local mean within these distances. On contrary, multipath fading  $r_0(x)$  occurs on a small-scale change of distance between the transmitter and receiver (of the order of less than signal wavelength), when constructive and destructive interferences of multiple propagation paths cause fast signal variations even within distances with constant local means, making the overall spatially variable received signal a product of  $m(x)$  and  $r_0(x)$ . However, for a number of algorithms (e.g. in handoff and power control or in coverage planning tools) it is very important to accurately estimate the local mean of the received signal, i.e. to average the overall signal so that the multipath-caused variations are smoothed sufficiently, while preserving the intact of those caused by shadowing.

However, the problem of estimation of accurate signal spatial local mean has not been addressed thoroughly. Namely, after the identification of spatial averaging problem in 1968 [1] and definition of averaging parameters in 1985 [2], only few papers address the issue, providing limited applicability analytical [3] - [8], [15] - [18] and empirical [9] - [12] guidelines for identified averaging parameters calculation. Thereby, no paper providing a detailed comprehensive literature overview and critical review of both analytical and empirical approaches for all averaging parameters calculation is published, neither

the ones giving closed-form expressions for their calculation in specific propagation conditions.

Accordingly, in Section II, to date published analytical and empirical approaches for calculation of all relevant averaging parameters are presented and reviewed. In Section III, closed-form expressions for determination of identified parameters are derived based on the presented approaches evaluated as the most accurate ones. Utilizing the existing empirical values of channel parameters from literature, in Section IV, the averaging parameters in various propagation environments are calculated using the derived expressions, indicating that even the most accurate ones can be used only in the specific propagation conditions. The conclusions are summarised in Section V.

## II. APPROACHES FOR LOCAL AVERAGING PARAMETERS CALCULATION - AN OVERVIEW

In order to estimate the radio signal spatial local mean, values of three basic averaging parameters [2]: averaging length -  $2L$ , the number of samples which must be averaged on  $2L - N$  and the necessary distance between samples -  $d$ , must be calculated for treated propagation environment. However, since  $2L = Nd$  [2], it is sufficient to provide the analytical expressions or empirical methods for determination of only two of these parameters, while the third one can be calculated using the given equation.

### A. Approaches for averaging length calculation

1) *Analytical approaches*: The first analytical approach for calculation of the length at which the overall signal should be averaged in order to determine its local mean is given in [3]. In there, it is recommended to estimate the local mean  $\hat{m}(x)$  by integrating  $r(x) = m(x)r_0(x)$  along a certain length  $2L$ , keeping in mind that  $\hat{m}(x)$  approaches to its true mean  $m(x)$  for properly chosen averaging length  $2L$ . Since  $\hat{m}(x)$  is Gaussian random variable with the mean  $m(x) = \langle r(x) \rangle$  and the variance  $\sigma_{\hat{m}}^2$  [3], wherein  $\sigma_{\hat{m}}^2$  can be expressed as a function of the autocorrelation function  $R_r(x)$  and the mean value  $\langle r(x) \rangle$  of  $r(x)$  [eq. (9), 3], the parameter called  $1\sigma_{\hat{m}}\text{spread} = f(R_r(x), \langle r(x) \rangle)$  [eq. (18), 2] was proposed as a metric for determination of the appropriate averaging length, delivered based on a condition that 68% of  $\hat{m}(x)$  values should be within the interval  $(m(x) - \sigma_{\hat{m}}, m(x) + \sigma_{\hat{m}})$  around its true mean  $m(x)$ . Accordingly, after inserting theoretical expressions for  $R_r(x)$  and  $\langle r(x) \rangle$  for chosen distribution of

$r(x)$ , with  $x = 2L$ , into  $1\sigma_{\hat{m}}\text{spread}(2L)$ , it is finally recommended to determine the appropriate value of averaging length by fixing  $1\sigma_{\hat{m}}\text{spread}(2L)$  to a value from which deviations of  $\hat{m}(x)$  from  $m(x)$  can be considered as acceptable.

Thereby, in [2] and [3], the proposed metric was used under the assumptions of constant  $m(x)$  on any  $2L$  (including values at which, in reality,  $m(x)$  starts to vary) and Rayleigh-distributed  $r_0(x)$  with its mean  $m(x)$ , which make  $r(x)$  also Rayleigh distributed, with its easy to calculate  $R_r(x)$  and  $\langle r(x) \rangle$ . That way,  $1\sigma_{\hat{m}}\text{spread}(2L)$  was obtained as monotonically decreasing function of  $2L$  (Fig. 5 in [2]) and  $40\lambda$  was pronounced as the appropriate averaging length, at which the estimation error drops to its acceptable value of  $1dB$  [3]. However, since multipath fading is Rayleigh distributed and shadow fading is constant within the length of  $40\lambda$  only in some urban environments, applicability of the above recommendation is limited to a very specific propagation conditions and cannot be used as a general reference for the local averaging.

Adopting the metric (in term of  $1\sigma_{\hat{m}}\text{spread}(2L)$  function) and the assumption of  $m(x)$  constancy on any  $2L$  from [3], in the subsequent years, described work has been extended by incorporating Rican [4], [5] and Nakagami-m [6] multipath fading models and the associated theoretical expressions for  $R_r(x)$  and  $\langle r(x) \rangle$ . As a result, sets (instead of just one) of monotonically decreasing  $1\sigma_{\hat{m}}\text{spread}(2L)$  functions dependent on parameter reflecting the amount of multipath in treated channel are obtained (Fig. 4 in [4] and [5] and Fig. 1 in [6]), enabling us to choose different values of preferable averaging lengths for different propagation conditions. Accordingly, in [4], the value of  $20\lambda$  is chosen to be the recommended averaging length for all channels with the amount of multipath smaller or equal to Rayleigh channels, since any further increase in  $2L$  does not bring significant improvement on the  $1\sigma_{\hat{m}}\text{spread}(2L)$  [4]. Similarly, in [5], lengths between  $20\lambda$  and  $40\lambda$  are recommended for local averaging in Rayleigh and better than Rayleigh microcell environments. However, in both papers, despite the extensive analytical calculations, recommendations for selection of preferable averaging lengths are generated by visual inspection of few obtained  $1\sigma_{\hat{m}}\text{spread}(2L)$  curves instead of being based on unambiguous mathematical criteria, which makes their accuracy questionable.

On the other side, recommendations that in suburban/rural environments with small amount of multipath accepted estimation error is at  $2L$  between  $10\lambda$  and  $20\lambda$ , in urban environments with Rayleigh multipath fading at  $40\lambda$  and in indoor, at  $20\lambda$  [6], are based on the established mathematical criterion ( $1 - 0.2 \text{ dB} \leq 1\sigma_{\hat{m}}\text{spread}(2L) \leq 1 + 0.2 \text{ dB}$ ) proposed in [6]. Besides, the expression for  $1\sigma_{\hat{m}}\text{spread}(2L)$ , delivered in [6], is used for averaging lengths calculation not just in channels with small amount of multipath, but also in worse than Rayleigh multipath conditions, thus providing the most comprehensive and accurate results regarding the spatial averaging among all previously listed. However, in [6], as well as in all the above presented papers, only multipath-caused

variations are considered when calculating  $1\sigma_{\hat{m}}\text{spread}(2L)$  and lengths at which local averaging should be done, while shadowing-caused variations are totally neglected by assuming the local mean  $m(x)$  unchanged for any  $2L$ . However, in reality, after the increase of distance, dependent on shadowing characteristics, local mean also starts to vary. Therefore, the presented recommendations [6] are conductive only as averaging length lower bounds that enable good enough averaging out of the multipath fading ( $2L_{mp}$ ), while for determination of the appropriate averaging lengths ( $2L_a$ ), shadowing-caused variations on  $2L$  also should be considered [7].

Accordingly, in [7], the system for measurements' processing, composed of envelope-detector, amplifier and low-pass filter is suggested, with filter parameters chosen to enable averaging out of multipath-caused variations, while preserving intact those caused by shadowing. Thereby, in order to select such a parameters, analytical expressions for their determinations (based on the assumptions that multipath fading is Rayleigh and shadow fading is lognormal distributed process) are derived [7]. They show that the local mean estimation error is a function of averaging length  $2L$ , as well as shadowing parameters: its standard deviation  $\sigma$  and correlation distance  $X_c$  [eq. (17), (A3) (A11), 7], being at first monotonically decreasing and than monotonically increasing function of  $2L$ . In these circumstances, for some fixed values of  $\sigma$  and  $X_c$ , the optimal averaging length is calculated by minimising estimation error obtained for chosen set of parameters, indicating that for typical urban environment in UHF band, the appropriate averaging length is  $2\lambda$ , while that for distances greater than  $2\lambda$ , the local mean estimation error starts to increase again. The above finding is very important because (opposite to described approaches which provide monotonically decreasing estimation error function, leading to conclusion that  $1\sigma_{\hat{m}}\text{spread}(2L) \rightarrow 0$  for  $2L \rightarrow \infty$ ) it provides the analytical evidence that averaging length cannot be adopted as large as desired in order to sufficiently smooth variations due to multipath. Accordingly, it become obvious that, beside multipath fading, shadowing-caused variations also must be considered when determining the appropriate averaging length for treated environment. However, despite the undeniable contribution of approach and findings from [7], obtained recommendations based on the proposed processing scheme strongly differ from all analytical and empirical results and, to date, have not found application in practice.

Finally, in [8], a novel approach for the appropriate averaging length  $2L_a$  calculation is proposed, suggesting its selection between the lower averaging bound  $2L_{mp}$  - from which multipath variations can be considered sufficiently smoothed (which can be calculated using the approach from [3]) and the upper averaging bound  $2L_{sh}$  - to which shadow fading can be assumed as highly correlated process (with its correlation coefficient lower or equal to 0.8). Accordingly, assuming the lognormal distributed shadow fading and Gudmundson correlation model [19], in [8], expression for calculation of  $2L_{sh}$ , dependent on shadow fading parameters  $\sigma$  and  $X_c$ , is delivered [eq. (8), 8]. It is then applied on the specific

values of fading parameters taken from the literature, providing recommendations for  $2L_{sh}$  in different environments (in indoor  $2L_{sh} \approx 1 m$ , in urban cells  $2L_{sh} \approx 5 m$  and in suburban/rural environments  $2L_{sh} \approx 30 m$ ) and possibility to calculate the appropriate value of averaging length  $2L_a$  for treated environment. However,  $2L_a$  values recommended in [8] are actually chosen between  $2L_{mp}$  and  $2L_{sh}$  only for suburban and rural cells with UHF band, while in urban and indoor cells,  $2L_a$  values equal to  $2L_{sh}$  and lower than  $2L_{mp}$  are chosen. Besides, approach proposed in [8] is not tested for  $2L_a$  determination in others than UHF frequency bands, which is why its validity domain must be additionally investigated.

2) *Empirical approaches:* Beside the presented analytical approaches, few empirical methods for determination of the appropriate averaging lengths are also proposed. According to the earliest one, spatial local averaging should be done at  $2L$  at which the overall measurement samples multipath distribution is closest to Rayleigh [9], [10]. Applying the proposed approach on field measurements performed in urban areas at 108 MHz, 455 MHz, 900 MHz [9] and 108 MHz [10], it is recommended to perform spatial averaging at  $20\lambda$ ,  $30\lambda$ ,  $60\lambda$  and  $40\lambda$  for treated frequencies. However, the above recommendations are delivered by visual inspection of empirically obtained distribution curves with no established metric and decision criteria. Besides, applicability of proposed method is limited to environments with Rayleigh distributed variations due to multipath, making the presented approach, as well as the obtained results [9], [10], inappropriate for further averaging.

On the other side, the empirical approach proposed in [11] provides much more accurate results. There, the true mean values  $m(x)$  are calculated by averaging 1000 of individual RF measurements while moving the transmitting and receiving antennas over a circular, horizontal region several wavelengths across, while the appropriate averaging length  $2L_a$  is determined as the lengths at which the difference between  $m(x)$  and  $\hat{m}(x)$  values estimated using described linear averaging at different  $2L$  lengths, is minimal. However, due to complexity of required measurement equipment and long duration of measurement procedure, presented approach is applicable only in indoor environments (where spatial local averaging should be done at  $10\lambda$  [11]).

Besides, the appropriate averaging length can be obtained as the one for which the empirical  $1\sigma_{\hat{m},spread}(2L)$  function, generated using  $R_r(x)$  and  $\langle r(x) \rangle$  values calculated from the measurement samples, is minimal [12]. However, to date, this approach is applied only on measurement samples propagating at low frequencies with MW and HF bands, providing the following recommendations: for signal propagating in suburban/rural area at 1359 kHz, averaging should be done at  $2\lambda$  [12], for those propagating at 810 kHz, 1060 kHz and 1260 kHz in urban cells between  $\lambda$  and  $2.5\lambda$  [13], while for signal propagating at 26 MHz, local mean estimation error is minimal at  $2L$  close to  $7\lambda$  [14].

However, despite the accuracy of proposed empirical methods ([11], [12]), they require acquisition of many overall signal

samples, which is time consuming and requires the existence of already installed target system in use. Contrary, for the most of the above mentioned applications, accurate a priori estimation of the local mean is necessary, which implies lower significance of empirical methods for  $2L_a$  estimation and motivates development of accurate analytical  $2L_a$  prediction.

### B. Approaches for number of samples calculation

1) *Analytical approaches:* Beside the necessity for determination of the appropriate averaging length, for accurate local mean estimation, it is also necessary to average the appropriate number of samples  $N$  on  $2L_a$ . Thereby, the first analytically delivered recommendation for determination of  $N$  can be found in [2], where it is suggested to estimate the local mean by using the simplest unbiased estimation method based on samples averaging. According to the suggested method, sampling average  $\bar{r}(x)$  can be determined by averaging the appropriate number of overall signal samples  $r(x)$  taken within the chosen averaging length [eq. (7), 6]. Assuming that the number of these samples is always greater than 6, according to central limited theorem,  $\bar{r}(x)$  is normally distributed, with an easy-to-calculate probability of its deviation of a true mean. Considering the above and assuming that signal samples are uncorrelated, it is then suggested to calculate  $N$  as the value at which 90% of sampling average is within  $\pm 1 dB$  of its true mean, which is shown to be dependent on the mean value  $m_r$  and the variance  $\sigma_r$  of  $r(x)$  [eq. (21), (22), (27), 2]. Thereby, since on appropriately chosen averaging length, the shadowing-caused variation can be neglected, it is finally concluded that  $r(x)$  distribution can be considered identical as assumed distribution of multipath fading and that for Rayleigh distributed variations due to multipath, it is necessary to average 36 of signal samples on chosen  $2L$ . However, in order to convert necessary metrics ( $m_r$  and  $\sigma_r$ ) characterising Rayleigh distributed samples, from linear to logarithmic scale, in [2], unusual transformation is performed [eq. (20), 2], limiting the accuracy of the above recommendation.

The mentioned shortcoming is overcome in [15] and [6], where numbers  $N$  for Rayleigh distributed samples taken in linear ( $N = 57$ ) and logarithmic domain ( $N = 85$ ) and for Nakagami-m distributed variations due to multipath ( $N = 20$  in rural cells and  $N = 40$  suburban and indoor cells) are calculated based on the described averaging procedure [2] and the assumption about uncorrelated samples, but with appropriately performed transformation between metrics characterising distributions in both scales.

Additionally, beside samples averaging estimator assumed in [2], [15] and [6], local mean can be determined using the minimum variances or maximum-likelihood estimators proposed in [16], [17], for Rayleigh [eq. (3), 16] and Nakagami-n [eq. (21), (31), 17] multipath fading channels. It is found that significantly fewer samples are required for given estimation accuracy when using the minimum variance and maximum-likelihood instead of sample averaging estimator. However, the derivation of these estimators assume perfect knowledge of the fading severity on the receiver, indicating that for the unknown

amount of multipath (which is the case in practical systems) samples averaging estimator yields better performance results and should be used for local averaging [17].

At the end of this subsection it is also important to stress that all listed approaches for calculation of  $N$  are based on the assumption that the overall signal samples are distant enough so they can be considered as independent. However, for small  $2L_a$  and large  $N$ , correlation between succeeding samples might occur. Thus, in [18], the effect of samples' correlation on number  $N$  is investigated, showing that for Rayleigh distributed multipath fading, it is necessary to average more correlated than uncorrelated signal samples for the same estimation accuracy (Fig. 4 from [18]). This finding should be considered in environments with severe fading conditions.

2) *Empirical approaches:* In addition to analytical approaches for determination of  $N$ , in [12], empirically based procedure relying on equation for calculation of  $N$  delivered in [15] (for receiver with a linear characteristics), is proposed. However, unlike in [15], [6], where  $m_r$  and  $\sigma_r$  are calculated based on the assumed theoretical multipath fading distribution, their values in [12] are determined from measurement samples, indicating that in MW band,  $N$  should be equal to 20, 11 [13] and 8 [12] in urban, suburban and rural cells, respectively, while in urban cell with signal propagating in HF band,  $N$  should be equal to 21 [14].

### C. Approaches for distance between samples calculation

1) *Analytical approaches:* When it comes to parameter  $d$ , in [2] and [6], its values are calculated as a ratio between previously determined averaging lengths and numbers  $N$ , confirming that, for thus obtained distances, overall signal samples can be considered uncorrelated (since the distances between them are greater than values which correspond to the first null of their autocorrelation coefficient [15]).

On the other side, in [4] and [5], expressions for calculation of  $d$  instead of  $N$  are delivered. In there, already described  $1\sigma_{\bar{m}}\text{spread}(2L)$  metric, calculated by replacing  $2L$  with  $Nd$  and fixing its length to  $20\lambda$ , is used for determination of  $d$  in Rician multipath channels. It is showing that  $d$  should be equal to  $0.5\lambda$ , since its further increment does not bring significant reduction in the  $1\sigma_{\bar{m}}\text{spread}(2L)$  [4]. However, since the above recommendation is obtained for fixed averaging length, its applicability is strictly limited to environments in which  $2L_a = 20\lambda$ .

2) *Empirical approaches:* Finally, in [12], empirical method for determination of distance  $d$  is proposed, suggesting that it should be equal to the length at which empirical correlation coefficient of multipath fading samples (obtained after the normalization of overall field samples with its estimated local mean) drops to 0.2.

### III. EXPRESSIONS FOR SPATIAL LOCAL AVERAGING PARAMETERS CALCULATION

Based on the above assessed literature, the most accurate and comprehensive analytical expressions for averaging parameters calculation are chosen as a reference for estimation

of signal spatial local mean. Accordingly, the appropriate averaging length  $2L_a$  should be determined as the value between:

$$2L_{mp} \leq 2L_a \leq 2L_{sh} \quad (1)$$

Thereby, assuming Nakagami distributed multipath fading and the fact that the acceptable estimation error obtained after its averaging on chosen length is close to 1 dB [3], lower averaging bound  $2L_{mp}$  can be calculated by solving the following equation:

$${}_2F_3\left(\frac{1}{2}, \frac{1}{2}; 1, 1, \frac{3}{2}; -2\pi\frac{2L_{mp}}{\lambda}\right) - 1/2 \left[ J_0\left(2\pi\frac{2L_{mp}}{\lambda}\right)^2 + J_1\left(2\pi\frac{2L_{mp}}{\lambda}\right)^2 \right] \approx 0.008 \quad (2)$$

(in which  $m$  reflects the amount of multipath,  ${}_2F_3(\cdot)$  presents hypergeometric function and  $J_n(\cdot)$   $n$ -th order Bessel function of the first kind), delivered by combining eq. (2) - (6) from [6] and solving the obtained integral.

Upper averaging bound  $2L_{sh}$ , for lognormal distributed shadow fading and Gudmundson correlation model [19] can be calculated from [eq. (8), 8] as:

$$2L_{sh} = -X_c \ln \left[ \frac{\ln(0.8 \exp(0.1325\sigma^2) + 0.2)}{0.1325\sigma^2} \right] \quad (3)$$

assuming that shadowing-caused variations can be neglected if their correlation coefficient is equal or greater than 0.8 [8].

Finally, the minimum number of uncorrelated samples, which should be averaged on chosen averaging length for obtaining 90% of estimated sampling average within  $\pm 1$  dB of its true mean, can be calculated by solving [eq. (8), 6], as:

$$N = 207.2 \left[ \frac{m - (\Gamma(m+0.5)/\Gamma(m))^2}{(\Gamma(m+0.5)/\Gamma(m))^2} \right] \quad (4)$$

while  $d$  can be determined as:

$$d = 2L_a/N \quad (5)$$

and should be large enough to guarantee uncorrelation between succeeding overall signal samples [15].

### IV. APPLICABILITY OF PROPOSED EXPRESSIONS FOR AVERAGING PARAMETERS CALCULATION IN DIFFERENT PROPAGATION CONDITIONS

In order to assess the appropriateness of listed analytical expressions for local averaging parameters calculation in certain propagation environment, adequate values for the relevant shadowing and multipath parameters ( $m$ ,  $\sigma$  and  $X_c$ ) characterizing each type of environment with UHF band (taken from [8] and listed in Table I) are inserted into (2) - (4) and the results, for the worst case multipath and shadowing conditions, are given in Table II.

From Table II it can be concluded that in suburban and rural outdoor environments in which signals are propagating



TABLE I  
SHADOWING AND MULTIPATH PARAMETERS VALUES

Types of cells	$f$ (Ghz)	$\sigma$ (dB)	$X_c$ (m)	$m$
Rural	0.9 - 2.4	$\leq 10$	250	4 - 8
Suburban	0.9 - 2.4	$\leq 9$	200	2 - 8
Urban	0.9 - 2.4	$\leq 8$	30	1
Indoor	1.8 - 5.2	$\leq 8$	5	2 - 4

TABLE II  
RECOMMENDED AVERAGING PARAMETERS VALUES

Types of cells	$2L_{mp}$	$2L_{sh}$	$N$
Rural	$9\lambda$	$95\lambda - 255\lambda$	14
Suburban	$20\lambda$	$85\lambda - 225\lambda$	28
Urban	$45\lambda$	$14\lambda - 37\lambda$	57
Indoor	$20\lambda$	$5\lambda - 14\lambda$	28

in UHF band, accurate local mean can be estimated using the averaging parameter values calculated by proposed analytical expressions (1) - (5). Namely, for these environments, the appropriate averaging length  $2L_a$  can be chosen in between  $2L_{mp}$  and  $2L_{sh}$ , since  $2L_{mp} < 2L_{sh}$ . Besides, in these cell types,  $N$  can be calculated using (4), since for all possible  $2L_a$  values, distances  $d$  obtained using (5) is large enough so the samples can be considered as uncorrelated.

However, from Table II it can be seen that in outdoor urban and indoor cells  $2L_{mp} > 2L_{sh}$  and that it is not possible to concurrently satisfy both criteria: to obtain multipath-caused estimation error close to 1 dB [3] and shadowing correlation equal to 0.8 [8], and to proceed with estimation of  $2L_a$  that would have been appropriate for these propagation environments. In these circumstances, it is possible to prefer one criterion over the other, which is done in [8] by choosing  $2L_a = 2L_{sh}$ . However, since it is not clear whether it is more important to preserve shadow fading constancy or to sufficiently average out multipath variation within the local mean, due to absence of accurate analytical background,  $2L_a$  values recommended in [8] cannot be taken as a reference.

Besides, (1) - (3) cannot be used for determination of the appropriate averaging lengths also in MW and HF bands, since for these frequencies, regardless of the type of the cell,  $2L_{sh} \ll \lambda$  and consequently,  $2L_{mp} > 2L_{sh}$ . Additionally, the existing expressions also cannot be used for determination of  $N$  and  $d$  in these bands, due to differences between their values calculated by (4) - (5) and those obtained empirically [12] - [14].

## V. CONCLUSION

In this paper, the most accurate and comprehensive expressions for determination of averaging parameters relevant for local mean estimation are selected based on the assessed literature on spatial averaging problem. Applying the empirically determined input parameters from the literature onto these expressions results in recommendations that local mean with neglected multipath and shadowing caused variations in rural and suburban cells in UHF band can be estimated by averaging 28 of equally distant overall signal samples

at the length between  $20\lambda$  and  $85\lambda$ . However, it is also shown that chosen expressions, as well as all others available in the literature, lack generality with respect to propagation environments and frequency bands, since neither of them can be used for averaging parameters calculation in UHF indoor and urban cells as well as in MW and HF bands. Accordingly, due to limited applicability of so far proposed expressions, it is recommended to develop a novel analytical approach for spatial averaging parameters calculation, which should be applicable under all propagation conditions.

## REFERENCES

- [1] R. H. Clarke: "A statistical Theory of Mobile-Radio Reception", The Bell System Technical Journal, Vol. 47, No. 6, pp. 957-1000, July 1968
- [2] W. C. Y. Lee: "Estimate of Local Average Power of a Mobile Radio Signal", IEEE Transactions on Vehicular Technology, Vol. 34, No. 1, pp. 22 - 27, February 1985
- [3] W. C. Lee, Y. S. Yeh: "On the Estimation of the Second-Order Statistics of Log normal Fading in Mobile Radio Environment", IEEE Trans. on Communications, Vol. 22, No. 6, pp. 869-873, Jun, 1974
- [4] C. Tepedelenioglu, A. Abdi, G. B. Giannakis, M. Kaveh: 'Estimation of Doppler Spread and Signal Strength in Mobile Communications with Applications to Handoff and Adaptive Transmission', Wireless Comm. and Mobile Computing, John Wiley and sons, pp. 221-242, 2001
- [5] M. D. Austin, G. L. Stuber: "Velocity Adaptive Handoff Algorithms for Microcellular Systems", IEEE Transactions on Vehicular Technology, Vol. 43, No. 3, pp. 549-561, August 1994
- [6] P. Njemcevic: "Local Average Signal Estimation in Nakagami-m Channels", in Proc. on IEEE 6th International Symposium on Communications, Control, and Signal Processing, Athens, Greece, 2014
- [7] A. J. Goldsmith, L. J. Greenstein, G. J. Foschini: "Error Statistics of Real-Time Power Measurements in Cellular Channels with Multipath and Shadowing", IEEE Transactions on Vehicular Technology, Vol. 43, No. 3, pp. 439 - 46, August 1994
- [8] P. Njemcevic: "A Novel Approach in Determination of the Appropriate Spatial Averaging Signal Length", Wireless Personal Communications, Vol. 82, No. 3, pp. 1851-1861, January 2015
- [9] J. D. Parsons, M. F. Ibrahim: "Signal Strength Predictions in Built-up Areas", IEEE Proceedings, Vol. 130, No. 5, pp. 377 - 384, August 1983
- [10] B. R. Davis, R. E. Bogner: "Propagation at 500 MHz for Mobile Radio", IEEE Proceedings, Vol. 132, No. 5, pp. 307 - 320, August 1985
- [11] R. A. Valenzuela, O. Landron, D. L. Jacobs: "Estimation Local Mean Signal Strength of Indoor Multipath Propagation", IEEE Transactions on Vehicular Technology, Vol. 46., No. 1, pp. 203-212, February 1997
- [12] D. Vega, S. Lopez, J. Matias, U. Gil, I. Pena, M. Velez, J. Ordiales, P. Angueira: "Generalization of Lee Method for the Analysis of the Signal Variability", IEEE Transactions on Vehicular Technology, Vol. 58, No. 2, pp. 506 - 516, February 2007
- [13] U. Gil, I. Pena, I. Angulo, P. Angueira, D. de la Vega, D. Guerra: "Statistical Characterization of the Medium Wave Field Strength Spatial Variability in Urban Environments Using The Generalized Lee Method", Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP), pp. 1 - 5, Barcelona, 12 - 16 April 2010
- [14] J. M. Matias, S. Lopez, I. Pena, I. Fernandez, P. Angueira: "Location correction factor for coverage planning tools for DRM", in the 26 MHz band", in Proc. on IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Las Vegas, USA, March 2008
- [15] P. J. D. Parsons: "The Mobile Radio Propagation Channel", John Wiley and Sons Ltd, England, 2000
- [16] D. Wong, D. C. Cox: "Estimating Local Mean Signal Power Level in a Rayleigh Fading Environment", IEEE Transactions on Vehicular Technology, Vol. 48, No. 3, pp. 956 - 959, May 1999
- [17] Y. Ko, M. S. Alouini: "Local Mean Signal Estimation over Nakagami-m Fading Channels", Wireless Communications And Mobile Computing, John Wiley and Sons, pp. 367 - 374, March 2006
- [18] B. Ai, Z. D. Zhong, G. Zhu, M. Zhao: "Novel Statistical Criteria for Local mean Power Estimation in Wireless Coverage Prediction", IET Microwaves, Antennas and Prop., Vol. 5, No. 5, pp. 596 - 604, 2011
- [19] M. Gudmundson: "Correlation models for shadow fading in mobile radio systems", Electronic Letter, Vol. 27, No. 23, pp. 21452146, 1991

# Microstrip Antenna with Inserted Slot for Dual-Polarized

Teeravisit Laohapensaeng  
 School of Information Technology  
 Mae Fah Luang University  
 Chiang Rai, Thailand  
 teeravisit.lao@mfu.ac.th

**Abstract**—Dual-polarized microstrip antenna with high isolation is presented. The high isolation can be achieved by inserting a diagonal rectangular slot at one of the patch corner. The antenna is excited by using two electric probes. The antenna is designed to operate at the frequency of 2.4 GHz. The experimental results have shown that, it has more than 30 dB isolation with a moderate antenna gain of 8dBi over the impedance bandwidth of 8.3 %. The results from this study will be applied to the antenna using for the Intelligent Transportation Systems (ITS) in further.

**Keywords**; square patch, inserted slot, isolation improvement

## I. INTRODUCTION

Microstrip antenna using two orthogonal probes feed for dual-polarized has gained considerable interest for frequency reusability and polarization diversity in wireless communication systems [1]. In order to satisfy the requirement for above applications, isolation of 30dB or more between two probes feed is required [2]. Generally, it is difficult to achieve high isolation for this configuration. Thus, substantial efforts have been made to overcome this problem. It is known that, a thin microstrip antenna using two orthogonal probes feed has a good isolation; however, the input impedance bandwidth of the antenna is narrow and the antenna gain is quite low. Thus, increasing the input impedance bandwidth and the antenna gain, an air-gap is adopted by inserting it between patch and the ground plane. Nevertheless, isolation between two probes fed is decreased [3]. To overcome the problem, a new configuration of microstrip antenna, which has high isolation, is proposed. The high isolation is achieved by inserting a slot into the radiating patch. It is formed as a band stop filter connecting between the two probes feed. In this paper, the electrical characteristics of the proposed antenna such as radiation-pattern, input impedance and antenna gain are studied and reported at the ISM band (2.4 GHz).

## II. ANTENNA STRUCTURE AND DESIGN

The geometry of the proposed antenna is shown in Fig.1. It consists of asquare patch which has a diagonal rectangular slot inserted near one of the patch corners, two orthogonal

probes feed and a ground plane. The square patch has a width of  $w$  and is printed on the top side of a thin dielectric substrate, which is placed above a ground plane at a height of  $h_a$ . The conducting posts are used as the supporter (see Fig.1). The substrate has a thickness of  $h_d$  and dielectric constant of  $\epsilon_R$ . The antenna is excited via the two direct probes, which is located at a distance  $D$  from the patch center. The inserted slot has a width of  $w_s$  and a length of  $l_s$ . It is located at between two probes. The width of the ground plane is  $g$ .

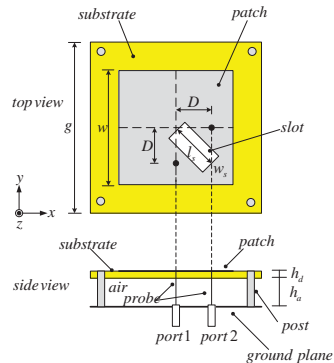


Figure 1 Antenna Structure

When the slot is absent, the antenna can be considered to be a square patch microstrip antenna which has an air-gap inserted between patch and ground plane. The antenna is excited by two orthogonal modes of  $TM_{10}$  (exciting via port 1) and  $TM_{01}$  (exciting via port 2). Their resonant frequencies are the same and mainly determined by the patch size. According to the literature [3], the patch width  $w$  can be calculated when a given resonant frequency  $f_r$  is defined as:

$$w = \frac{c_o}{2f_r \sqrt{\epsilon_{eff}}} - \frac{2(h_d + h_a)}{\sqrt{\epsilon_{av}}} \quad (1)$$

Where  $c_0$  is the speed of light in free space ( $\sim 3 \times 10^8$  m/s),  $\epsilon_{eff}$  is the effective dielectric constant and  $\epsilon_{av}$  is the average dielectric constant, which can be calculated from the equations (2) and (3) respectively.

$$\epsilon_{av} = \frac{\epsilon_R (h_d + h_a)}{h_d + \epsilon_R h_a} \quad (2)$$

$$\epsilon_{eff} = \frac{\epsilon_{av} + 1}{2} + \frac{\epsilon_{av} - 1}{2} \left( 1 + \frac{10(h_d + h_a)}{w} \right)^{-1} \quad (3)$$

The proposed antenna is designed to operate at a frequency of 2.4 GHz. The thin substrate FR4 with thickness of 0.8 mm and dielectric constant of 4.3 is used. By using the equations as shown above, the square patch microstrip antenna which an air-gap is inserted between patch and the ground plane can be designed. Detailed dimensions of the antenna are depicted in Table.1.

Table 1: Dimensions of the antenna

Design frequency = 2.4GHz	
Substrate type = FR4, $\epsilon_R = 4.3$	
$h_f = 0.8$ mm	$h_a = 6$ mm
$w = 47$ mm	$g = 80$ mm

Fig.2 (a) shows the simulated return loss and isolation of the antenna as the function of the 1.8 - 3 GHz frequency range when the probe distances  $D$  are 12 mm and 16 mm, respectively. It can be seen that the probe distance is only for providing the good impedance matching. As the distance  $D$  increases the input return loss at the designed frequency ( $f_0$ ) increases. In both cases, the isolation has a maximum value of the order of 30 dB at the frequency of 2.2 GHz (denote by  $f_o$ ). Therefore, the isolation gets worse below 20 dB over the impedance bandwidth (defined by return loss  $\geq 10$ dB). To improve the isolation to be 30 dB or more over the impedance bandwidth at the designed frequency, a diagonal rectangular slot has been introduced. It is inserted into the radiating patch between two probes (see Fig.1). The inserted diagonal rectangular slot will act as a band stop filter connecting between two ports [4]. Therefore, the high isolation between the two ports should be obtained. Fig.2 (b) shows the return loss and the isolation of the antenna with the same dimensions as the previous antenna, except the slot is included and the probe distance is fixed to 16 mm. As illustrated in the figure the presence of the slot, mainly affects the isolation. The high isolation frequency point  $f_o$  is moved up from 2.2 GHz to 2.4 GHz, while the input return loss of the antenna seems to be maintained at the same order of both the frequency and amplitude when compared to the case without the slot.

In order to obtain the appropriate slot dimensions to gain the accepted isolation between the two ports ( $\geq 30$ dB), the contour of isolation as the function of various slot dimensions are calculated. The result is shown in Fig.3. From the result, it

might be concluded that to obtain the accepted isolation when the dimensions of the antenna as listed in Table.1, the width and the length of the slot must be in the range 13 - 15 mm and 14 - 20 mm, respectively. In this design the slot width is selected to be 14 mm, and the slot length 18 mm. The high isolation between the two ports of 42 dB at the designed frequency (2.4GHz) is obtained. It is increased by 20 dB compared to the case without the slot.

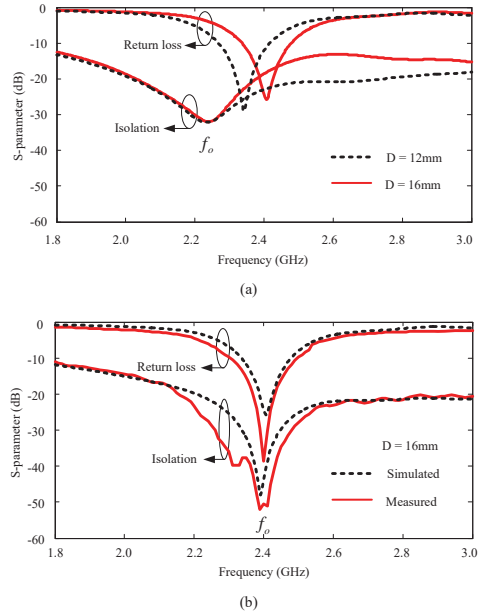


Figure 2 Return loss and Isolations of the antenna (a) without inserted slot and (b) with inserted slot

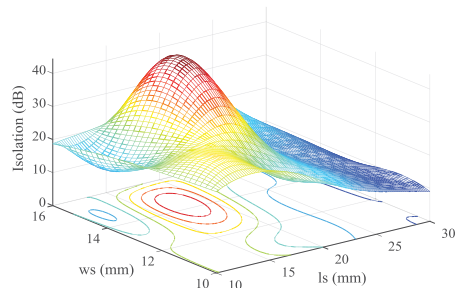
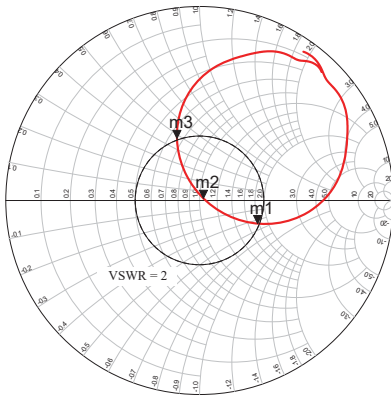


Figure 3 Contour of isolation for various slot dimensions



Marker	Frequency (GHz)	Input Impedance (Ohm)
1	2.30	91.45 + j 23.00
2	2.40	51.90 + j 02.20
3	2.50	32.07 + j 23.98

Figure 4 Input impedance

Fig.4 shows the simulated input impedance of the antenna versus the frequencies in the 1.8 – 3 GHz range. The input impedance of the antenna is 51 Ohm at the center frequency. The antenna acts as the capacitor when the operating frequency is lower than the center frequency and opposite when the frequency is higher than. VSWR is a term related to the ability to power transferring between the antenna and the feeding network. Practically,  $VSWR \leq 2$  is accepted. As can be seen that, to gain  $VSWR \leq 2$  the antenna must be operated in the frequencies 2.30 – 2.50 GHz range.

### III. EXPERIMENT RESULTS

In order to verify the validity of the design, a 2.4 GHz prototype antenna was fabricated and tested. The square patch with a diagonal rectangular slot was printed on the FR4 substrate. The FR4 has a dielectric constant of 4.3 and a thickness of 0.8 mm. The slot width and the slot length are 14 mm and 18 mm, respectively. The patch is placed above a ground plane at a height of 6 mm. The conducting posts are used as the supporter. The antenna is excited through two 50 Ohm SMA connectors from underneath the ground plane, which is located at a distance of 16 mm from the patch center. A photograph of the prototype antenna is shown in Fig.5.

The input return loss and the isolation of the prototype antenna were measured in the 1.8 – 3 GHz frequency range by HP8270C Network Analyzer. The measured results are compared with the simulated results as shown in Fig.2 (b) by

the dot lines. As can be seen in the figure the measured results are slightly higher than the simulated one. However, both results show the same trend. The isolation is more than 30 dB over the impedance bandwidth of 8.3% from 2.3 – 2.5 GHz.

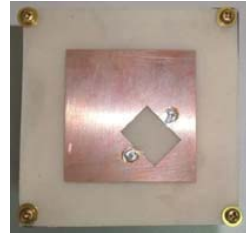


Figure 5 Prototype antenna

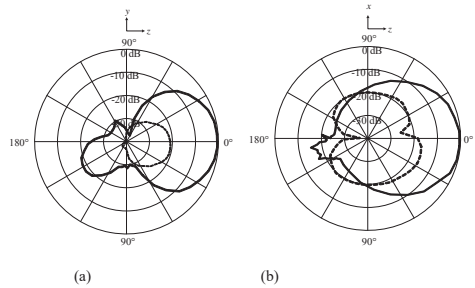


Figure 6 Radiation patterns for port 1 excitation at 2.4GHz  
(a) E-plane (b) H-plane

Fig. 6 shows the measured co-and cross-pol radiation patterns for port 1 of the prototype antenna at 2.4 GHz. It was measured in the anechoic chamber. The far field range of 1 m was used. A standard dipole antenna was employed as the transmitted antenna. Both the transmitted antenna and the antenna under test were installed on a 1.5 m high wooden stand. The equipments to measure the radiation file of the antenna is shown in Fig.7.

As shown in Fig. 6, the proposed antenna possesses the peak antenna gain of around 8 dBi. The E-plane half-power beam-width is  $57.2^\circ$  and the cross-pol (dot line) is less than 20 dB when compared to the co-pol (solid line) level. The H-plane half-power beam-width is  $60.3^\circ$  and the cross-pol (dot line) is less than 17 dB when compared to the co-pol (solid line) level.

Fig.8 shows the measured antenna gain as a function of frequency in the 1.8 – 3.0 GHz frequency range. The Friis transmission formula was used in the calculations to achieve the measured antenna gain. The measured antenna gain was done in the anechoic chamber. As shown in the figure, the

proposed antenna has the moderate antenna gain of around 8 dBi over the impedance band width (defined by return loss  $\geq 10$ dB) from 2.3 – 2.5 GHz.

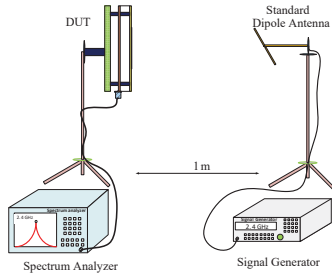


Figure 7 Setup the measured of radiation field

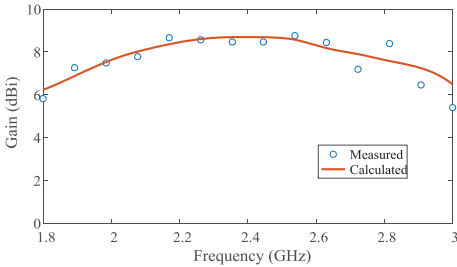


Figure 8 Antenna gain

IV. CIRCULAR-POLARIZAD ANTENNA

In this research work, the author extends the study of the proposed antenna by feeding it through a 3dB Quadrature Hybrid, to form it as a circular - polarized antenna. Then, the input return loss was measured. The measured results compared to the simulated result as shown in Fig. 9.

As it can be seen in the figure, the measured results are slightly narrower than the simulated one. However, both results show the same trend. The slightly different result is might be due to the imperfect of the 3dB Quadrature Hybrid. To verify the being of the circularly polarization wave, the axial ratio of the antenna was measured. The result is shown in Fig.10. It is seen that, the axial ratio is less than 3 dB from 1.8 - 2.9 GHz. It may conclude that the proposed antenna feeding through 3dB Quadrature Hybrid can acts as the circular-polarized antenna.

V. CONCLUSION

A dual-polarized microstrip antenna with high isolation is proposed. The high isolation is achieved by inserting a diagonal rectangular slot into the radiating patch.

Experimental results show that the antenna has more than 30 dB isolation with the moderate antenna gain of 8dBi over the impedance bandwidth of 8.3 % from 2.3 – 2.5 GHz. Also, the proposed antenna can be formed as the circular-polarized antenna when the antenna is feed through the 3dB Quadrature Hybrid. The results of this study will be applied to the antenna using for the smart wireless systems e.g., the Intelligent Transportation Systems (ITS) [5], in further. For instance, the antenna will be integrated with the adaptive circuitry to form the automatic switchable polarized-antenna [6] for improving the signal to noise ratio or connected to the dual - radio transceiver for selecting to use the best signal.

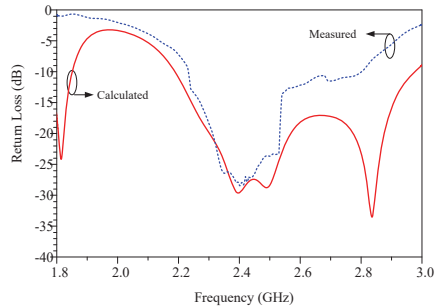


Figure 9 Return loss of the circular-polarized antenna

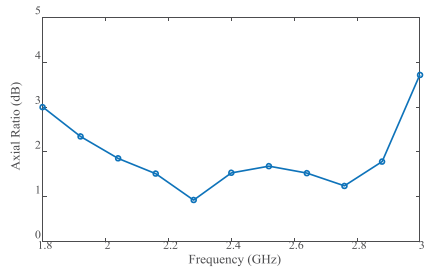


Figure 10 Axial ratio

REFERENCES

- [1] Wong, H., Lau, K.L., and Luk, K.M., "Design of dual-polarised L-probe patch antenna array with high isolation", *IEEE Trans. Antennas Propag.*, 52, (1), pp. 45-52, 2004
- [2] H.-W. Son, "Design of dual-polarised microstrip antenna with high isolation using capacitive feeds", *ELECTRONICS LETTERS*, Vol.45, No.11, May 2009.
- [3] Girish Kumar and K. P. Ray, "Broadband Microstrip Antennas", *Artech House*, 2003, pp. 29 – 60

- [4] K. Boonying, T. Laohapensaeng, S. Kosulvit and C. Phongcharoenpanich, "A Flat Antenna for Polarization Diversity", *JICTEE-2010*, Luangprabang Lao PDR, December 21-24, pp.528-530.
- [5] Stefano Maddio, "A Compact Circularly Polarized Antenna for 5.8 GHz Intelligent Transportation System", *IEEE Antennas and Wireless Propagation Letters*", 2016
- [6] T. Laohapensaeng and K. Boonying, "Dual-Polarized Microstrip Antenna with High Isolation Using an Inserted Slot", *2011 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, December, 2011

# *A Contemporary Survey on Accuracy Demands of Mobile Navigation Systems*

<sup>1</sup>Karthick Nanmaran, <sup>2</sup>B. Amutha, <sup>3</sup>N. Gokul Ashok

<sup>1,2</sup>School of Computing, SRM University, India

<sup>3</sup>Larsen & Toubro Infotech, India

{<sup>1</sup>karthicknanmaran, <sup>2</sup>bamutha62, <sup>3</sup>gokulashokn}@gmail.com

**Abstract**— The aim of this paper is to provide a brief outlook of the GPS positioning devices used to guide users on ground. When users make their movement, mobile devices are used by them not only for communicating with other people whereas they are used to guide them using GPS technology. The devices are operated using radio frequencies in wireless mode of operation and GPS provide the location information of the user using the number of satellites connected at that point of time. Localization accuracy of the mobile based devices varies from device to device due to so many reasons like environmental conditions, number of satellite connectivity, sampling interval and also the pedestrian movements. This paper provides a thorough investigation on various causes of localization accuracy variations on demand and a way to improve localization accuracy in mobile based GPS appliances. We also interrogate the noise reduction filters used for improving localization accuracy in outdoor environment and provide a design solution for improving localization accuracy in GPS denied environments.

**Keywords**— GPS, Kalman Filter, Android, Visibility, Mobility, Localization accuracy

## I. INTRODUCTION

Determining the exact location of a person is very important if we want to provide Location Based Services (LBS) to the person. GPS plays an important role in zeroing on the location of a person anywhere on Earth. The accuracy is high in rural areas and low in urban contexts. This is due to the existence of high rises in cities, called otherwise as urban canyons. They often lead to Multipath reflection of GPS signals and hence it becomes difficult to determine where the user is actually present. If the Location Based Service also includes view shed analysis, the problem expands since the GPS might place the user on top of buildings when actually the user is walking along a pavement on the side of a building [1]. He will not be able to see anything what the service suggests. Hence increasing the accuracy of the GPS signal becomes essential. For view shed analysis, it is important not only to know the location but also to determine the facing direction of the user with his mobile phone.

## II. STATE OF THE ART

### A. Localization on Android

To build a public safety system using GPS enabled Smart phones and the android operating system, a proof of concept was proposed in [2]. The Android Operating System (AOS)

and software development environment have evaluated several of its capabilities by constructing a working application. This application collected speed and location information from the GPS receiver, used the Google Maps Application Programming Interface (API) to determine the location of nearby schools, and sounded an alarm if a person drove over the speed limit in a school zone. The platform proved capable of supporting a melding of different services, and the broad applicability of the Smartphone to the public safety problems was recognized.

In [3] Kalman filter techniques have been introduced and the mechanisms for improving the location accuracy thereby reducing the location error through an example was established.

A realizable taxonomy was depicted when the user is in moving. It also shows the simple graphical interface while not overwhelming the user with uncomfortable sensors with the help of Android OS. In [4] the facilities available in android platform for implementing LBS services were discussed in detail. Many applications which use location based services have hit the market every day [5] Research on Mobile Location Service Design Based on Android paper presents the design method of a location based mobile service. The design example shows the ease of implementing self-location, drawing the driving trace, performing query and flexibly controlling the real-time map on Android. A Formal Model to Analyze the Permission Authorization and Enforcement in the Android Framework [6] was proposed using Android permission scheme. On road vehicle navigation was dealt with the limited accuracy requirements in [7].

A Heave Compensation Algorithm Based on Low Cost GPS Receivers provides the information about how low cost receivers are able to provide the required location accuracy. [8] This was the first formalization of the permission Scheme enforced by the Android framework. Quality assessment of a network-based RTK GPS service in the UK proposes the idea behind combining network in to the localization technique of GPS Transceivers. [9]

In the research, how to display the data from database by List View on Android [10] was presented. They presented an optimal method to access and display data from a database using the mapping mechanism. In 2007, a comparative survey on location based systems, the methods and the mathematical concepts have been discussed [11]. Autonomous vehicles in an urban place have been analyzed using GPS receivers in [12]

Android Based Wireless Location and Surrounding Search System Design was presented by Li Xu Dong, Tang Hai and Yan Gaoshi of University of Electronic Science and Technology of China (UESTC). Personal position and monitoring using dead reckoning with step counts have been analyzed using the GPS receivers and Gyroscope. [13]

### B. Localization using Position Sensor

Most previous works focused on constructing taxonomies of location determination techniques by using specific type of positioning sensor. In an article describing the location systems for ubiquitous computing, Hightower et al., (2001) have developed a taxonomy for mobile computing devices in order to identify opportunities for new location-sensing techniques. Several evaluation properties have been listed: precision, accuracy, scale, cost, and limitations.

Kjaergaard (2007) stated that it was not much help in specific question to radio location fingerprinting by proposing specific taxonomy for general properties of location fingerprinting systems which are: scale, output, measurements, and roles of accuracy, precision, complexity, robustness, scalability and cost. Furthermore, the list was improvised by Gu et al. (2009), in her article by introducing several evaluation criteria for assessing indoor positioning systems, namely security and privacy, cost, performance, robustness, complexity, user preferences, commercial availability, and limitations.

### C. Localization using Mathematical algorithms

The focus is for location determination in mobile Positioning technology that can be used on mobile navigation system. In this survey, we focus only on location determination by using standalone embedded GPS on Mobile phone in structured environment since it needs integration with other device or sensor in order to make it survive in unstructured environment.

A geometric map-matching algorithm makes use of the geometric information of the spatial road network data by considering only the shape of the antenna. In the geometric map-matching algorithm, the technique based on simple search algorithm is most commonly used. There are three types, which are

- point-to-point matching,
- point-to-curve matching and
- Curve-to-curve matching.

Point-to-point matching refers to the matches of each position fixes to the closest 'node' or 'shape point' of a road segment. Point-to-curve matching [12] and White et al, [13]) refers to the matches of the point on to the closest curve in the network. Meanwhile, curve-to-curve matching [14] and White et al, [15]) compares the vehicle's trajectory against known roads. A topological map-matching algorithm is an algorithm that using the links geometry such as: points, lines, and polygons as the links connectivity and contiguity.

A probabilistic algorithm is referring to the information definition of an elliptical or rectangular confidence region around a position fix which is obtained from a navigation

sensor which is GPS. The error region is superimposed on the road network to identify a road segment. If the error region contains more than one street, this algorithm will perform a weighted search on the candidate streets. Advanced map-matching algorithms are referred to as those algorithms that use more refined concepts such as a Kalman Filter or an Extended Kalman Filter by Kim et al, [16].

**Dumpster-Shafer's mathematical theory of evidence** by Yang et al, [17], a flexible state space model and a particle filter by Gustafsson et al, [18], an interacting multiple model by Cui et al, [19]), a fuzzy logic model by Kim et al, [19], the application of Bayesian inference by Pyo et al, [20] provide lot of information about the models needed to detect location accuracy in different formats. RADAR [21] is an in-building RF-based user location and tracking system uses the nearest neighbor in signal space (NNS) technique to predict the user's location. The system also uses signal propagation modeling approach to build the radio map, the goal was to reduce the system dependence on empirical data. The authors ignored the Floor Attenuation Factor (FAF) which was proposed earlier and adopted the Wall Attenuation Factor (WAF) instead. They discovered that there is an inverse relationship between the amount of additional attenuation and the number of walls separating the transmitter and the receiver. The accuracy of the system was about 2{3 m. Horus [22], a probabilistic WLAN location determination system which was designed with the goal of high accuracy and low computational cost.

In [23], the authors presented a hybrid indoor positioning method that uses ray-tracing model for modeling the multipath effects. In [24], the authors proposed an indoor location determination system that uses non line of sight (NLOS) scheme and one bound scattering paths.

Most indoor positioning system based on TOA or AOA metrics requires sophisticated devices to measure time or angle. Moreover, TOA location determination systems use the TOA measurement of the first path to determine the location which in turn is difficult to be calculated accurately in indoor environments [6]. All previous approaches involve identifying current location using GPS. This method of location identification is not popularly used, which can create problems in some locations as inside thick buildings where satellite signals do not reach.

## III. LOCALIZATION ACCURACY EVALUATION

In general, considering the third party external GPS receivers, smart phones with Bluetooth plays an important role for location data transfer. Dual 150S is being used in remote locations with good location accuracy. The location accuracy it provides is around 2.5 Meters.

The device can be made as a wrist watch, dash board device or in other forms to view the satellites. To get 2-3-meter location accuracy, external Bluetooth GPS receiver is essentially required. Depending upon the accuracy requirement mapping grade accuracy can be used. Table 1 indicate the parameter requirement for the GPS receivers.



**Table 1: Parameter for GPS Receiver requirement**

EVALUATION PARAMETERS						
Location	Signal to Noise Ratio [SNR]	Number of Satellites	Signal strength from satellite	Measurement Interval	Differential Correction	Drift in Accuracy
Any	Lower	More	Weak	Longer	Needed	More
Any	Higher	Less	Strong	Lesser	Not Needed	Less
<b>Open area -Landslide</b>	6.0	4 to 5 or greater Up to 14	Multi path Signal	10 seconds with 1 second sampling interval	1m for 100 km	better
<b>Indoor</b>	5	Not Possible	Direct Signal	Reduced sampling interval through means	1 cm accuracy	less
<b>Semi Indoor</b>	4-5	Possible	Direct and reflected Signal	6 seconds with 1 second sampling interval	50 cm accuracy	more
<b>LOS</b>	5-6	Min of 3 Possible or greater	Direct Signal	5seconds with 1 second sampling interval	10 cm accuracy	less
<b>NLOS</b>	3-4	No visibility	Reflected Signal	Reduced sampling interval through means	1-meter accuracy	more
<b>Multi Floor</b>	3	Not Possible	Reflected Signal	Reduced sampling interval through means	1- meter accuracy	better
<b>Room</b>	2-3	Not Possible	Direct Signal	5seconds with 1 second sampling interval	1-5 cm accuracy	better
<b>Room to room</b>	2-3	Not Possible	Reflected Signal	5seconds with 1 second sampling interval	10 cm accuracy	better
<b>Dense Area</b>	2-3	5-6	Reflected Signal	10 seconds with 1 second sampling interval	1-meter accuracy	more
<b>Forest</b>	3-4	At least 5-6	Direct plus Multipath Signal	10seconds with 1 second sampling interval	1-meter accuracy	more
<b>Sea</b>	5-6	5-6 or greater	Reflected Signal	5seconds with 1 second sampling interval	1meter accuracy	less
<b>Highly recommended</b>	6	4	Direct Signal	12 seconds with 1 second	cm accuracy	nil
<b>&lt;1.5 is not recommended</b>	1-2	1-2 or less	Multipath Signal	Not possible	nil	nil

**A. Erratic effects in locating systems**

Real-time locating is affected by a variety of errors. The major reasons are physical and may not be reduced by improving the technical equipment. The only escape is mathematical intelligence to improve.

**B. None or no direct response**

Many RTLS systems have a very mundane requirement: they require direct and clear wireless visibility. For those systems, where there is no visibility on the path from mobile tags to resident nodes there will be no result or a non-valid result from locating engine. This applies to satellite locating as well as other RTLS systems such as angle of arrival and time of arrival. If the locations in the tracking area contain distinct measurement fingerprints, line of sight is not necessarily needed.

**Incorrect location**

The measured location may appear entirely faulty. This is a generally result of simple operational models to compensate for the plurality of error sources. It proves impossible to serve proper location after ignoring the errors.

**C. Locating backlog:**

*Real time* is no registered branding and has no inherent quality. A variety of offers sails under this term. As motion causes location changes, inevitably the latency time to compute a new location may be dominant with regard to motion.

**D. Temporary Location Error**

Location will never be reported *exactly*, as the term *real-time* and the term *precision* directly contradict in aspects of measurement theory as well as the term *precision* and the term *cost* contradict in aspects of economy.

**E. Steady Location Errors**

Recognizing a reported location steadily apart from physical presence generally indicates the problem of insufficient over-determination and missing of visibility along at least one link from resident anchors to mobile transponders.

**F. Location Jitter**

Noise from various sources has an erratic influence on stability of results. The aim to provide a steady appearance increases the latency contradicting to real time requirements.

**G. Location Jump**

As objects containing mass have limitations to jump, such effects are mostly beyond physical reality. Jumps of reported location not visible with the object itself generally indicate improper modeling with the location engine.

**H. Location Creep**

Location of residing objects gets reported moving, as soon as the measures taken are biased by secondary path reflections with increasing weight over time.

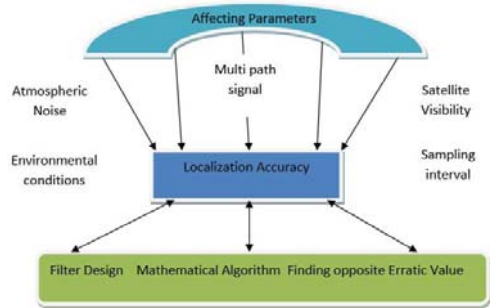


Figure 1: GPS Accuracy affects

Recent advances in communication technologies have a great impact on location determination systems. Location determination systems are deployed in almost every building, from hospitals where the location of patients and doctors or any medical equipment can be determined, or sending information to customers based on their location, to organize the trace and reducing congestion in the highways. To do so, the below mentioned approach can be of maximum use.

**Environment Model Building**

**Spatial Modelling:** Constructing the cartographic map of the buildings in the environment -[2Dor3D]

**Fixed Location Data from Anchor:** A GPS transceiver system or an access point mechanism can be structured for providing known location coordinate of the environment.

**Development of an Accurate Positioning Algorithm Using RSSI:** Constructing the ZigBee network to exchange data through RSSI. Using collaborative localization, the entire environment can be positioned by proper mathematical modeling of the network.

**Network for Dynamic Data Collection:** Constructing the ZigBee network by properly placing the modules as data collector nodes, router nodes and Coordinator nodes.

**Web Mapping Mechanism:** A GIS mapping mechanism with GPRS to send the data to Internet using web mapping services. It consists of developing augmented tools for emergency rescue operations along with the cartographic map of the Indoor/Semi Indoor.

**IV. MOBILE NAVIGATION ACCURACY IN PEDESTRIAN MOVEMENTS**

The GPS Resolution depends mostly upon the Hardware than the Software, But the Mobile Phone is so small. This Will Reflect On The GPS Accuracy, as 2 Meters Accuracy With Some Android Phones.

**V LOCALIZATION ACCURACY MEASURES AND APPLICATION OF KALMAN FILTER:**

The Kalman filter is a tool that can estimate the variables of a wide range of processes. In mathematical terms we would say that a Kalman filter estimates the states of a linear system. The

Kalman filter not only works well in practice, but it is theoretically attractive because it can be shown that of all possible filters, it is the one that minimizes the variance of the estimation error. Kalman filters are often implemented in embedded control systems because in order to control a process, you first need an accurate estimate of the process variables. The Kalman filter not only works well in practice, but it is theoretically attractive because it can be shown that of all possible filters, it is the one that minimizes the variance of the estimation error.

Motion Effects on GPS Receiver Time Accuracy: Motion Parameters

Direction and speed parameters are used by some systems in order to describe motion. Motion parameters consist of a way to describe the rate of change and the future value of the location of an object or person. Direction refers practically to persons because their motion, in contrary to the motion of objects, is not deterministic. This parameter is used as a “compass” and the information we want to extract from it is to determine to which objects and persons the user is aiming at, while moving in space. Direction is usually used with parameters  $x$  and  $y$  in order to have more specific conclusions.

The Kalman filter has two distinct phases: **Predict and update**. The predict phase uses the time delay estimate from the previous time-step to produce an estimate of the time delay at the current time-step from each satellite.

In the update phase, measurement information at the current time-step is used to refine this prediction to arrive at a new, presumably more accurate time delay estimate, again for the current time-step.

The “slower” the solution rate setting of the filter, the more integration (smoothing) of the data stream results. These three elements of the Kalman filter are important concepts when trying to understand what happens when the GPS receiver is moving.

The Kalman filter is a filter used to predict the current state  $k$  which can be derived from the previous state at  $(k - 1)$ .

Where,  $F_k$  is the state transition input which is applied to the previous state  $x_{k-1}$ ; and  $B_k$  is the control-input which is applied to the control vector  $u_k$ .

$w_k$  is the process noise which is a zero mean multivariate normal distribution with covariance  $Q_k$ . At time  $k$  an observation  $z_k$  of the true state  $x_k$  is made.

where  $H_k$  is the observation model which maps the true state space into the observed space and  $v_k$  is the observation noise which is assumed to be zero mean Gaussian white noise with covariance  $R_k$ .

The initial state, and the noise vectors at each step  $\{x_0, w_1, \dots, w_k, v_1 \dots v_k\}$  are all assumed to be mutually independent values.

## VI. CENTIMETER ACCURACY TO AVOID ACCIDENTS

1. Bluetooth & Wireless LAN interconnectivity architectures. Wireless networks of small range will be utilized in order to provide information in various bandwidth ranges and to ensure compatibility with existing terminal devices. The use of Bluetooth for local positioning will also be explored for heading positions/directions.

2. Layout mapping and navigation tracking technologies will be used in order to facilitate the navigation planning and routing services.

3. Indoor GPS will be used to support accurate indoor positioning and tracking. Indoor GPS will be used to support very accurate low-power, low-cost indoor positioning, which is otherwise impossible to achieve with complementary indoor positioning technologies.

### GPS for pedestrian movements

When the receiver or the antenna is actually not stationary, the measurements are constantly changing and the self-survey fix can no longer be relied upon to generate low phase error to the time mark (IPPS). In fact, the more quickly the receiver moves in space the greater the error would be. Acceleration is most critical. Each sample is assumed to be received at the surveyed location. When the time delay is estimated the satellite position will seem wrong compared to where receiver expected it to be. That is, the measurement values will differ greatly from the values predicted by the Kalman filter. In our receiver, the Kalman filter tracks rates of change in the various parameters, and attempts to get a convergent solution based on all measurements which haven't been excluded. Measurements excluded are those that exceed or do not track anticipated values. An estimated confidence is assigned to measurements at have remained locked for a period of time. As the number of measurements discarded grows, too few samples will remain to result in a valid solution, on at any point in time. Failure to achieve a GDOP is an “unlock” condition.

Cricket has a position estimation accuracy of 10 cm and an orientation accuracy of 3 degrees. Environmental effects are not just caused by obstacles. Greater reflections in the shielded room cause higher multipath effects. In this case the small size of the room means the primary multipath signals arrive with net constructive superimposed signals. In addition to multipath, fading and shadowing of the RF channel, signal strength measurements are also affected by the following factors: 1. **Transmitter variability**; 2. **Receiver variability**; 3. **Antenna orientation**;

## VII. CONCLUSION

In the next generation of mobile navigation system, people will need more alternative types of context information of the environments on the mobile phone, not just only limited to communication services. The usage of location context is becoming more popular nowadays. By utilizing and improving existing location determination techniques, this will enable location wares to be more intelligence in order to upgrade the quality of life. The taxonomy for location determination on

mobile navigation system which is crucial to the establishment of ubiquitous positioning on mobile phone. In this paper we have argued for the potential applicability of mobile positioning technologies and applications in indoor environments. Based on a review of the issue from a technological, as well as an implementation perspective, we have developed the hypothesis that context-aware applications can prove beneficial for all stakeholders in a number of application areas, such as supermarkets, museums, and exhibitions.

At the same time, our analysis has shown that the road to full-scale development and application of such services is full of challenges and still unanswered research questions. A first set of such challenges involves the technological aspects of indoor positioning. Another set of challenges involves the potential transformation of industry value chains (for example, in supermarkets) and inter-party relationships (for example, in exhibitions). The introduction of context-aware services may result in a significant shift of the relative bargaining power between stakeholders, and hence have a considerable impact on future business models and market structures. Although such changes are expected to favor the end customer (shopper, visitor, and so on) in all cases, further research will undoubtedly be required to explore the dynamics of context-aware service provision and investigate the likely impact on the viability of future business ventures. Acknowledging the above, we have embarked on an empirical investigation that will complement our theoretical insight to provide a robust and holistic approach to the research problems of indoor mobile positioning. Despite the relatively early stage of the empirical aspects of our research, the results so far suggest strong indications to support our hypotheses. In any case, we have shown that the theoretical investigation alone provides ample space on which to base the prediction that the future of m Business holds significant potential for indoor environment applications that will complement and integrate basic outdoor location service provision.

Future research and development will be vital to develop this technology. Tracking blind pedestrians, multiple objects, differentiating between friend, foe, or civilian vehicles, and reducing false detection, and estimating the value of a target based upon size are areas of prospective advances this type of research would be capable of.

## REFERENCES

1. Paul Hightower, "Motion Effects on GPS Receiver Time Accuracy" Instrumentation Technology Systems, 2008
2. Robin Wentao Ouyang, Albert Kai-Sun Wong and Kam Tim Woo, "GPS Localization Accuracy Improvement by Fusing Terrestrial TOA Measurements" IEEE Communications Society IEEE ICC 2010 proceedings
3. Welch & Bishop, "An Introduction to the Kalman Filter," UNC-Chapel Hill, TR 95-041, July 24, 2006
4. Wan Bejuri et al "Ubiquitous Positioning: A Taxonomy for Location Determination on Mobile Navigation System" Signal & Image Processing: An International Journal(SIPIJ) Vol.2, No.1, March 2011 DOI: 10.5121/sipij.2011.2103 24
5. B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, "Global Positioning System (GPS). Theory and practice," *Wien: Springer, 1992, 1992.*
6. J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing. Computer", vol. 34, no. 8, pp. 57–66, 2001.
7. Z. Sun, B. George, & M. Ronald, "On-Road Vehicle Detection: A Review", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no.5, pp. 694-711, 2006.
8. S. Blake, C. Hill, T. Moore, C. Hide, D. Park, "A Heave Compensation Algorithm Based on Low Cost GPS Receivers", *The Journal of Navigation*, vol.61, no.2, pp. 291-305. 2008. Signal & Image Processing: An International Journal(SIPIJ) Vol.2, No.1, March 2011
9. J. Aponte, X. Meng, C. Hill, T. Moore, M. Burbidge, A. Dodson, "Quality assessment of a network-based RTK GPS service in the UK". *Journal of Applied Geodesy*. vol. 3. no.1. pp.25- 34, 2009.
10. J.S. Greenfield, "Matching GPS observations to locations on a digital map". In *proceedings of the 81st Annual Meeting of the Transportation Research Board*, January, Washington D.C., 2002.
11. F. Gustafsson, et al, "Particle filters for positioning, navigation, and tracking", *IEEE Transactions on Signal Processing*, vol.50, no.2, pp.425–435, 2002.
12. Y. Cui & S.S. Ge, "Autonomous vehicle positioning with GPS in urban canyon environments", *IEEE Transactions on Robotics and Automation*, vol.19, no.1, pp.15–25, 2003.
13. C. D. Cliff Randell and H. L. Muller, "Personal position measurement using dead reckoning," in *Proceedings of the Seventh International Symposium on Wearable Computers*, Narayana swami, Ed. IEEE Computer Society, October 2003, pp. 166–173.
14. Q. Ladetto et al "Two different approaches for augmented GPS pedestrian navigation," in *International Symposium on Location Based Services for Cellular Users*, Locellus, 2001.
15. G. Abwerzger et al, "Demonstrating a GPS/EGNOS/Loran-C navigation system in difficult environments as part of the ESA project SHADE," in *EURAN 2004Conference*, Munich, Germany, June 2004.
16. J. A. Farrell and M. Barth, "*The Global Positioning System and Inertial Navigation*". New York: McGraw-Hill, 1999.
17. S. S. Ge et al, "Terrestrial navigation based on integrated GPS and INS," in *Proc.SPIE Conf. Enhanced and Synthetic Vision*, vol. 3364, Orlando, FL, Apr.13–17, 1998, pp. 348–358.
18. M. Tsakiri, A. Kealy, and M. Stewart, "Urban canyon vehicle navigation with integrated GPS/GLONASS/DR systems," *Navigation: J. Inst.Navigation*, vol. 46, no. 3, pp. 161–174, 1999.
19. J. D. Weiss et al, "GPS/INS integration in a severe urban environment," in *Proc. IEEE Position Location, Navigation Symp.*, 1998, pp. 432–440.
20. J. K. Ray, M. E. Cannon, and P. Fenton, "GPS code and carrier multipath mitigation using a multi antenna system," *IEEE Trans. Aerosp. Electron.Syst.*, vol. 37, pp. 183–195, Jan. 2001.
21. C. W. Jang, J. C. Juang, and F. C. Kung, "Adaptive fault detection in real-time GPS positioning," *Proc. Inst. Elect. Eng.—Radar, Sonar, Navigation*, vol. 147, no. 5, pp. 254–258, 2000.
22. K. Venkatraman, et al, "A hybrid method for improving GPS accuracy for land vehicle navigation system"... Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on PP74-79, IEEE.
23. K Nanmaran, B Amutha, "Situation Assisted Indoor Localization using Signals of Opportunity" International Conference on Indoor Positioning and Indoor Navigation, IPIN, Vol 7,2014.
24. K Nanmaran, W Mackaness, "An EGNOS Assisted GPS based Pedestrian Navigation System" Proc. Int. Conf. on Communication Engineering and Network Technologies, CEN ELSEVIER 2012.

# Modeling Service Interaction in M2M Connectivity Management

Evelina Pencheva

Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 e-mail: enp@tu-sofia.bg

Anastas Nikolov

Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 e-mail: nikolov.anastas@gmail.com

Ivaylo Atanasov

Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 e-mail: iia@tu-sofia.bg

Kamelia Nikolova

Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 e-mail: ksi@tu-sofia.bg

**Abstract**— The increased amount of connected devices and Machine-to-Machine (M2M) applications places business and technical challenges for network operators. The complexity of connectivity challenge yields for appropriate connectivity management solutions. The explosion of M2M services may result in undesired service interaction. Despite of the considerable progress in service interaction management, there is a lack of knowledge on the kind of interaction in real M2M communication systems. In this paper, we present formal definitions of different M2M connectivity management services and suggest a method for detecting service interaction using standard reasoning by description logic.

**Keywords**— *Device management; Connectivity management; Service interaction; Description logics; Satisfiability*

## I. INTRODUCTION

It is obvious and undisputable that the trend for increase of the number of smart devices equipped with sensors and actuators will be constant. Being smart, the devices need to connect in order to communicate with other things. Machine-to-Machine (M2M) communications are viewed as a frontier of future next generation networks. The ubiquitous penetration of M2M communications in different application area has led to more and more M2M applications. The increased amount of connected devices and M2M applications places business and technical challenges for network operators [1].

On one hand, device management includes functions like automated device configuration, over-the-air firmware updates, remote reboots, diagnostics and troubleshooting, security and integrity. Different protocols and proprietary solutions have fragmented the M2M market and have added complexity, time and cost to integration process [2], [3]. An abstraction required for scalable platform that adheres to standards and addresses a broad range of common M2M functions is provided by OMA Lightweight M2M [4]. On the

other hand, the explosion of M2M services and applications may result in service interaction. Service interaction, or feature interaction, manifests itself as a function of services which is neither exactly the sum of every service nor behaves as expected [5]. Instances of the service interaction problem have been studied in different M2M applications like home automation [6], automotive systems [7], service systems [8] and in other fields. The compositionality and modularity [9] are in the base of the problem instances, while the difference between the individual views, interpretations and eventual solutions, is considerable. An example for such significant difference might be given when comparing the views on service interactions of automotive systems engineering and of service systems in aspects like functionality, parallelism, structure etc.

Despite of the progress in developing approaches for modeling, detecting, and resolving service interactions, there is a lack of sufficient knowledge on the kind of service interactions that occur in real-world M2M systems [10]. In our previous works, we studied different aspects of service interaction in CAMEL networks [11]-[14]. Customized Applications for Mobile Enhanced Logic (CAMEL) is service delivery platform for GSM and UMTS networks. Our research focused on human call related behavior. In [11] and [12], we studied service interactions based on CAMEL originating and terminating basic call state models respectively and reasoned on interactions between services available for calling and called parties. In [13] and [14], we stressed on CAMEL mobility management models to study interaction between services as a result of subscriber's mobility. CAMEL models are not applicable into the world of M2M communications where devices use data transfer.

In this paper, we present a model of device connectivity and a formal method for specification of connectivity management (CM) services by means of description logic. Our

approach allows the formal definition of the basic CM model as well as definitions of services (Location-based CM and Quality of service-based CM) by means of refinement. Service interactions are considered as a contradiction problem and may be discovered automatically by standard reasoning algorithm on description logic.

The paper is structured as follows. In Section II, we briefly present the OMA Trap Framework which allows interoperable way for device management using any kind of events worthwhile for managing and monitoring the networked services or applications deployed on devices, or faults on the software and hardware, etc. Section III makes an introduction to description logics and then discusses its usage for CM model description. The approach to service definition and the algorithm for inference of service interaction are presented in Section IV and Section V. The conclusions summarize the contribution.

## II. OMA DIAGNOSIS AND MONITORING TRAPS

The OMA Lightweight M2M protocol is targeted at constrained devices with low power microcontrollers and limited amount of memory, as well as at more powerful embedded devices. It sets a protocol between a server located in a public or private data center, and a client which resides at the device. The interface between the LWM2M client and server allows device management. The focus in this paper is on device connectivity management using connectivity observation and bearer selection.

Devices may be connected using cellular bearers, wireless bearers or may use wireline ones. The server may observe line voltage and signal strength at the device side. For this purpose, the server establishes an observation relationship with the device in order to set the observation policy. The device sends periodical or triggered reports, containing the information until the server cancels the observational relationship. The server may query the device about multiple parameters, related to its connectivity, including about the network bearer in use, and about available network bearers. When, for example, the device is using cellular connectivity and supports WLAN connectivity, but WLAN coverage is available, then the server may instruct the device for making selection of bearer with the preferred WLAN among the other options.

OMA DiagMon Trap Events specification defines a number of standardized traps [15]. Geographic trap is active when a device is in a specific geographic area. Collection of measurement data (e.g. signal strength and quality) is important for the connectivity management when the device communicates over the wireless network. Received power trap can help the connectivity optimization process by triggering the server when the received power of the device drops below the server-specific value. Whenever the received power by the device drops below a server-specified value (TrapActivePower), it causes this trap to go active. Alternatively, when device senses power rises above another server-specified value (TrapInactivePower), it causes this trap to go inactive. In cases that the trap goes active or inactive, the device notifies the registered server. If the device exposes quality of service (QoS) metrics functionality, it may report to the server QoS related

events. Different QoS parameters can be measured, as far as the delays and jitter are critical for real-time applications like streaming, and bandwidth is important for applications such as video surveillance, transportation services, and industrial control. By means of these traps, in the next section we describe the connectivity model formally, using descriptive logic.

## III. DEVICE CONNECTIVITY MANAGEMENT MODEL

Description logic is a formal language used for knowledge representations and reasoning about it. The basic syntactic blocks, used to represent the knowledge base, are atomic concepts, atomic roles and constants. The basic components of the knowledge base are Terminology box (TBox), which introduces terminology in the application domain and the Aseerption box (ABox), which contains assertions about constants. Typical reasoning on knowledge base is to determine whether a description is non-contradictory or whether one description subsumes another description.

Our approach to definition of atomic concepts is to represent the device states and bearer related facts in the CM model as concepts. Fig.1 shows the device connectivity model which is based on OMA Received power trap.

Let us assume, that there is a finite set of bearer indices which represent the possible bearers that may be used by a particular device. The following concepts are defined:

*disconnected*, device is disconnected;  
*connected<sub>b</sub>*, device is connected using bearer *b*;  
*marginal<sub>b</sub>*, device's received power is below a server-specified value;  
*changeRequired*, device needs to change the used bearer;  
*inArea*, device is in a specified area;  
*preferred<sub>b</sub>*, bearer *b* is preferred bearer in the area;  
*qosAcceptable<sub>b</sub>*, quality of service of bearer *b* is acceptable;  
*qosUnacceptable<sub>b</sub>*, quality of service of bearer *b* is not acceptable;  
*available<sub>b</sub>*, bearer *b* is available;  
*unavailable*, there are no available bearers.

The transitions that change the device state are defined as atomic roles:

*signalDrop*, received power of used bearer drops below server-specified value;  
*signalRise*, received power of used bearer rises above server-specified value;  
*enter*, device enters the specified area;  
*exit*, device exits the specified area;  
*qosDecrease*, QoS of used bearer becomes unacceptable;  
*timerExpiry*, time guarded hysteresis of the received power is over;  
*connect*, device connects to the network;  
*disconnect*, device disconnects from the network.

Our terminology box contains expressions showing the changes in CM model and statements specifying the relationship between the events that cause transitions.

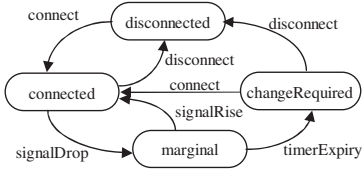


Fig. 1 Basic device connectivity model

$$\text{disconnected} \sqcap \text{available}_b \sqsubseteq \exists \text{connect. connected}_b \quad (1)$$

$$\text{connected}_b \sqsubseteq \exists \text{signalDrop. marginal}_b \quad (2)$$

$$\text{marginal}_b \sqsubseteq \exists \text{signalRise. connected}_b \quad (3)$$

$$\text{marginal}_b \sqsubseteq \exists \text{timerExpiry. changeRequired} \quad (4)$$

$$\text{changeRequired} \sqcap \text{available}_c \sqsubseteq \exists \text{connect. connected}_b \quad (5)$$

$$\text{changeBearer} \sqcap \text{unavailable} \sqsubseteq \exists \text{disconnect. disconnected} \quad (6)$$

$$\text{connected}_b \sqsubseteq \exists \text{disconnect. disconnected} \quad (7)$$

We need expressions that describe the device mobility, based on OMA Geo trap:

$$\neg \text{inArea} \sqsubseteq \exists \text{enter. inArea} \sqcap \text{preferred}_b \quad (8)$$

$$\text{inArea} \sqcap \text{preferred}_b \sqsubseteq \exists \text{exit. } \neg \text{inArea} \quad (9)$$

Let us denote by DEV the set of all devices. By CMS we denote the states  $s_i$  in the CM model. The assertion box contains one statement presenting the initial state for each device:

$$s_0: \sqcap_{d \in \text{DEV}} (\text{disconnected} \sqcap \text{available}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcup \text{disconnected} \sqcap \neg \text{available}_b \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcup \text{disconnected} \sqcap \neg \text{inArea} \sqcap \text{available}_c \sqcup \text{disconnected} \sqcap \text{unavailable}).$$

To express the fact that each device is in exactly one state at any moment we use the statement:

$$\top \sqsubseteq \neg (\sqcup_{d1, d2 \in \text{CMS}. d1 \neq d2} (s1 \sqcap s2)) \sqcap (\sqcup_{d \in \text{CMS}} s)$$

The device state changes by means of actions defined as action functions. An action function  $\text{Func}_{\text{CMS}}$  for given state corresponds to the possible transitions in the CM model. For example, the expression  $\text{Func}_{\text{CMS}}(\text{connected}_b) = \{\text{signalDrop}\} \cup \{\text{disconnect}\} \cup \{\text{enter}\} \cup \{\text{exit}\}$  means that if the device is connected, the signal strength of the used bearer may drop, or the device may disconnect, enter or exit the area.

The fact that each device can change the CM state only by means of certain actions is represented by the following statement: for all  $s \in \text{CMS}$ , and all  $R \notin \text{Func}_{\text{CMS}}(s)$ ,  $s \sqsubseteq \forall R.s$ .

#### IV. SERVICE MODELS

Services are modeled as transformations on the knowledge base using contexts  $C[\phi]$  as subformula  $\phi$  of any formula  $\psi$ .

##### A. Location-based Bearer Selection

The Location-based Bearer Selection (LBS) service assumes that there is a predefined geographic area in which a preferred bearer is used. The refinement for LBS service is defined by the following statements:

$$C_1[\text{LBS} \sqcap \text{disconnected} \sqcap \text{available}_b \sqcap \text{inArea} \sqcap \text{preferred}_b] \sqsubseteq \exists \text{connect. } C_2[\text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b] \quad (10)$$

$$C_3[\neg \text{LBS} \sqcap \text{disconnected} \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_b] \sqsubseteq \exists \text{connect. } C_4[\text{connected}_c \sqcap \text{inArea} \sqcap \text{preferred}_b] \quad (11)$$

$$C_5[\text{LBS} \sqcap \text{disconnected} \sqcap \neg \text{available}_b \sqcap \text{inArea} \sqcap \text{preferred}_b] \sqsubseteq C_6[\text{disconnected}] \quad (12)$$

$$C_7[\text{LBS} \sqcap \text{connected}_b \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_c] \sqsubseteq \exists \text{connect. } C_8[\text{connected}_c] \quad (13)$$

$$C_9[\neg \text{LBS} \sqcap \text{connected}_b \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_c] \sqsubseteq C_{10}[\text{connected}_c] \quad (14)$$

$$C_{11}[\text{LBS} \sqcap \text{connected}_b \sqcap \neg \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_c] \sqsubseteq \exists \text{disconnect. } C_{12}[\text{disconnected}] \quad (15)$$

$$\text{LBS} \sqsubseteq \neg (\text{connected}_c \sqcap \text{inArea} \sqcap \text{available}_b \sqcap \text{preferred}_b) \quad (16)$$

##### B. Quality of Service-based Bearer Selection

The Quality of Service-based Bearer Selection (QBS) service requires bearer change if the QoS available on the used bearer decreases under predefined value. The refinement for QBS service is defined by the following statements:

$$C_1[\neg \text{QBS} \sqcap \text{connected}_b] \sqsubseteq \exists \text{qosDecrease}_b. C_2[\text{connected}_b] \quad (17)$$

$$C_3[\text{QBS} \sqcap \text{connected}_b] \sqsubseteq \exists \text{qosDecrease}_b. C_4[\text{connected}_b \sqcap \text{qosUnacceptable}_b] \quad (18)$$

$$C_5[\text{QBS} \sqcap \text{connected}_b \sqcap \text{qosUnacceptable}_b \sqcap \text{available}_c] \sqsubseteq \exists \text{connect. } C_6[\text{connected}_c] \quad (19)$$

$$C_7[\text{QBS} \sqcap \text{connected}_b \sqcap \text{qosUnacceptable}_b \sqcap \text{unavailable}] \sqsubseteq \exists \text{disconnect. } C_8[\text{disconnected}] \quad (20)$$

$$\text{QBS} \sqsubseteq \neg (\text{connected}_c \sqcap \text{qosUnacceptable}_c) \quad (21)$$

Possible service interaction may occur when the device is in the specified area and the QoS available on preferred bearer in use decreases.

##### C. Other Services Related to Connectivity Management

In addition to diagnostic and monitoring traps, described in Section III, OMA defines also Call drop trap and Data speed

trap. Whenever a call drop occurs in the predefined period, it causes Call drop trap to go active. For M2M communications this trap is related to data session drops. Whenever an uplink or downlink average data speed reaches the lower/higher threshold value, it causes Data speed trap to go active/inactive.

There are also events, related to Policy and Charging Control (PCC), that may trigger bearer change [16]. Such events include e.g. out of credit (credit is no longer available), usage report, enforcement of Application Detection Control rule, etc.

## V. REASONING ON SERVICE INTERACTION

When introducing new services, it is important to find out whether a new service is contradictory to existing concepts i.e. whether it is satisfiable or unsatisfiable with respects to the axioms in TBox, representing the CM model.

We use a tableau method defined in [17].

The tableau  $t \triangleq \{ \langle b | p : C \rangle \}$  is a set of prefixed formulae where the prefix of given formula is consisted of a binary string  $b := \varepsilon | (110)^*$  and a string of alternating names  $p := n(Rm)^+$ , and  $C$  is concept. Here  $\varepsilon$  is the empty string,  $n$  and  $m$  are names of individuals,  $R$  stands for the names of roles, and  $()^+$  denotes one or more occurrences.

Tableau Method	
AND:	$\frac{\langle b   p : C \cap D \rangle}{\langle b   p : C \rangle \quad \langle b   p : D \rangle}$
OR:	$\frac{\langle b   p : C \cup D \rangle}{\langle b_M 0   p : C \rangle \quad \langle b_M 1   p : D \rangle}$ $b_M$ maximal for $b$
SOME:	$\frac{\langle b   p : \exists RC \rangle}{\langle b   pRn : C \rangle}$ $pRn$ new (unless $pR$ exists in the branch)
KB:	$\frac{\vdots}{\langle b   p : \neg C \cup D \rangle}$ $p$ present in $b$ and $C \sqsubseteq D \in T$

### A. Detection of Interaction between LBS and QBS

The tableau algorithm for detecting interactions between LBS and QBS services proceeds as follows:

Applying AND to the start formula produces four cases:

$$\langle \varepsilon | s_0 : \sqcap_{d \in \text{DEVICES}} \rangle$$

$$(\text{disconnected} \sqcap \text{available}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcup$$

$$\text{disconnected} \sqcap \neg \text{available}_b \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcup$$

$$\text{disconnected} \sqcap \neg \text{inArea} \sqcap \text{available}_c \sqcup$$

$$\text{disconnected} \sqcap \text{unavailable})$$

1. In case of  $\text{disconnected} \sqcap \text{available}_b \sqcap \text{inArea}_a \sqcap \text{preferred}_b$

1.1 Applying KB to rule (10) produces

$$\langle \varepsilon | s_0 : \neg \text{disconnected} \sqcup \neg \text{available}_b \sqcup \neg \text{inArea} \sqcup \neg \text{preferred}_b$$

$$\sqcup \exists \text{available}_b, (\text{connected}_b \sqcap \text{qosAcceptable}_b) \rangle$$

1.2. Applying OR gives two branches:

1.2.1  $\langle 0 | s_0 : \neg \text{disconnected} \rangle$  which is closed because of the appearance of  $\langle 0 | s_0 : \text{disconnected} \rangle$  in this segment earlier.

1.2.2  $\langle 0 | s_0 : \neg \text{available}_b \rangle$  (closed).

1.2.3  $\langle 0 | s_0 : \neg \text{inArea} \rangle$  (closed).

1. 2.4  $\langle 0 | s_0 : \neg \text{preferred}_b \rangle$  (closed).

1.2.5  $\langle 1 | s_0 : \exists \text{connect}_b, (\text{connected}_b \sqcap \text{inArea})$

1.3 Applying SOME gives  $\langle 1 | s_0 \text{connect } s_1 :$

$$(\text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b) \rangle$$

1.4 We derive rule (18) and applying KB it produces

$$1.4.1 \langle 1 | s_0 \text{connect } s_1 : \neg (\text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b) \sqcup$$

$$\exists \text{qosDecrease}_b, (\text{connected}_b \sqcap \text{qosUnacceptable}_b \sqcap \text{preferred}_b)$$

$$\rangle \text{ and after OR}$$

1.4.2  $\langle 10 | s_0 \text{connect } s_1 :$

$$\neg (\text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b) \rangle \text{ (closed)}$$

1. 4.3  $\langle 11 | s_0 \text{connect } s_1 : \exists \text{qosDecrease}_b, (\text{connected}_b \sqcap \text{qosUnacceptable}_b \sqcap \text{inArea} \sqcap \text{preferred}_b) \rangle$  and applying SOME

$$\langle 11 | s_0 \text{connect } s_1 \text{qosDecrease}_b s_2 :$$

$$\text{connected}_b \sqcap \text{qosUnacceptable}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \rangle \rangle$$

1. 5. Next derivation is the rule (19) for which we apply KB and the result is

$$1. 5.1 \langle 110 | s_0 \text{connect } s_1 \text{qosDecrease}_b s_2 : \neg \text{connected}_b \sqcap$$

$$\text{qosUnacceptable}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \rangle \text{ (closed)}$$

1. 5.2  $\langle 111 | s_0 \text{connect } s_1 \text{qosDecrease}_b s_2 :$

$$\exists \text{connect}. \text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \rangle \text{ and after applying}$$

SOME it produces  $\langle 111 | s_0 \text{connect } s_1 \text{qosDecrease}_b s_2$

$$\text{connect } s_3 : \text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \rangle$$

which contradicts to LBS  $\sqsubseteq \neg (\text{connected}_c \sqcap \text{inArea} \sqcap \text{preferred}_b)$ .

2. In case of

$$\text{disconnected} \sqcap \neg \text{available}_b \sqcap \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_b$$

the device remains disconnected as to rule (12).

3. In case of  $\text{disconnected} \sqcap \neg \text{inArea} \sqcap \text{available}_c$

3.1 Applying KB to the rule (1) and eliminating the closed cases gives

$$3.1.1 \langle 1 | s_0 : \exists \text{connect}_c, (\text{connected}_c \sqcap \neg \text{inArea}) \rangle$$

to which applying SOME results in  $\langle 1 | s_0 \text{connect } s_1 : \text{connected}_c$

$$\neg \text{inArea} \rangle$$

3.2 We derive rule (8) and applying consecutively KB, OR

and SOME gives two branches

$$3.2.1 \langle 1 | s_0 \text{connect } s_1 \text{enter } \text{connected}_c \sqcap \text{inArea} \sqcap$$

$$\text{preferred}_b \sqcap \text{available}_b \rangle$$

for which we derive rule (13) and again applying KB, OR and SOME produces



$\langle 1 \mid s_0 \text{ connect } s_1 \text{ enters } s_2 \text{ connect } s_3: \text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcap \text{available}_b \rangle$

3.2.2  $\langle 1 \mid s_0 \text{ connect } s_1 \text{ enter } \text{connected}_c \sqcap \text{inArea} \sqcap \text{preferred}_c \sqcap \neg \text{available}_c \rangle$  for which we derive

$\text{connected}_b \sqcap \neg \text{available}_c \sqcap \text{inArea} \sqcap \text{preferred}_c \sqsubseteq \exists \text{disconnect}$ .  
 $\text{disconnect}$  and applying KB, OR and SOME produces  
 $\langle 1 \mid s_0 \text{ connect } s_1 \text{ enters } s_2 \text{ disconnect } s_3: \text{disconnected} \rangle$

3.3 To  $\langle 1 \mid s_0 \text{ connect } s_1 \text{ enter } s_2 \text{ connect } s_3: \text{connected}_b \sqcap \text{inArea} \sqcap \text{preferred}_b \sqcap \text{available}_b \rangle$  we apply the similar steps as those in (1.1.5) and it gives  $\langle 1 \mid s_0 \text{ connect } s_1 \text{ enter } s_2 \text{ connect } s_3 \text{ qosDecrease}_b \text{ s}_4 \text{ connect } s_5: \text{connected}_c \sqcap \text{inArea} \sqcap \text{preferred}_b \rangle$  which contradicts to LBS.

4. In case of  $\text{disconnected} \sqcap \text{unavailable}$ , the device remains disconnected.

The result is closed tableau which means that  $\delta_{QBS}(\delta_{LBS}(\text{CMS}))$  interacts on activation  $\{QBS\} \cup \{LBS\}$ .

It is important to mention that the service interaction can be detected automatically since the programmability of the algorithm.

## VI. CONCLUSIONS

Interactions are normal and typical when applying the compositional approach to M2M services. However, some interactions turn out to be unexpected or even undesired and this makes M2M service interaction modeling a tool for detecting such issues in advance.

We propose a method for formal description of services for M2M connectivity management and an approach to service interaction detection. Once detected at the specification phase, service interactions may be avoided by applying policies and rules.

The presented results outline a possible solution and the approach seems to be promising as far as the scalability is achievable because of algorithm's programmability.

## REFERENCES

- [1] L. Latvakoski, A. Ilvari, P. Vitic, B. Jueh, M.B. Alaya, T. Monteil, Y. Lopez, G. Talavera, J. Gonzalez, N. Granqvist, M. Kellil, H. Ganem, T. Vaisanen, "A Survey on M2M Service Networks", *Computers*, 2014, vo.3, pp.130-173, doi: 10.3390/computers3040130
- [2] D. Schulz, R. Gitzel, "Seamless maintenance - Integration of FDI Device Management & CMMS," *IEEE Int. Conf. on Emerging Technologies & Factory Automation (ETFA)*, 2013, pp.402-407.
- [3] C.S. Shih, C. T. Chou, K. J. Lin, B. L. Tsai, C. H Lee, D. Cheng, C. J. Chou "Out-of-Box Device Management for Large Scale Cyber-Physical Systems," *IEEE Int. Conferences iThings, GreenCom, and CPSCom*, 2014, pp.402 - 407.
- [4] S. K. Datta, C. Bonnet, "A lightweight framework for efficient M2M device management in oneM2M architecture," *Int. Conf. on Recent Advances in Internet of Things (RioT)*, 2015, pp.1-6.
- [5] C. Pereira, A. Aguiar, "Towards Efficient Mobile M2M Communications: Survey and Open Challenges," *Sensors* 2014, no. 14, pp.19582-19608; doi: 10.3390/s14101958.
- [6] C. Maternaghan, K. Turner, "Policy Conflicts in Home Automation," *Computer Networks*, vol.57 issue 12, pp.2429-2241.
- [7] A.L.Dominguez. "Detection of Feature Interactions in Automotive Active Safety Features," PhD thesis, School of Computer Science, University of Waterloo, 2012.
- [8] Y. B. Lin *et al.*, "EasyConnect: A Management System for IoT Devices and Its Applications for Interactive Design and Art," in *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 551-561, 2015.
- [9] P. Zave. "Modularity in Distributed Feature Composition. In *Software Requirements and Design: The Work of Michael Jackson*," pages 267-290. Good Friends Publishing, 2010
- [10] S. Apel, C. Kastner, B. Garvin. "Exploring Feature Interactions in the wild: the new feature interaction challenge," *FOSD, ACM*, 2013, pp.1-8.
- [11] Pencheva, E., I. Atanasov. "Detection of CAMEL Feature Interaction", *International Journal on Information Technology and Security*, 2010, (1), pp. 25-42.
- [12] Atanasov, I., E. Pencheva. "CAMEL Service Interaction Detection", *International Journal on Information Technologies and Control*, 2010, (4), pp.2-9.
- [13] Atanasov, I., E. Pencheva. "A Formal Approach to Service Interaction Detection in Mobile Networks". *Proc. of 10th WSEAS Int. Conf. on Software Engineering, Parallel and Distributed Systems (SEPADS '11)* Cambridge, UK, 2011, pp.118-123
- [14] Atanasov, I., E. Pencheva. "Reasoning on Service Interaction in Mobile networks", *International Journal of Computers and Communications*, 2011, vol.5 (2), pp.59-66.
- [15] Open Mobile Alliance, "Diagnostics and Monitoring Trap Events Specifications", 2013, OMA-TS-DiagonTrapEvents-V1\_2-20131008-A
- [16] 3GPP TS 23.203 Policy and Charging Control Architecture, Release 13, v13.7.0, 2016.
- [17] C. Areces, W. Bouma, M. de Rijke, "Feature Interaction as a Satisfiability Problem", <http://www.loria.fr/~areces/content/papers/files/fi-mascots.pdf>

# Research and Experiment of the Indoor Position System Based on Smartphones

Yingxin Lou

Beijing University of Posts and Telecommunications  
Beijing, China  
316594230@qq.com

Zhimiao Wang

Beijing University of Posts and Telecommunications  
Beijing, China  
3248699172@qq.com

**Abstract**—In recent years, indoor positioning techniques have been researched to support automatic guidance for visitors in public buildings such as museums, galleries, etc. But many indoor positioning techniques cannot distinguish human's moving patterns and external infrastructures are needed to improve accuracy. Therefore, a smartphone-based indoor positioning system with Improved Pedestrian Dead Reckoning (IPDR) and Magnetic Calibration is presented. The IPDR algorithm estimates walking distance and direction in real time and reports relevant data to magnetic calibration component. Based on the updated information, the final location of the user could be calculated utilizing the relevant data, the real-time data of magnetic field sensor and the two-dimensional magnetic map acquired in advance. We evaluate this method using experimental measurements in practice.

**Keywords**—*Indoor Positioning; Improved Pedestrian Dead Reckoning(IPDR); Pocket Pattern; Magnetic Calibration;*

## I. INTRODUCTION

Nowadays, many smartphone-based positioning algorithms have been proposed to provide indoor Location-Based Services (LBS) [1]. The first reason is that the accuracy of Global Positioning System (GPS) become worse indoors due to the satellite signal attenuation. As we all know, GPS plays the dominant role in outdoor LBS and it provides the most accurate outdoor-positioning service for smartphones at present. However, GPS has inherent limitation that its accuracy rapidly declines in indoor environment or urban canyon areas. According to surveys of people's habits of calling, there are only 10-20% calls from outdoors, while more than 70% calls originate from indoors which presents great potentials of indoor LBS. So a new accurate indoor positioning algorithm is needed for people.

A well-known Smartphone-Based Pedestrian Dead Reckoning (PDR) has been proposed by Schindhelm [2] to make up for the shortage of GPS. Due to the rapid development of smartphones, a variety of sensors such as accelerometer, magnetometer, gyroscope, proximity sensor, barometer, temperature sensors and Wi-Fi module are built in smartphone. Thus, PDR algorithm applies the data from these sensors to detect a certain kind of pattern which we call stop-and-go pattern and finally calculates the position. So PDR algorithm

has been widely applied in retail, health-care, museum-visiting, etc. The step number, stride length and direction could be estimated from the relevant sensor data.

In this paper, we study the navigation system in detail. First, we use the Improved Pedestrian Dead Reckoning to calculate the route when people begin to walk. Second, we will collect the guide magnetic map and store it before the people start. Third, when the people arrive the calibration points, we will reposition the walking route and display corresponding information. Through several experiments, our system is validated to be effective and practically providing accurate user position in indoor environments.

The rest of the paper is organized as follows. In Section 2, we provide an overview of our proposed system. Section 3 proposes a novel inertial positioning method based on the Improved Pedestrian Dead Reckoning algorithm. Section 4 describes an efficient method of magnetic calibration. In section 5, the experiment is implemented in a corridor environment and the results are analyzed. Section 6 is the conclusion and future work.

## II. SYSTEM ARCHITECTURE

In this section, we will describe our proposed system architecture shown in Fig.1. We proposes a method of indoor positioning which is combined of improved accuracy pedestrian dead reckoning based on inertial sensor and track calibration based on geomagnetic sensor. Pedestrian dead reckoning algorithm is consisted of three parts including gait detection based on dynamic threshold, step estimation based on linear regression and direction detection based on pocket pattern [3]. Meanwhile, we calibrate the position which is obtained from pedestrian dead reckoning algorithm using the method of geomagnetic matching to eliminate the accumulated error for making more precise positioning. Once the state of pedestrian changing from motion to static, improved accuracy pedestrian dead reckoning module will send real time position to geomagnetic calibration. Then we will match the geomagnetic map stored offline. At last, we give the calibrated real-time positioning back to improved accuracy pedestrian dead reckoning module.

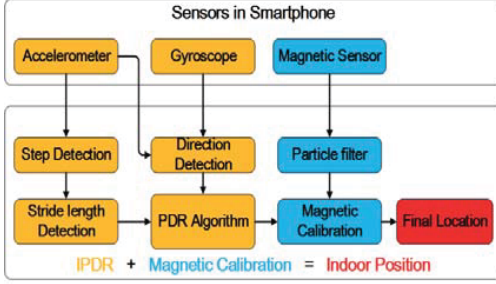


Fig. 1. System architecture

We can come to a conclusion that this paper designed system and method are correct by collecting data in the actual field and it is simulated on the MATLAB platform. This system can achieve a more stable and higher accuracy indoor positioning.

### III. IMPROVED PEDESTRIAN DEAD RECKONING

In this section, we will introduce the improved accuracy pedestrian dead reckoning algorithm, including gait detection based on dynamic threshold, step estimation based on linear regression and direction detection based on pocket pattern. Compared to traditional PDR, our IPDR has better accuracy and stability.

#### A. Gait Detection

Traditional gait detection in PDR only adopts time window detection algorithm. However, we will add three algorithms to make better gait detection. It includes mean filter, step start point detection and dynamic threshold updating.

During the process of collecting data, we attempt to flat the microphone thus its direction is identical to acceleration sensor. So it can better represent people's walking gait. According to acceleration sinusoidal change regulation, a sinusoidal time is a step. The synthetic acceleration magnitude  $A$  at time  $t$  is  $A = \sqrt{a_x^2 + a_y^2 + a_z^2}$  [4]. However, there are some questions. First, although people are in static station, but the microphone still has some tiny vibration so that vertical acceleration is still changing. Second, when people are in stepping station, it also has tiny vibration influencing the vertical acceleration. Third, acceleration sensor's sampling frequency is 50 Hz [5] and the data will interrupt if people walk too fast. The last one, we will get false data if the microphone's coordinate axes do not strictly consistent to the people' because of shaking factor. In view of the above problems, we propose mean filter, step start point detection and dynamic threshold updating in addition to time window detection.

Mean filter is used to remove noise. In this paper, due to people holding and stride factors, there are some vertical acceleration peak values in a cycle influencing gait detection step judgment. Therefore, we use mean filter whose window factor is  $M$  and window size is  $2M+1$ . Then, we will slide the

window along acceleration signal to get filtered acceleration. The mean filter is summarized in the following equation.

$$A_i = \left( \sum_{k=i-M}^{i+M} a_k \right) / (2M + 1) \quad (1)$$

where  $A_i$  is the acceleration value after smoothing filter,  $a_k$  is the vertical acceleration value.

Step start point detection aims at judging when the people begin to walk. When the person is at rest, due to the slight vibration, the vertical direction acceleration value will be slightly shocked [6]. In that case, it can be seen in static station. Therefore, according to the acceleration of pedestrian walking change, we will observe the vibration amplitude and regard the acceleration value between the  $9.5 \sim 10.1 \text{ m/s}^2$  [7] a slight vibration, that is, the acceleration of the accuracy is  $0.3 \text{ m/s}^2$ . If the acceleration value exceeds the range of  $9.5 \sim 10.1 \text{ m/s}^2$ , the amplitude of floating is large enough so that to be thought a starting point. As seen in Fig. 2.

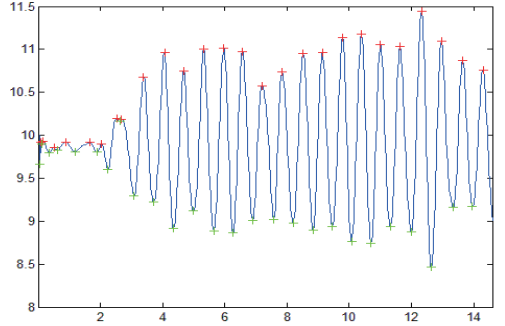


Fig. 2. Step start point detection

Dynamic threshold updating aims at detecting step start point, as seen in Fig.3. The threshold value is adjusted dynamically according to the state continuity and similarity of pedestrians. In order to get more accurate judgment stepped starting point and the effective number of steps, we need to deduce the dynamic threshold according to the value of a step wave crest and trough. The dynamic threshold updating is summarized in the following equation [8].

$$T_n = \alpha T_{thr} + \beta \frac{MAX_i + MIN_i}{2} + \gamma \quad (2)$$

where  $T_n$  is the dynamic threshold of current cycle,  $T_{thr}$  is the dynamic threshold of the last period,  $MAX_i$ ,  $MIN_i$  are the maximum value and minimum value of the acceleration in the current step cycle,  $\alpha$ ,  $\beta$ ,  $\gamma$  are the pre-trained parameters. Initial dynamic threshold  $T_{thr}$  is gravity acceleration  $g$ .

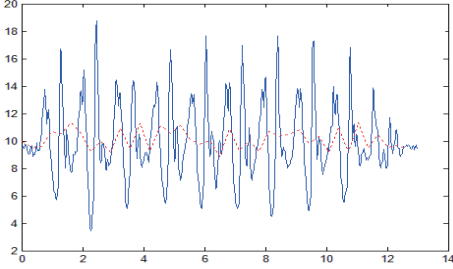


Fig. 3. Dynamic threshold

Time window detection algorithm is the most important foundation in gait detection. The normal pedestrian walking frequency is 0.5~5Hz, which means time is 0.1~1s. By detecting the peaks and troughs, the time interval between the peaks and troughs is half a cycle. Therefore, if interval between the peaks and valleys is between the 0.1~1s, we then think it an effective step. If the time interval is less than 0.1s or more than 1s, we cancel the step count.

### B. Step Estimation

Traditional gait detection in PDR adopts fixed step detection model. In our paper, we will change it to linear regression model.

According to their height, weight, walking state, pavement condition and other factors, each step has a strong randomness. If we only use the fixed step size 0.7m, then after a long period of time, it will have a lot of errors. The linear step model is a linear estimation model, which is established among the stride length, stride frequency and acceleration variance. When pedestrians stride frequency in the range of a standard step, there will be a near linear relationship with frequency and acceleration variance. So we use adaptive step size algorithm, the adaptive step size algorithm is summarized in the following equation.

$$l_i = a * f_i + b * v_i + c \quad (3)$$

where  $l_i$  is the every step length,  $a$ ,  $b$ ,  $c$  are pre-trained parameter,  $f_i$  is the step frequency,  $v_i$  is acceleration variance, as in the following.

$$f_i = \frac{1}{t_i - t_{i-1}} \quad (4)$$

$$v_i = \frac{1}{N_s} \sum_{t=t_{i-1}}^{t_i} (a_t - \bar{a}_i)^2 \quad (5)$$

where  $t_i$  is the step starting time,  $N_s$  is the number of acceleration sampling points,  $a_t$  is the amplitude of acceleration sampling point,  $\bar{a}_i$  is the average amplitude of acceleration sampling point.

### C. Direction Detection

Traditional direction detection in PDR only has the microphone flat model. However in our paper, we will provide another pocket pattern, which is used more commonly in real condition.

As noted earlier, we assume that microphone is in flat model so that the coordinate system between acceleration sensor and pedestrian is identical. So the direction which is the rotation angle of Z axis is calculated  $\Delta\theta_z = \omega\Delta t$  [9]. However, in the actual process of pedestrian walking, it is not necessary that people strictly keep the phone flat and may tilt or put in the pocket. Thus, the direction coordinate system is changed and should be redefined.

First, we have the integral of x, y, z axis and calculate the angular displacement  $\theta_x$ ,  $\theta_y$ ,  $\theta_z$ , as in the following[10].

$$\theta_x = \int_{t_{begin}}^{t_{stop}} \omega_x dt \quad \theta_y = \int_{t_{begin}}^{t_{stop}} \omega_y dt \quad \theta_z = \int_{t_{begin}}^{t_{stop}} \omega_z dt \quad (6)$$

Then we can calculate  $\theta_x$ ,  $\theta_y$ ,  $\theta_z$  proportion of z axis component in the pedestrian coordinate system which is similar to acceleration component. Taking  $\theta_x$  for example, its rotation angle component on z axis is in the following formula.

$$\theta_z = \frac{\overline{a_x} \theta_x}{\sqrt{\overline{a_x}^2 + \overline{a_y}^2 + \overline{a_z}^2}} \quad (7)$$

where  $\overline{a_x}$  is the x axis acceleration average projected to z axis, so as to  $\overline{a_y}$ ,  $\overline{a_z}$ ,  $\theta_z$  is the x axis rotation angle projected to z axis.

In the last, we project all the x, y, z axis rotation angle to z axis in pedestrian coordinate system. The direction formula is summarized in the following equation. As shown in Fig.4.

$$\theta_z = \frac{\overline{a_x} \theta_x + \overline{a_y} \theta_y + \overline{a_z} \theta_z}{\sqrt{\overline{a_x}^2 + \overline{a_y}^2 + \overline{a_z}^2}} \quad (8)$$

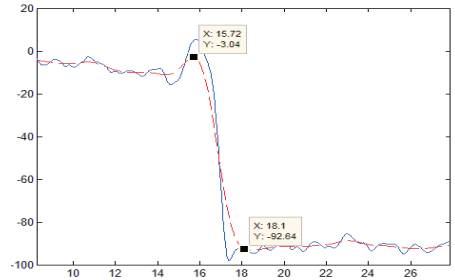


Fig. 4. Angular displacement around Z-axis based on pocket pattern

#### IV. MAGNETIC CALIBRATION

After the IPDR algorithm, in this section, we will propose a magnetic calibration algorithm based on magnetic map matching and particle filter, which will correct the pedestrian track more precisely.

##### A. Magnetic Map

In modern architecture, the steel and iron in the reinforced concrete structure will disturb the current earth's magnetic field in the local area and the phenomenon of abnormal magnetic field changes with the change of its position. Use the geomagnetic characteristic information as features to build specific indoor environmental magnetic field. Then we compare measurement of geomagnetic characteristics correction with the pre-stored magnetic field correction of data matching to determine the position and correct the track of calculation result.

This experiment is on the first floor of No.3 building of Beijing University of Posts and telecommunications. Because of the building's reinforced concrete structure, it has unique magnetic field characteristics in some feature points. The feature points can be set for the magnetic calibration points. In this paper, the Teslameter11th software is used to measure the magnetic induction intensity and the resolution is 0.1 uT, the sampling height is 1.2m, the sampling frequency is 50Hz.

TABLE I. MAGNETIC MAP

Number	Calibration Points		
	Magnetic coordinate( $\mu T$ )	Position coordinate(m)	Position feature
1	(49.9, 52.0)	(0, 0)	starting point of front gate
2	(40.2, 43.1)	(16, 0)	turning point
3	(36.0, 32.1)	(16, -8)	between two large circular columns
4	(39.2, 36.8)	(16, -24.5)	empty place
5	(41.4, 41.5)	(16, -35)	turning point
6	(35.1, 35.2)	(33, -35)	turning point
7	(29.6, 32.8)	(33, -15)	between a column and a staircase.
8	(25.1, 29.3)	(33, -4)	between a column and an escalator
9	(39.3, 43.5)	(33, 15)	close to a small door
10	(38.0, 39.7)	(10, 15)	turning point
11	(44.3, 46.0)	(10, 40)	empty place
12	(47.0, 50.1)	(10, 65)	ending point of eastern gate

As we can see the magnetic map in the table I, we propose to measure positive left and right directions magnetic points to constitute a two-dimensional coordinates. Improving the dimension can make better accuracy of position calibration as shown in the experiment.

##### B. Particle Filter Algorithm

The basic idea of particle filter: according to the experience of conditions distribution on the system state vector, we will

produce a collection of random samples in the state space sampling which is called the sample set for particle. Then we will constantly adjust the weight of the particle size and the sample position according to observed values. Finally, we will adjust the particle information modification conditions of initial experience distribution to estimate the system states and parameters.

Particle filter has three basic steps of operation [11]: sampling (from not containing observations of the state space to generate new particles), weight calculation (based on observations to calculate the weight of each particle, resampling (abandon particles with small weight and replace the particles with larger weight). The three steps constitute the basic particle filter algorithm. The system state model and observation model of particle filter are shown respectively in formula (9) and (10).

$$x_k = \Phi x_{k-1} + \Gamma \mu_{k-1} \quad (9)$$

$$Z_k = Z_{kt} + v_k \quad (10)$$

where  $x_k$  is the pedestrian coordinates,  $\mu_{k-1}$  is system noise,  $Z_k$  is observed value,  $Z_{kt}$  is real value,  $v_k$  is observation noise, and all the parameters is in k time.

##### C. Magnetic Calibration

After the magnetic map established, we will go on the magnetic calibration based on particle filter algorithm. The map process is divided into four phases. First, the initialization of the particle filter is activated by detecting the first sampling correction point. The probability of all the particles that fall outside the space or in the space of the wall and the arrangement is directly set to zero [12]. Second, predicting the particles, as shown in the following formulas.

$$x_t^i = x_{t-1}^i + l_t^i \sin(\theta_t^i) + n_t \quad (11)$$

$$y_t^i = y_{t-1}^i + l_t^i \cos(\theta_t^i) + n_t \quad (12)$$

where  $x_{t-1}^i$ ,  $y_{t-1}^i$  is the t-1 time position coordinate for particle i,  $l_t^i$  is the average walking length acquired IPDR,  $\theta_t^i$  is the rotating direction for particle i,  $n_t$  is the zero mean Gauss noise. Third, establishing the magnetic field characteristic observation model, as shown in the following formula [13].

$$Z_t = Z_{rt} + n_g \quad (13)$$

where  $Z_t$  corresponds to the actual position of the current position of the magnetic induction intensity measurement ( $x_m, y_m$ ),  $Z_{rt}$  corresponds to the actual position of the current position of the magnetic induction intensity real value ( $x_r, y_r$ ). The last, we will realize the particle filter algorithm including initialization, sampling, weight calculation, normalization, resampling and output [14]. The weight calculation and normalization are shown respectively in formula (14) and (15).

$$\omega_i^j = p(Z_i | X_i^j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(Z_i - X_i^j)^2}{2\sigma^2}\right] \quad (14)$$

$$\omega_i = \frac{\omega_i^j}{\sum_{i=1}^N \omega_i^j} \quad (15)$$

V. EXPERIMENTS AND RESULTS

In this section, the operation of the proposed IPDR with magnetic calibration system is demonstrated. First, we describe the sensors in the smartphone. Then, we present results of a set of simulations.

To test our IPDR system, the SAMSUNG GALAXY SIIIIN9008 smartphone with various built-in sensors is used to implement this system. The device is equipped with a MPU6500 3-axis accelerometer, a MPU6500 gyroscope sensor, a YAS5323-axis magnetic field sensor and a BOSCH barometer sensor. The sampling frequencies are set to 50Hz for all sensors. The experiment is conducted on the first floor of No.3 teaching building in Beijing University of Posts and Telecommunications. The buildings magnetic map was already collected in the smartphone in advance.

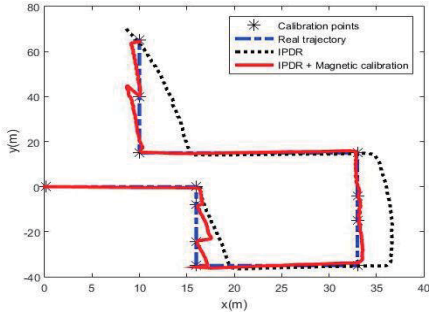


Fig. 5. Results of indoor position test

Pictured above is the simulation of the whole system. The asterisks are calibration points and the curves respectively stand for real trajectory, IPDR and IPDR with magnetic calibration. The black dotted line stands for IPDR which can be seen that maximum distance between path calculation and the real path is about 5m. This is because despite we improve the precision of modules, the gyroscope in the using process still has the cumulative deviations leading the path to a little error. The solid red line represents IPDR with magnetic calibration. As we can see, after magnetic calibration, the direction is still deflect because of gyroscope cumulative error. The biggest difference between the corrected track and the real track is within 2m.

It can be seen from the simulation results that the proposed algorithm can have a very good positioning accuracy in practical applications, and it has a certain practicality and feasibility.

VI. CONCLUSION AND FUTURE WORK

In this paper, an Improved Pedestrian Dead Reckoning combined with Magnetic Calibration algorithm is provided to achieve higher accuracy. In IPDR, we propose more normal pocket pattern direction detection algorithm instead of traditional flat pattern. In Magnetic Calibration, we propose a two-dimensional magnetic coordinates calibration points instead of only one dimension. According to experimental results and the comparison with other typical indoor location methods, our IPDR system with map matching performs well and is especially suited for indoor scenarios. In the future, we will learn to solve the problem of gyroscope cumulative error.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (61671077, 61671264).

REFERENCES

- [1] K. Sagawa, H. Inooka and Y. Satoh, "Non-restricted measurement of walking distance," in Systems, Man, and Cybernetics, 2000 IEEE International Conference on, vol.3, pp. 1847-1852, 2000.
- [2] C. K. Schindhelm, F. Gschwandtner, and M. Banholzer, "Usability of apple iphones for inertial navigation systems," in PIMRC, 2011 IEEE 22nd International Symposium on, pp. 1254-1258, IEEE, 2011.
- [3] S. H. Shin, C. G. Park, J. W. Kim, et al., "Adaptive Step Length Estimation Algorithm Using Low-Cost MEMS Inertial Sensors," in Sensors Applications Symposium, 2007. IEEE, pp. 1-5, 2007.
- [4] W. Kang, S. Nam, Y. Han, and S. Lee, "Improved heading estimation for smartphone-based indoor positioning systems," in PIMRC, 2012 IEEE 23rd International Symposium on, pp. 2449-2453, IEEE, 2012.
- [5] Y. Jin, M. Motani, W.-S. Soh, and J. Zhang, "Sparsetrack: Enhancing indoor pedestrian tracking with sparse infrastructure support," in INFOCOM, 2010 Proceedings IEEE, pp. 1-9, IEEE, 2010.
- [6] Z. Xiaojun, L. Qun and C. Guihai, "APT: Accurate outdoor pedestrian tracking with smartphones," in INFOCOM, 2013 Proceedings IEEE, pp. 2508-2516, 2013.
- [7] E. Le Grand and S. Thrun, "3-Axis magnetic field mapping and fusion for indoor localization," in Multisensor Fusion and Integration for Intelligent Systems, 2012 IEEE Conference on, pp. 358-364, 2012.
- [8] CHO Seong-yun, PARK Chan-gook. "Measurement system of walking distance using low-cost accelerometers," in Proceedings of the 4th Asian Control Conference, pp.16-21, 2002.
- [9] L. Perry-Lube and M. Lefkowitz, "Explorer - mobile navigation and interpretation at the american museum of natural history," in Museums and the Web 2011, Trant and Bearman,Eds., 2011.
- [10] K. Seong-Eun, K. Yong, Y. Jihyun, and S. K. Eung, "Indoor positioning system using geomagnetic anomalies for smartphones," in Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on, pp. 1-5, 2012.
- [11] B. Li, B. Harvey, and T. Gallagher, "Using barometers to determine the height for indoor positioning," in IPIN, 2013 International Conference on, pp.1-7, IEEE, 2013.
- [12] E. Akeila, Z. Salcic, A. Swain, A. Croft, and J. Stott, "Bluetooth-based indoor positioning with fuzzy based dynamic calibration," in TENCON 2010-2010 IEEE Region 10 Conference, pp. 1415-1420, IEEE, 2010.
- [13] Z. Mu, X. Yubin, and M. Lin. "Radio-map establishment based on fuzzy clustering for wlan hybrid kann/ann indoor positioning," China Communications, vol.7, no.3, pp.64-80, 2010.
- [14] L. Binghao, T. Gallagher, A. G. Dempster, and C. Rizos, "How feasible is the use of magnetic field alone for indoor positioning?" in Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on, pp.1-9, 2012.

# Relay Assisted Device-to-Device Communication: Approaches and Issues

Uyoata Uyoata\* and Mqhele Dlodlo†

Department of Electrical Engineering, University of Cape Town, Cape Town, South Africa

Email: \* uytuyo001@myuct.ac.za, † Mqhele.dlodlo@uct.ac.za

**Abstract**—Enabling technologies for 5G and future wireless communication have attracted the interest of industry and research communities. One of such technologies is Device-to-Device (D2D) communication which exploits user proximity to offer spectral efficiency, energy efficiency and increased throughput. Data offloading, public safety communication, context aware communication and content sharing are some of the use cases for D2D communication. D2D communication can be direct or through a relay depending on the nature of the channel in between the D2D devices. Apart from the problem of interference, a key challenge of relay aided D2D communication is appropriately assigning relays to a D2D pair while maintaining the QoS requirement of the cellular users. In this article, relay assisted D2D communication is reviewed and research issues are highlighted. We also propose matching theory with incomplete information for relay allocation considering uncertainties which the mobility of the relay introduces to the set up.

**Index Terms**—Device-to-Device(D2D)communication, relay allocation, channel uncertainty, mobile relaying.

## I. INTRODUCTION

Wireless communication has evolved in to a ubiquitous technology with widespread applications. It is estimated that future wireless communication technologies will need to cater for much higher speed and reduced latency than the present 4G networks. To enable the evolution towards 5G networks and beyond, D2D communication within cellular networks has been proposed as an enabling technology and has been included in third generation partnership project (3GPP)s Proximity Services (ProSe) standard [1]. D2D communication could be deployed as an overlay [2] to an existing cellular network where it uses orthogonal channels or as an underlay in which available resources are shared [3]. The former reduces interference while the later makes for better spectral efficiency. The gains associated with D2D communication include hop gain, reuse gain, and proximity gain [4]. While direct communication between devices is not new, D2D communication employs the already available network and so overcomes the manual pairing and access point definition associated with unlicensed spectrum communication. Studies in D2D communication have considered the feasibility of D2D underlying cellular networks [5], resource allocation, spectrum sharing [6] and coverage extension [7]. Cooperative communication has also been considered in [8] [9]. D2D communication can be direct between two mobile devices or relay assisted in which either a fixed low power relay is employed to forward signals or a mobile relay is used. The need to introduce relay to D2D communication arises when the distance between

the nodes is too far for direct communication or when the channel is severely impaired. Instances where relay assisted D2D communication can be used include: coverage extension and in content sharing networks.

For cellular communication, 3GPP has standardized radio relay technologies specifying Layer 1 Relay, Layer 2 Relay and Layer 3 Relay. The standardization of D2D communication is still in the works and associated relay techniques have not yet been standardized. A cell with multiple technologies is shown in Fig.1. This article reviews the approaches to relay aided D2D communication with a view of highlighting important research issues. This paper is structured as follows, Section II describes LTE-A relay technologies, an overview of D2D communication is given in section III, whereas resource allocation and performance analysis in relay aided D2D communication are considered in section IV. Section V outlays relay selection approaches in D2D communication while the application scenarios are presented in section VI. Research issues are discussed in section VII and the effect of channel uncertainties due to mobility is studied in section VIII. The paper is concluded in section IX.

## II. LTE-A RELAY TECHNOLOGY

Relay technology has been standardized by 3GPP for LTE-A with the aim of improving cell edge throughput and improving cell coverage. A relay provides access to UEs in the downlink and a wireless backhaul to the base station in the uplink. Radio relay technologies can be broadly classified as Layer 1 Relay, Layer 2 Relay and Layer 3 Relay. A layer 1 relay is simply a repeater station which amplifies its received radio signal and forwards it to the destination. Layer 2 relays amplifies received signals only after successful decoding/encoding and demodulation/modulation. Layer 3 relay is similar to Layer 2 relay but having radio protocols similar to an LTE base station. [10]. A summary of the features of the relay technology types is given in table 1. To adapt LTE-A relay technologies to D2D communication, consideration has to be made of whether the device acting as a relay is a fixed infrastructure of the type in Table 1 or a UE.

## III. D2D COMMUNICATION

A technology envisioned for future wireless communication is device to device (D2D) communication. Introduced in [11] as a means of allowing nearby devices communicate through multi-hops instead of a single hop via the base station D2D

TABLE I  
LTE-A RELAY TECHNOLOGY

Relay Technology	Features	Pros	Cons
Layer 1 Relay	Amplify and Forward	Simplicity of implementation Minimal processing delay	Noise and interference amplification
Layer 2 Relay	Decode and Forward	Noise and interference removal	Processing delay due to decoding and encoding
Layer 3 Relay	Decode and Forward Modulation/Demodulation User data regeneration processing	Noise and interference removal	Additional processing delay due to: decoding and encoding and user data processing

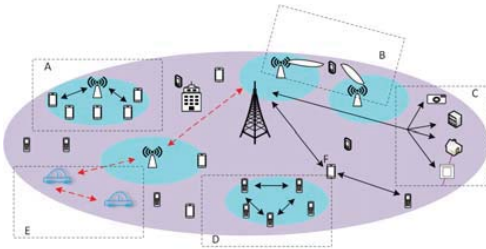


Fig. 1. Multiple technologies overlaid in a macro cell; A: Fixed relay assisted D2D communication; B:mmW communication;C:Massive machine type communication; D:Mobile relay assisted D2D/D2D Mesh;E: short range inter-vehicular communication; F: Mobile relaying

communication is now a technology item for LTE-Advanced. D2D communication has received attention in the academia, industry and has seen standardization efforts by the 3GPP. An actual implementation of a cellular D2D communication was FlashLinQ; a creation of Qualcomm that exploits the orthogonality in OFDM to provide peer-to-peer communication among devices in close proximity [12]. Primarily the instances where D2D communication can be strongly advocated for is in public safety communication when the cellular infrastructure is damaged, content sharing in close proximity (like multimedia in dens area), gaming among others. It allows for ad-hoc networks to be formed between user devices with little or no coordination from the macro cell or operator base station. D2D communication extends the fixed relaying which is already a part of the LTE-Advanced standard [13] to mobile relaying. Mesh networks in which device communicated directly are not new; Bluetooth, WLAN, HiperLAN2, TETRA, Infra-red, NFC are technologies which have allowed such communication in the unlicensed band although these largely are deployed in an unplanned nature. The 3GPP in an effort to integrate D2D communication in the licensed band into its standard has created study groups to research in this area [14]. Deployment of D2D communication in the same band as the macro-cell base station (and by extension other co-channel small cells), interference resource allocation, security valid concerns. Although deployed within the cellular network, D2D communication is not completely similar to UE to eNodeB(or BS) links especially with respect to the channel model. In D2D communication, both the transmitter and receiver are

mobile in contrast to the stationary cellular BS thus affecting the nature of the channel fading and the temporal correlation of shadowing. Furthermore the transmitter and receiver heights are lower with respect to base station heights and the distance between the devices are smaller considering envisioned dense deployment [1]. Resource management wise, managing D2D communication depends on whether the devices are within the coverage of a network or not. For the case where the D2D UEs are within the coverage of an eNodeB/BS, the base station provides management and control functions whereas outside the coverage areas, the devices can act in ad hoc mode or a device can assume the position of a cluster head although the later drains the battery of the device and so there could be incentives to encourage such. But from a public safety network perspective, an incentive may not be considered. D2D communication has been shown to offer various gains to the overall network performance by way of, increased user throughput [1], [15], [3], traffic offloading [3], [16], spectrum utilization [2], [3].

Although D2D communication targets close proximity devices, in a situation where the channel between the devices is impaired, relaying can be used to assist in forwarding the intended signals. Relays can be fixed or mobile. When relays are in motion at high speeds, obtaining accurate channel state information becomes difficult due to the channel variation rate. In the literature, the approaches in analysing relay aided D2D communication can be categorised into: resource allocation, relay selection and applications.

#### IV. RESOURCE ALLOCATION AND PERFORMANCE ANALYSIS APPROACHES

The envisaged dense and heterogeneous nature of future wireless networks have motivated wide research into resource allocation and interference management. For D2D communication, algorithms for mode selection, resource allocation, and power control have also been proposed. While most resource allocation algorithms have been centralised, distributed techniques have been developed for the seemingly ad-hoc nature of D2D communication. In [26], resource allocation for relay based D2D communication is considered in which devices pass messages D2D nodes exchange messages with the relay to determine suitability of transmission. The LTE-A L3 relay is used and sum rate is the metric for performance measure although relay selection is not considered. Still using LTE L3 Relay [27], numerical optimization is used for a centralised resource allocation in relay aided D2D communi-



cation. The optimization problem is formulated as a mixed integer non-linear problem for which relaxations are made for ease of analysis. Perfect channel estimations are assumed and a distance threshold is obtained beyond which using a relay does not offer the target gains. Still using the LTE-A L3 Relay as a platform the authors in [27] propose an algorithm robust against channel uncertainties which is modelled using a bounded uncertainty set with known probability [20]. The assumed sources of uncertainty are channel gains and interference at receiving ends. Although assuming a perfect channel can ease analysis, such analysis serve as benchmarks to those that consider imperfect channel conditions. Game theoretical approaches have been used for resource allocation in wireless communication. In [28] a promising approach namely matching theory is used for distributed resource allocation. The channel uncertainty is modelled as in [20] and LTE L3 Relay is also used. While resource allocation has attracted the interest of various researchers, other articles have focused on theoretical analysis and performance evaluation of relay based D2D communication. Using Monte Carlo simulations, Kiran Vanganuru et al [21] demonstrate the potential of mobile relaying in improving the capacity and coverage of a network. Decode and forward (DF) cooperative protocol is used at the mobile relay and the eNB centrally controls the assigning of relays. Considering multi-hop through UEs, Donghoon Lee et al [18] performed outage probability and ergodic capacity analysis for decode and forward D2D communication showing the gains thereof. The analysis assumes a single D2D pair and a single interfering cellular UE. Stochastic analysis is the approach employed by Akeam Al-Hourani et al [23] with the aim of proposing the optimum relay region within which energy is saved. Departing from other relay region formulation results, it proposes an elliptical relay region. Furthermore, a distance threshold is presented that provides the boundary beyond which relay assisted communication is not preferable.

## V. RELAY SELECTION IN RELAY AIDED D2D COMMUNICATION

Most of the literature considering resource allocation in relay aided D2D communication have either focused on a single fixed relay or multiple relays whose presence are either viewed as sources of interference or as part of a multi-hop chain. For efficient resource utilization, selecting a single relay out of a cluster of relays is also of importance. In [22] a non-centralised approach to relay selection is proposed in which a mobile relay is selected such that the interference from d2d communication to cellular UEs are minimized. The relay selection is timer based such that the relay experiencing least interference has the least time to forward signals from a D2D transmitter. Perfect channel estimation is assumed and multiple cellular links with a single D2D link is considered. Extending the work in [22], Chen Zhengwen et al [24] formulated relay selection as an optimization problem in which the capacity of the relay-D2D receiver link is optimized while ensuring the QoS requirement of cellular UEs. The relay selection is done jointly with resource allocation to limit interference to

cellular UEs. The paper assumes cognitive sensing at the relays and D2D nodes. Weichen Xia et al [29] proposed the use of both CSI and the distance between nodes as criteria for relay selection depending on whether D2D communication is indoor or outdoor respectively. Firstly a cluster of candidate relays is created using the distance/CSI criteria, then the specific relay is selected based on a defined channel efficiency. The metrics of performance are error probability and outage probability. Resource allocation is not the focus of the amplified and forward based work and perfect channel knowledge is assumed. In Boijang Ma et al [30], the problem of relay allocation is formulated as a matching problem between d2d pairs and candidate fixed relays. Bipartite graph is used and the aim is to minimize the power consumption of D2D communication. Yicha Chen et al [31] formulated the relay selection problem as bipartite graph for which an auction algorithm was proposed in which D2D pairs submitted bids for the candidate relay. The proposed algorithm shows increased throughput as against centralised relay allocation algorithm although complexity analysis was not performed. The issue of relay incentives is addressed in [32] where UEs through a learning algorithm determine the relay policy which utilizes their utility. The literature uses token passing between UEs and the proposed learning algorithm are tested in scenarios of various mobility. Perfect channel knowledge is assumed and the decision to relay is based on optimizing the relay utility under amplify and forward cooperative protocol.

## VI. RELAY ASSISTED D2D COMMUNICATION APPLICATION SCENARIOS

The use cases for relay assisted D2D communication include public safety communication and popular content distribution. Hao Xu [33] investigated the use of relay assisted D2D communication in distributing popular video contents within clusters of devices. Inter-cluster interference is considered and a UE relay is introduced when the distance between the D2D nodes is exceeds a threshold. The channel is assumed perfect and the effect of group mobility is not considered. Although dedicated public safety communication systems such as Terrestrial Trunked Radio (TETRA) are already available, their speeds do not suit emerging public safety scenarios. G. Fodor et al [34] proposed an architecture for d2d communication for public safety and disaster scenarios application. They proposed a cluster based approach where a cluster assumes the functionalities similar to base station functionalities for nodes that are out of coverage. Such roles include resource management and synchronisation. In situations where the elected cluster head is out of coverage, a UE within coverage assumes the roles. The architecture proposed serves full, partial and out of coverage scenarios. The application of mobile devices specifically smartphones in multi-hop relaying in disaster areas is proposed in [35] where adaptive algorithm which switches between two routing techniques namely mobile ad-hoc network (MANET) and disruption tolerant network(DTN) depending on the distribution of the mobile devices is presented. Message relaying through a developed mobile application was tested between

devices and an unmanned aircraft system was used to aid relay to a distant location. Mobile relaying can present security risks. In [25] beamforming at fixed relay is proposed as a technique to counter attempts by eavesdroppers to intercept information between a D2D pair. Physical layer network coding is employed at the relay node. Moreover channel uncertainty is considered between the eavesdropper and the entities involved in D2D communication although perfect channel estimation is assumed between the D2D entities.

## VII. RESEARCH ISSUES

Given the forgoing, research in relay based D2D communication is thriving although there are research issues that need be highlighted. While most research works in this area have focused on fixed relays a few have studied the use of mobile phones as the relay. Authors studying the use of UEs as relays generally assume that the UEs are within efficient relaying distance and have not analysed the efficient relay region for mobile relays. In studying mobile relaying most authors have assumed perfect channel estimation at the relay UE and the transmitting UE. This ignores the uncertainty which channel ageing due to mobility could introduce into the channel formulation. Where uncertainty is considered in the literature, uncertainty is introduced in the interference channel while the channel between a D2D transmitter and relay is assumed perfect. For direct communication such relaxation could hold considering the inter-distance, but for relay based communication, the channel uncertainties may occur. Some works have studied D2D communication in the light of 3GPP LTE/LTE-A standards. They have done so by using already standardized techniques such the 3GPP power control and Layer 3 Relay. Doing so can help fast track standardization procedure for relay aided D2D communication. While research into ensuring secured direct communication between D2D nodes is gaining steam, few works have considered security in relay assisted D2D communication. An overview of security in D2D communication with proposed solutions is given in [36]. Moreover while resource allocation, relay selection and performance analyses have been the focus of most papers, energy efficiency of relay assisted D2D communication has not been so much focused on. Furthermore most works assume that the relays and the D2D devices have single antennas. A few works have considered designing beamforming precoders for MIMO D2D communication to avoid intra-cell interference. Without loss of generality, a feature of static nodes in wireless communication is perfect channel estimation which allows for coherent detection. Mobility of users in a wireless network introduces time variation leading to errors in the channel estimation. When channel uncertainty introduced by device mobility is considered in D2D communication and a relay protocol such as the two way amplify and forward is used for a time division duplex system, the self-interference at the UEs are not completely cancelled due to the channel uncertainties. This makes the analysis less trivial. In the next section we show the effect of such channel uncertainty in relation to the outage probability. A summary of the research

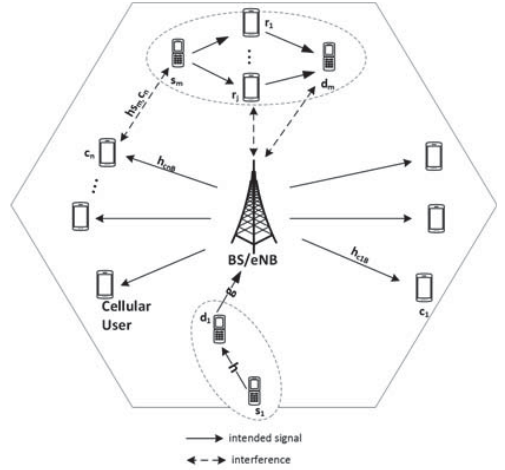


Fig. 2. Relay Assisted D2D Communication System Model

works on relay assisted D2D communication is provided in Table II.

## VIII. RELAY ASSISTED D2D COMMUNICATION AND CHANNEL UNCERTAINTIES

To show the effect of channel variation, a coverage extension scenario is considered in which there is a D2D UE  $s_1$  at the cell edge as in Fig.2. By using direct communication between  $s_1$  and an appropriately selected idle devices  $d_1$  in the range of  $s_1$ , the coverage of BS can be extended to  $s_1$ . Fig. 2 shows the system model. Time division duplex with two time slots is considered. In two way relaying, the relay broadcasts in the second time slot a scaled combination of the signals it received from two transmitters in the first time slot such that at the two transmitters cum receivers, the self-interference can be subtracted or cancelled and the target signal extracted. The channels between  $s_1$  and the  $d_1$  is given by  $h$  and between  $d_1$  and the BS/eNB is given by  $g$ .

### A. Problem Formulation

For a static or fixed relay,  $h$  and  $g$  can be modelled as circularly complex with zero mean and a variance value but for this case in which  $d_1$  is mobile, the channels must capture the uncertainty which relay mobility introduces. When the relay is in motion, the channel estimation is not perfect due to the time variation of the channel. Hence the channels  $h$  and  $g$  can be expressed as [37]

$$h(n) = a_1 h(n-1) + \Delta h(n) \sqrt{1 - a_1^2} \quad (1)$$

$$g(n) = a_2 g(n-1) + \Delta g(n) \sqrt{1 - a_2^2} \quad (2)$$

TABLE II  
 KEY RELATED LITERATURE

Paper	Approach	Solution Proposed	Uncertainty	Optimality	coordination	Relay Mobility
[26]	Resource allocation	message passing	No	Asymptotically optimal	Distributed	Fixed
[18]	Theoretical evaluation and performance analysis	Not applicable (NA)	No	NA	Centralised	Mobile
[21]	✓	✓	✓	✓	Centralised	Fixed
[27]	Resource allocation	Numerical optimization using Karush-Kuhn-Tucker	✓	Asymptotically optimal	Centralised	Fixed
[28]	✓	Matching Theory	Yes	Near Optimal	Distributed	Fixed
[20]	✓	Numerical optimization using Karush-Kuhn-Tucker	✓	✓	✓	Mobile
[24]	Resource allocation and Relay selection	Numerical analysis	No	Optimal	Distributed	✓
[32]	Relay selection	Numerical analysis	No	NA	Distributed	✓
[29]	✓	Performance analysis	✓	✓	Distributed	✓
[25]	Interference Management	MIMO beamforming	Yes	✓	Centralised	Fixed
[31]	Relay allocation	auction based bipartite matching	No	Optimal	Centralised	Mobile
[30]	✓	Bipartite matching Hungarian algorithm	✓	Suboptimal	Centralised	Fixed

Where  $a_i$ ,  $i=1,2$  indicates the channel variation rate related to the Doppler shift by the zeroth order Bessel function of the first kind (i.e.  $a[n] = J_0(2\pi f_D T[n])$ ), while  $\Delta h(n)$  and  $\Delta g(n)$  are the time varying component of the channel  $h$  and  $g$  and are independent and identically distributed with distribution  $\mathcal{CN}(0, \sigma_i^2)$ ,  $i=1,2$  [38]. Estimation (that is the knowledge the receiver has) of the channels are given by  $\hat{h}$  and  $\hat{g}$  are complex random Gaussian processes described as,  $\hat{h} \sim \mathcal{CN}(a_1 h_o(1 - a_1^2) \sigma_1^2)$ ,  $\hat{g} \sim \mathcal{CN}(a_2 g_o(1 - a_2^2) \sigma_2^2)$ . Without loss of generality it can be inferred that at an instant  $t_o$  where there is no channel variation or rather at the start of transmission,  $\hat{h}=h_o$  and  $\hat{g}=g_o$ .

Assuming for sake of tractability of analysis that the interference at  $s_1$  and BS are ignored, the measured SINRs at the  $s_1$  and BS can be expressed as:

$$\gamma_1 = \frac{\beta^2 \alpha_1^2 P_2 |h_o|^2 |g_o|^2}{2\beta^2 \alpha_1^2 P_1 |h_o|^2 (1 - \alpha_1^2) + \beta^2 \alpha_1^2 |h_o|^2 \sigma_2^2 + \beta^2 \alpha_1^2 P_2 |g_o|^2 (1 - \alpha_1^2) + \sigma_1^2} \quad (3)$$

$$\gamma_2 = \frac{\beta^2 \alpha_1^2 P_1 |h_o|^2 |g_o|^2}{\beta^2 \alpha_1^2 P_1 |g_o|^2 (1 - \alpha_1^2) + \beta^2 P_1 |g_o|^2 + \sigma_2^2} \quad (4)$$

Where  $P_i$  is the transmit power of node  $i$ ,  $\beta$  is the amplification factor at  $s_1$  associated with amplify and forward relaying,  $\sigma_n$  is the noise variance at node  $n$ . Denote data rates at  $s_1$  and BS as  $R_1$  and  $R_2$  respectively. Therefore for the data rate at the outage probability  $P_{out}$  given a target/threshold data rate of  $R_{th} = 5 \text{ bits/s/hz}$  can be expressed as;

$$P_{out} = P_r(\min(R_1, R_2) < R_{th}) \quad (5)$$

The probability that the cell edge user  $s_1$  is in outage when a helper node with which it forms a D2D pair is in mobility is investigated in this section. Firstly the impact of such mobility is demonstrated for selected variation rate of the channel ( $a_1 = 0.998, 0.899, 0.799$ ). Fig 3 shows a sharp increase in outage probability as the values of  $a_1$  approaches

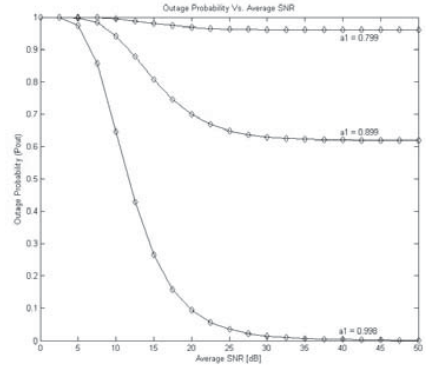


Fig. 3. Probability of outage vs. SNR

0 which is not counter intuitive. Fig. 3 demonstrates the effect of time variation in mobile relaying. Considering that future cellular network will be dense and the mobility of devices are largely independent, the variation rate of the channels between a D2D transmitter and possible receivers (in this case potential mobile relay) vary. The effect of relay mobility points to the need to factor in the uncertainty which mobility introduces into D2D communication channel in analysis and so develop robust algorithms for relay allocation. An approach that can be viable is matching theory with incomplete information.

## IX. CONCLUSION

In this work, a review of key literature on relay assisted has been provided highlighting addressed issues and open areas to further explore. We have made a case for the channel uncertainties in D2D communication resulting from device mobility and suggested matching theory with incomplete infor-

mation as a viable approach to such uncertainties. In a sequel, matching with incomplete information will be employed to achieve desired relay allocation to the D2D links.

REFERENCES

[1] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3gpp device-to-device proximity services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, April 2014.

[2] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Oct 2008, pp. 1537–1541.

[3] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, Dec 2009.

[4] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Mikls, and Z. Turnyi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, March 2012.

[5] C. H. Yu, O. Tirkonen, K. Doppler, and C. Ribeiro, "On the performance of device-to-device underlay communication with simple power control," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, April 2009, pp. 1–5.

[6] T. Koskela, S. Hakola, T. Chen, and J. Lehtomaki, "Clustering concept using device-to-device communication in cellular system," in *2010 IEEE Wireless Communication and Networking Conference*, April 2010, pp. 1–6.

[7] A. Abrardo, G. Fodor, and B. Tola, "Network coding schemes for device-to-device communications based relaying for cellular coverage extension," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 670–674.

[8] C. Ma, G. Sun, X. Tian, K. Ying, H. Yu, and X. Wang, "Cooperative relaying schemes for device-to-device communication underlying cellular networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 3890–3895.

[9] S. Shalmashi and S. B. Slimane, "Cooperative device-to-device communications in the downlink of cellular networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2014, pp. 2265–2270.

[10] M. Iwamura, H. Takahashi, and S. Nagata, "Relay technology in lte-advanced," *NTT DoCoMo Technical Journal*, vol. 12, no. 2, pp. 29–36, 2010.

[11] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, Mar 2000, pp. 1273–1282 vol.3.

[12] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, "Flashing: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1215–1228, Aug 2013.

[13] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis, and X. Shen, "Relaying operation in 3gpp lte: challenges and solutions," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 156–162, February 2012.

[14] 3GPP, "3rd generation partnership project; technical specification group sa; study on architecture enhancements to support proximity services (prose) (release 12)," *TR 23.703 V0.4.1*, 2013.

[15] N. Golrezai, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, July 2014.

[16] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "Proximity-based data offloading via network assisted device-to-device communications," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, June 2013, pp. 1–5.

[17] S. Wen, X. Zhu, Y. Lin, Z. Lin, X. Zhang, and D. Yang, "Achievable transmission capacity of relay-assisted device-to-device (d2d) communication underlay cellular networks," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, Sept 2013, pp. 1–5.

[18] D. Lee, S. I. Kim, J. Lee, and J. Heo, "Performance of multihop decode-and-forward relaying assisted device-to-device communication underlying cellular networks," in *Information Theory and its Applications (ISITA), 2012 International Symposium on*, Oct 2012, pp. 455–459.

[19] Y. Li, L. Sun, and W. Wang, "Exploring device-to-device communication for mobile cloud computing," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 2239–2244.

[20] M. Hasan, E. Hossain, and D. I. Kim, "Resource allocation under channel uncertainties for relay-aided device-to-device communication underlying lte-a cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2322–2338, April 2014.

[21] K. Vanganuru, S. Ferrante, and G. Sternberg, "System capacity and coverage of a cellular network with d2d mobile relays," in *MILCOM 2012 - 2012 IEEE Military Communications Conference*, Oct 2012, pp. 1–6.

[22] X. Ma, R. Yin, G. Yu, and Z. Zhang, "A distributed relay selection method for relay assisted device-to-device communication system," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, Sept 2012, pp. 1020–1024.

[23] A. Al-Hourani, S. Kandeepan, and E. Hossain, "Relay-assisted device-to-device communication: A stochastic analysis of energy saving," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[24] C. Zhengwen, Z. Su, and S. Shixiang, "Research on relay selection in device-to-device communications based on maximum capacity," in *Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on*, vol. 3, April 2014, pp. 1429–1434.

[25] K. Jayasinghe, P. Jayasinghe, N. Rajatheva, and M. Latva-aho, "Physical layer security for relay assisted mimo d2d communication," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 651–656.

[26] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication: A message passing approach," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6326–6341, Nov 2014.

[27] —, "Resource allocation for network-integrated device-to-device communications using smart relays," in *2013 IEEE Globecom Workshops (GC Wkshps)*, Dec 2013, pp. 591–596.

[28] —, "Distributed resource allocation for relay-aided device-to-device communication under channel uncertainties: A stable matching approach," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3882–3897, Oct 2015.

[29] W. Xia, S. Shao, and J. Sun, "Relay selection strategy for device to device communication," in *Information and Communications Technologies (IETICT 2013), IET International Conference on*, April 2013, pp. 318–323.

[30] B. Ma, H. Shah-Mansouri, and V. W. S. Wong, "A matching approach for power efficient relay selection in full duplex d2d networks," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[31] Y. Chen, S. He, F. Hou, Z. Shi, and X. Chen, "Optimal user-centric relay assisted device-to-device communication: an auction approach," *IET Communications*, vol. 9, no. 3, pp. 386–395, 2015.

[32] N. Mastronarde, V. Patel, J. Xu, L. Liu, and M. van der Schaar, "To relay or not to relay: Learning device-to-device relaying strategies in cellular networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1569–1585, June 2016.

[33] H. Xu, Y. Pan, N. Huang, Z. Yang, and M. Chen, "Relay-assisted device-to-device communications for video transmission in cellular networks," in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, Dec 2015, pp. 144–149.

[34] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahm, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, pp. 1510–1520, 2014.

[35] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-smartphone: realizing multihop device-to-device communications," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 56–65, April 2014.

[36] M. Alam, D. Yang, J. Rodriguez, and R. A. Abd-alhameed, "Secure device-to-device communication in lte-a," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 66–73, April 2014.

[37] K. S. Gomadam and S. A. Jafar, "Impact of mobility on cooperative communication," in *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006.*, vol. 2, April 2006, pp. 908–913.

[38] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, July 2005.

# Energy-efficient Management System for Smart Air Conditioners

Sarang Gosavi and Neeli Prasad

International Technological University, San Jose, CA, USA

**Abstract** – Energy management is one of the most sought out topic. The research on energy consumption and management is continuously undertaken. In this paper, a novel control mechanism is designed, which enables air conditioning to automatically adjust its temperature on the basis of user's preference, user's location and comparing outside and inside temperature by using a proposed adaptive algorithm. This will be an efficient energy management system for smart air conditioners use case in Internet of Things (IoT) application Smart Home.

**Keywords** – *efficient energy management; SMART A/C; IoT; Network Model; GPS; Bluetooth*

## I. INTRODUCTION

Air conditioning (A/C) has been around for about a century and for many people it is just used for comfort. For people in hotter regions of the world, it is used to combat the extreme heat. Air conditioning is said to give major contribution to global warming, due the use of Halogenated Chloro Fluoro Carbons (HCFC's) and Hydro Fluoro Carbons (HFC's) as in a refrigerant.

Usage of electricity for air conditioning in USA is more than all other countries combined. The residential air conditioning needs alone devour 6-8% of the country's electricity, or about 290 terawatt hours per year, more electricity than is used in Mexico, with population of 125 million. Hot states, e.g. Florida and Texas, have 98%+ of homes using air conditioning. The air conditioners in our cars alone use some 655,000 barrels of gasoline per day, more gasoline than Germany and Italy use in total combined [1].

China has been selling 50-75 million home air conditioners per year, in a country where coal is 77% of electricity. India, where coal is 70% of electricity, air conditioning already accounts for nearly 45% power use in Mumbai, 23 million people sweltering in perhaps the world's largest hottest city, still consuming just 7-10% of the electricity that New Yorkers do. Dr. Michael Sivak, a leading expert at the University of Michigan, has found that the potential cooling demand in Mumbai is almost 25% of entire USA [2].

In Middle East, for example, about 55% of Saudi Arabia's peak summer power consumption goes to air conditioning, using over 1 billion barrels of oil a year.

The ways to use air conditioners in an energy efficient manner can be linked to the smart devices off the shelf and reachable by everyone. Many countries, such as USA, UK, Japan, etc., who are already at an advanced stage of the design and

development of newer technologies, emphasizes on energy management and use of green energy.

Global warming potential (GWP) is a relative measure of how much heat a greenhouse gas traps in the atmosphere [3]. It compares the amount of heat trapped by a certain mass of the gas in question to the amount of heat trapped by a similar mass of carbon dioxide (CO<sub>2</sub>). The larger the GWP, the more a given gas warms the Earth compared to CO<sub>2</sub> over that time period. Chlorofluorocarbons (CFC's), HFC's, HCFC's, Perfluorocarbons (PFC's), and sulfur hexafluoride (SF<sub>6</sub>) are called high-GWP gases, because, for a given amount of mass, they trap substantially more heat than CO<sub>2</sub>. Table 1 gives a relative measure of GWP for various gases [3, 11].

This paper proposes a design for the energy management system for residential air conditioning system that controls the temperature of the room based on the number of people in the room. This works along with an application installed in the smart device. This also gives real-time data about the energy consumed by the air conditioner unit, and also the amount of carbon usage.

Sensors with Bluetooth are deployed in the hall that detects the number of occupants. This information will be sent to a server that not only stores these values in the database but also triggers different functions on the AC Controller Application. This application decides on the basis of a number of occupants, e.g., what should the temperature of the room be to provide a comfortable experience for everyone present in the room?

Table 1: Gases and GWP [3]

GAS	GWP
Carbon dioxide (CO <sub>2</sub> )	1
Methane (CH <sub>4</sub> )	21
Nitrous oxide (N <sub>2</sub> O)	310
Sulfur hexafluoride (SF <sub>6</sub> )	23,900
Perfluorocarbons (PFCs)	6,500
Hydrofluorocarbons (HFCs)	
HFC-23	11,700
HFC-32	650
HFC-125	2,800
HFC-134a	1,300
HFC-143a	3,800
HFC-152a	140

Further, in the paper, Section II gives a detailed description of the technologies used to implement the proposed idea. In

Section III discusses the network model and the components used for implementation, supplemented by a block diagram, and finally Section IV concludes the paper.

## II. PROPOSED USE CASE DESCRIPTION

GPS will calculate the user's location and the time (see Figure 1 and 2); how long the user will take to reach home? The A/C has already pre-configured settings based on the user's preferences. The sensor in the A/C unit will then automatically set the home temperature based on the arrival time of the user. Bluetooth technology is used to count the number of people in a room. Sensors attached to the A/C units calculate the amount of power consumed by the unit and then transmit the data to the wireless device through an application installed on it. This data is further stored and used in calculating the usage of CO<sub>2</sub>, energy, etc. It also predicts the energy bill for that month. The user can do various tasks to control the A/C through the application installed on his/her smart device. These functions can be turning it on or off, sets its temperature, put it in the sleep mode, to save power, etc. The application can also calculate average A/C usage for over a period of time and suggest a maximum bill amount for each month accordingly (see Figure 1).

By taking the consideration of cost per KWh set by the power agency, user can set the threshold limit and get alerts when energy consumption is maxed.



Figure 1: Smart residential Air Condition

The possible features of the application:

- Adjusting of the temperature while the user is still on his way to home
- Automatically switching off the A/C when nobody presents in the room or house.
- A/C will run only in the room where user is present by determining his location inside the house.

- Shows the energy consumption by A/C and suggest the reasons. The more user know, the more he can save.
- View and edit user's A/C temperature schedule. Get extreme temperature alerts before the pipes freeze or pet overheats.
- Earn points when you choose a temperature that saves you energy usages.

## III. NETWORK MODEL

**GPS application** - GPS and Wi-Fi technologies, both will be used in this Smart A/C (see Figure 3 and 5) [4, 7]. Here GPS is used to locate the user's presence as this information will be used by application in smartphone and according to data smart A/C will take decision. When the smartphone is in Wi-Fi range, A/C connects through Wi-Fi but when it gets out of the Wi-Fi range, it will connect through GPS. An application is installed on the user's smartphone that connects him with the A/C. This application will track user's location and sync with the home A/C. It will calculate user's distance from home and take appropriate actions that are already preconfigured [8].

For example, when the user leaves for his office in the morning, the distance will get increased from home as he moves ahead. The application will send a notification to the smart A/C controller, which checks the status of the A/C. If it finds the A/C is still on, it will switch it off. Similarly, when the user is on its way back to home, the application will calculate the time user takes to reach home. It will then alert the A/C controller, which switched-on the A/C and starts cooling the house on a very slow pace. In this way, when the user reaches home, he will receive a pleasant environment. Also, by turning the A/C in a slow pace will save a great amount of energy as the controller slowly adjust the temperature to the desirable settings instead of instantaneously trying to jump to it in the case if user reaches home and turn-on the A/C on the full speed. This approach would create a pleasant climate at home when the user reaches and he would not have to turn on the A/C and wait till it becomes pleasant.

If the user is going away and nobody in the house it will turn off itself and also it will send a signal to the main controller to turn off fans and lights in the house. The AC would be smartly calculating users daily pattern roaming in the house i.e. when the user reaches home which room he enters in which room, how much time he spends usually in that room and how and when does he move to other rooms etc. this pattern would be used to efficiently switch on and off different rooms AC at different times rather than switching all the A/C outlets at once, this would save a lot of electricity.

The Mobile app would send current temperature of user's location frequently to A/C. This data would be used to

calculate what would be the best temperature user would feel most comfortable in. for example, if the user has been in 90F, 85F, 87F, and 92F AC this data would be sent to AC from user's app, AC would then compare the temperature near user's house and the data received from user's mobile app and would calculate the best temperature and automatically sets the temperature. During summer if the inside temperature is getting hotter then it will automatically turn on the A/C exhaust system which would take hot air present inside the house to outside till the house's internal temperature drops by a couple of degrees then it would turn on the cooling system. This would save a good amount of power and electricity required to cool the house.

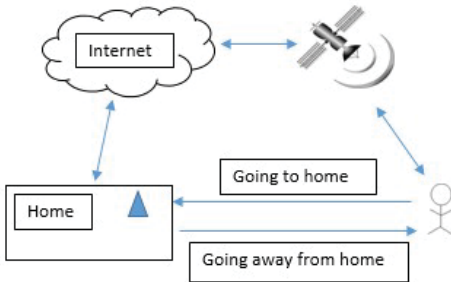


Figure 2: Application using GPS to locate user's location and send info to A/C controller. [13]

**Bluetooth** - In this technology, Bluetooth beacons are used, which have a long battery life. Human body has 60% water, and when Bluetooth signals pass through human body, they get obstructed [9]. This causes attenuation in Bluetooth beacons. So when a person walks between a Bluetooth transmitter and receiver, the attenuation in the Received Signal Strength Indication (RSSI) is used to indicate that a person has crossed the path [9].

The variation in RSSI is detected and the information is passed on to a microcontroller, which is programmed to operate as a counter that increments by 1 when there is attenuation in the RSSI. A system using infrared sensors needs to be installed by specialists, but the motion detector using Bluetooth beacons can be easily set up [9]. This technology suits well for offices and enterprises, where we have a large number of people.

The Bluetooth beacon transmits a signal containing a unique ID, and the Bluetooth receiver measures the RSSI of the appropriate signal. The variation in RSSI is detected and the information is passed on to a microcontroller, which is programmed to operate as a counter that increments by 1 when there is attenuation in the RSSI. A system using infrared sensors needs to be installed by specialists, but the motion detector using Bluetooth beacons can be easily set up. This

technology suits well for offices and enterprises, where we have a large number of people.

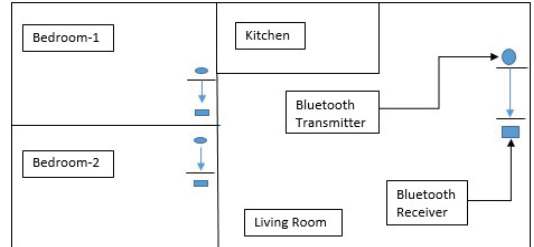


Figure 3: People counter implemented using Bluetooth

**Wi-Fi** - Wi-Fi signals can be used for more than just providing Internet access; they can actually be used to detect the presence of a person in a given space. This involves tracking the device of the person. In this case, the user is required to carry a smart device that is connected to the same Wi-Fi [5, 6]. If the user is in the living room along with his smart device, then the wireless sensors attached to the air conditioner can detect the user presence and turn on the air conditioner only for the living room. But the drawback of this technology is that the user is required to carry the smart device all the time. Figure 4 shows a detailed block diagram to detect a person's presence in space.

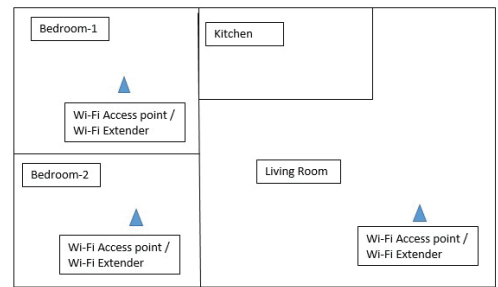


Figure 4: Wi-Fi to detect a user's presence

**Voice Recognition and Identification system** - In the above Wi-Fi technology user has to carry his/her mobile device either smartphone/ smartwatch/ iPad or any smart wearable device so that technology will locate and recognize the user but what if the user doesn't carry any device? For this drawback, we have added voice recognition and identification system so that if the user forgot its device or if device battery dies then using user's voice it will recognize the user and according to his/her preference, it will adjust the room temperature [12].

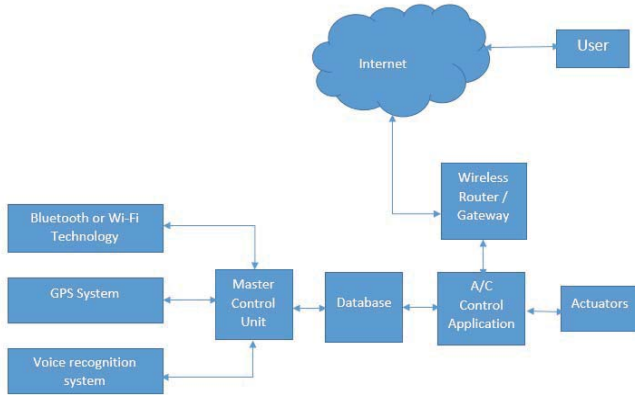


Fig 5: Smart A/C Network model

**Algorithm (software)** - In this smart air conditioning effective algorithm will run on a chip integrated with the smart A/C (see Figure 5), which will calculate the how much time is required and how much energy required to set the temperature in a room based on outside temperature. The system will also locate user's location and it will automatically adapt the user's preference to set the temperature. And according to time required by user to reach home it will gradually set the user's preferred temperature. A/C will use exhaust system to let go inside hot air outside by using this technique time and energy required to cool the inside temperature is less.

For example, if the outside temperature is 78 degrees Fahrenheit, and the user generally sets to 67-70 degrees, then when user reaching home the system will automatically turn on and gradually set the comfortable temperature that mean it will directly not set user's preference because if there is more temperature difference then air conditioner required more energy to set the user's requirement but instead it will gradually set the temperature like first it will set to 75 degrees Fahrenheit then it will set to 72 degrees Fahrenheit. In this way, the system will gradually set the user's preference. By doing this technique we can save a lot of energy.

IV. CONCLUSION

Save Money, Save the Earth. We are already aware that due to energy production and use of urban smog, oil spills, acid rain, and global warming causing most of the environmental problems. Figure 6 gives an idea about the phasedown targets to reduce the use of HFC's as a refrigerant [10].

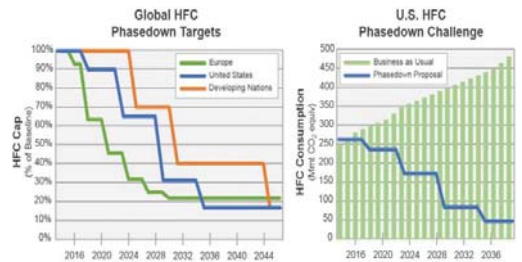


Figure 6: GFC Reduction targets in U.S and Global level. [10]

But each of us can make difference by taking energy use into account in our household purchasing and maintenance decisions. For an average household with two cars, energy used in the house accounts for more than half of its total energy expenditures.

In fact, choosing energy-efficient appliances like the air conditioning system proposed in this paper is one way of immediately reduce our contribution to global climate change. Carbon dioxide (CO<sub>2</sub>) is the primary gas contributing to global warming and virtually all energy-using equipment results in CO<sub>2</sub> emissions either directly or indirectly. Most of our electricity is likely to come from burning coal or gas at the power plant. A/C systems producing more CO<sub>2</sub> than vehicles. So by replacing a 20-year-old Air Condition system with a new, energy-efficient model, we will reduce our home's CO<sub>2</sub> contribution by about one ton every year while saving about \$80 in reduced electric bills.



The next steps, is to implement the proposed system and study the impacts.

REFERENCES

[1] <http://www.forbes.com/sites/judeclemente/2015/11/01/global-warming-means-more-air-conditioning-and-more-energy-demand/#6e05a4b426dd>

[2] Michael Sivak, "Potential energy demand for cooling in the 50 largest metropolitan areas of the world: Implications for developing countries", *Journal of Energy Policy*, Elsevier, Vol. 37, issue 4 (Apr. 2009), p. 1382-1384.

[3] Green House Protocols, <http://www.ghgprotocol.org>

[4] "Make your AC smart with tado", *tado*, 2016. [Online]. Available: <https://www.tado.com/us/>.

[5] "Never Lost Indoors – The Promise Of Wi-Fi® Location | Wi-Fi Alliance". *Wi-fi.org*. N.p., 2016. Web. 30 July 2016.

[6] Wenz, John. "The Surprising Way Wi-Fi Can Count The People In A Room". *Popular Mechanics*. N.p., 2015. Web. 30 July 2016.

[7] "Quirky + GE Aros Smart Window Air Conditioner - - Amazon.Com". *Amazon.com*. N.p., 2016. Web. 30 July 2016.

[8] H. Kim, K. Lee and S. Lee, "Location-based human-adaptive air conditioning by measuring physical activity with a non-terminal-based indoor positioning system", *Building and Environment*, vol. 62, pp. 167-173, 2013.

[9] Kyohei Sugino, Yusuke Niwa, Shun Shiramatsu, Tadachika Ozono, Toramatsu Shintani, "Developing a Human Motion Detector using Bluetooth Beacons and its Applications", *Information Engineering Express, International Institute of Applied Informatics*, vol. 1, No. 4, pp. 95-105, 2015.

[10] "Energy Department Releases Roadmaps on HVAC Technologies, Water Heating, Appliances, and Low-GWP Refrigerants", *Energy.gov*, 2016.

[11] "Addressing Global Climate Change | Steady State Revolution", *Steadystaterevolution.org*, 2016.

[12] Control your TV or air conditioner with voice commands, "Control your TV or air conditioner with voice commands, courtesy of Smart IR Remote", *Phone Arena*, 2014.

[13] Satellite clipart: <http://www.gograph.com/vector-clip-art/satellite.html>

# *Energy-Harvesting Aware Clustering Scheme for Wireless Sensor Networks with Non-uniform Energy-Harvesting rate*

Daiki MAEMOTO, Kazuo MORI, Kosuke SANADA, and Hideo KOBAYASHI

Graduate School of Engineering, Mie University  
1577 Kurimamatiya-cho, Tsu-shi, Mie, 514-8507, Japan

Email: maemoto@com.elec.mie-u.ac.jp, {kmori, k.sanada, koba}@elec.mie-u.ac.jp

**Abstract**— In cluster-based energy harvesting wireless sensor networks (EH-WSNs), clustering schemes considering harvesting power at energy harvesting nodes have been studied for reducing power consumption and prolonging network lifetime. However, the data collection rate would degrade in the EH-WSNs with geographical non-uniform energy-harvesting rate since isolated sensor nodes (SNs) increase due to concentration of cluster heads (CHs) at high energy-harvesting rate area. The EHGF has been proposed to suppress the performance degradation, however, this scheme results in non-uniform residual power between CHs caused from equal consumed power due to no consideration of difference in the energy-harvesting rate among CHs selected at small areas. This non-uniform residual power leads to short network lifetime. This paper proposes a clustering control scheme to reduce non-uniform residual power between the CHs. In the proposed scheme, each CH controls its cluster size based on its estimated available consuming power so as to reduce its communication load when locating at the area with low energy-harvesting rate. The simulation results show that the proposed scheme improves the network lifetime compared with the EHGF.

**Keywords**—WSN, energy harvesting, non-uniform energy harvesting rate, cluster size control

## I. INTRODUCTION

Wireless sensor networks (WSNs) consists of sensor nodes (SNs), which are distributed over sensing area, and a sink node (SINK), which plays a role of a central control station. SNs transmit sensed data to the SINK through multi-hop communication. WSNs have been employed in environmental monitoring systems, for example, forest and farm monitoring, and are expected to be applied for various practical uses [1]. Since SNs generally operate by battery with limited power source, and may be deployed to the locations where the battery is hard to replace, the WSNs require power saving mechanisms in order to prolong their network lifetime. Then, energy harvesting wireless sensor networks (EH-WSNs) [2,3] has been studied to prolong the network lifetime. The EH-WSNs enable to prolong their network lifetime by utilizing power re-chargeable function at energy harvesting nodes (EHNs) through energy-harvesting mechanism. However, the harvesting power is generally small and unstable depending on local circumstance because EHNs mainly employ solar energy generation. Thus EH-WSNs require power saving operation depending on the amount of the harvesting power.

For the power saving in EH-WSNs, a cluster based network has been investigated, where representative nodes,

cluster heads (CHs), selected from a given area collect sensed data from their subordinate SNs and transmit the aggregated data to the SINK. The cluster based network can reduce power consumption due to a reduction in the communication range at the SNs. However, the CH has to deal with large amount of the data communications, and the power consumption for the data communications concentrates at the CHs. Therefore, periodic cluster reforming (here after called “clustering”) needs to be performed every fixed time duration (here after called “round”), since all nodes have to share the power consumption by changing a role of the CH periodically to prolong the network lifetime.

For the clustering technique in EH-WSNs, energy harvesting-geographic aware fidelity (EHGAF) [4] has been proposed as an enhanced version of adapting geographic aware fidelity (GAF) [5], which has been originally proposed for the WSNs without energy-harvesting function. The EHGAF divides network area into some small square areas and selects one CH from each small area. This conventional scheme employs residual power at the EHNs as CH selection criteria and selects the EHN with the largest estimated residual power as a CH in the next round. Therefore, the EHGAF can suppress the network performance degradation due to the power depletion at the CHs. Furthermore, the CHs can deploy uniformly over the network area even under the condition with geographical non-uniform energy-harvesting rate. Therefore, CHs can keep shorter communication range from the SNs, leading to small power consumption and an improvement in data collection rate. However, this conventional scheme doesn't consider a difference in the amount of estimated residual power between the CHs for the selection of the connecting CH at the SNs. Thus, all the CHs consume almost the same power for their data communications since each CH subordinates almost the same number of SNs in despite of its different estimated residual power. As a result, the residual power results in non-uniformity between the CHs since the ratio of power consumption to harvesting power gets larger in the area with lower energy-harvesting rate under geographical non-uniform energy-harvesting condition. Therefore, the network lifetime decreases since the EHNs at the lower energy-harvesting area are likely to deplete their power resource.

The objective of this paper is to provide a clustering scheme with longer network lifetime by reducing non-uniformity in the residual power between the EHNs in the EH-WSNs with geographical non-uniform energy-harvesting rate. For achieving the objective, this paper proposes a novel clustering control scheme, which reduces the communication

load at the CHs locating at the lower-harvesting area and then suppresses non-uniform residual powers between the CHs by controlling their cluster size based on estimated available consuming power. The simulation results show that the proposed scheme can achieve larger network lifetime compared with the conventional one and can provide a solution against the problem with the conventional scheme.

## II. RELATED WORK

This section describes the existing clustering schemes, GAF [5] for WSNs, and EHGF [4], which is enhanced from the GAF so as to be applied to EH-WSNs.

### A. Geografic Aware Fidelity (GAF)

The GAF intends to reduce power consumption at SNs by deploying the CHs uniformly over the network area so as to provide almost the same communication distance for any data communications.

The GAF divides the network area into some small square areas, and then selects a CH from each small area every fixed time duration (round). This scheme can reduce power consumption by turning off the wireless communication function at the nodes except the CHs during the fixed time duration until the next round.

In the CH selection, each EHN periodically transmits a control packet containing numerical rank, which indicates fitness to play a role of a CH, and compares the rank included in the packets received from others with its own rank. If its own rank is lower than one of the received ones, the EHN changes its operation state into sleep state and then stops its communication function since it judges other one is suitable to be in charge of a CH. This operation repeats during the fixed time duration, and then one EHN is eventually selected as a CH from each small area. The GAF can select the EHN having the highest residual power by employing the amount of residual power as the CH selection criteria (rank).

In the EH-WSNs, however, the GAF results in inappropriate CH selection, where the EHN with the large amount of residual power but lower re-charging power by the energy harvesting function is also selected as a CH since the CH selection is performed in spite of the amount of harvesting power at the EHNs. Therefore, some selected CHs are likely to deplete their power resource until the next round.

### B. Energy harvesting GAF (EHGAF)

Against the above problem, the EHGAF have been studied in order to select the EHN having the large amount of both residual and harvesting power as a CH.

The EHGAF selects the EHN with the highest rank as a CH at the start of every round, by employing the rank of the amount of estimated residual power at the end of the current round. The rank of the EHN  $i$  in a given round  $r$ ,  $\hat{E}_{rst}^i(r)$ , is given by:

$$\hat{E}_{rst}^i(r) = E_{rst}^i(r) + E_{hrv}^i(r) - E_{CH}^i(r), \quad (1)$$

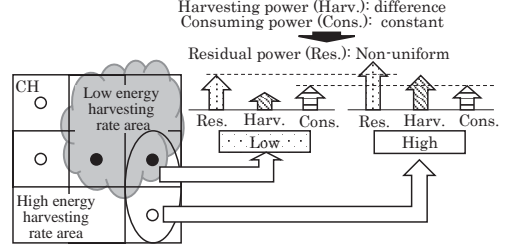


Fig.1 Problem with the EHGAF

where  $E_{rst}^i(r)$  is the amount of residual power at the start of the round  $r$ ,  $E_{hrv}^i(r)$  and  $E_{CH}^i(r)$  are the estimated amount of harvesting power and consuming power during the round  $r$ , respectively.  $E_{CH}^i(r)$  is calculated by:

$$E_{CH}^i(r) = m^i * (E_{Tx}^i(r) + E_{Rx}^i(r)), \quad (2)$$

where  $m^i$ ,  $E_{Tx}^i(r)$  and  $E_{Rx}^i(r)$  are the number of subordinate SNs in each cluster, the amount of power consumption for data transmission and reception, respectively. The EHGAF can achieve longer network lifetime by employing the rank of the amount of the estimated residual power.

However, the EHGAF utilizes the estimated residual power for the CH selection, but does not for the selection of the connecting CH at the SNs. In geographical non-uniform energy-harvesting rate conditions, as shown in Fig.1, the EHNs cannot recover their residual power at the area with low energy-harvesting rate since the amount of power consumption at the CHs are constant in spite of their  $\hat{E}_{rst}^i(r)$ , and consuming power may be larger than harvesting power. Therefore, the network lifetime gets shorter since the selected CHs are likely to deplete their power resource when locating at the lower energy harvesting area.

## III. CLUSTERING CONTROL SCHEME CONSIDERING GEOGRAPHICAL NON-UNIFORM ENERGY-HARVESTING RATE

This section proposes a clustering control scheme, which shares the power consumption for the data communications at CHs, by controlling the number of their subordinate SNs based on the amount of estimated available consuming power. The proposed scheme consists of three control phases, dividing the network area into some small areas, selecting a CH from each small area, and controlling cluster size based on the estimated available consuming power at selected CHs. The network area division and the CH selection basically employ almost the same control as the EHGAF.

### A. Network area division

This phase divides the network area into some small square areas, like the GAF and the EHGAF, in order to deploy the CHs uniformly over the network area, and selects a CH from each small area every round.

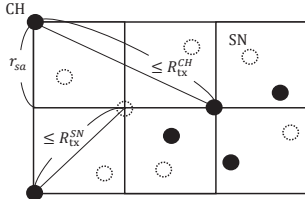


Fig.2 Network area division

In the network area division, as shown in Fig.2, the CHs need to be deployed within the transmission range of SNs,  $R_{tx}^{SN}$ , so as to successfully receive the data transmitted from SNs. In addition, the CHs also need to be within the transmission range of the neighboring CHs,  $R_{tx}^{CH}$ , so as to successfully transmit data through the multi-hop communication. Therefore, the side length of the divided small areas,  $r_{sa}$ , is given by:

$$r_{sa} \leq \min\left(\frac{R_{tx}^{SN}}{\sqrt{2}}, \frac{R_{tx}^{CH}}{\sqrt{5}}\right). \quad (3)$$

### B. CH selection

This phase selects the EHN having the highest rank from each small area as a CH, also like the EHGAF. In the proposed scheme, the amount of estimated available consuming power,  $\hat{E}_{usbl}^i(r)$ , is employed as the rank. The  $\hat{E}_{usbl}^i(r)$  is an estimation of the power being able to consume without power depletion in the current round and is given by:

$$\hat{E}_{usbl}^i(r) = E_{rst}^i(r) + \hat{E}_{hrv}^i(r), \quad (4)$$

where  $E_{rst}^i(r)$  is the amount of residual power at the EHN  $i$  at the start of the round  $r$ , and  $\hat{E}_{hrv}^i(r)$  is the amount of estimated harvesting power during the round  $r$ . In the proposed scheme, since the amount of harvesting power  $\hat{E}_{hrv}^i(r)$  is expected to be almost the same as that in the last round, the  $\hat{E}_{hrv}^i(r)$  in the current round employs the  $E_{hrv}^i(r-1)$  generated during the last round  $r-1$ .

### C. Cluster size control

This phase controls the cluster size based on the  $\hat{E}_{usbl}^i(r)$ , as shown in Fig.3, so as to share the communication load at the CHs depending on the  $\hat{E}_{usbl}^i(r)$  since the  $\hat{E}_{usbl}^i(r)$  at the selected CHs are different due to their energy-harvesting rate. For the selection of the connecting CH at SNs, the SNs generally employ the reception power level of the beacon transmitted from the surrounding CHs as the selection criteria. If the beacon transmission power at the CH,  $P_{bcn}^i$ , is lower than those at others, the CH can reduce its cluster size and decrease the number of its subordinate SNs. Therefore, the proposed scheme sets the small  $P_{bcn}^i$  for the CHs with small  $\hat{E}_{usbl}^i(r)$  and then reduces the power consumption at such CHs by reducing their communication loads due to the small cluster size.

The proposed scheme defines two levels of the beacon-transmission power  $P_{bcn}^i$ , and selects appropriate  $P_{bcn}^i$  for

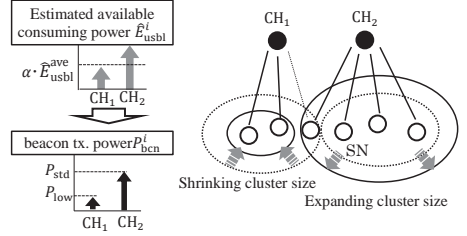


Fig.3 Cluster size control

each CH based on its  $\hat{E}_{usbl}^i(r)$ . The  $P_{bcn}^i(r)$  at the CH  $i$  in the round  $r$  is determined by threshold-based control employing the mean value  $\hat{E}_{usbl}^{ave}(r)$  over the  $\hat{E}_{usbl}^i(r)$  at all the CHs as a threshold, which is given by:

$$P_{bcn}^i(r) = \begin{cases} P_{low} & ; \hat{E}_{usbl}^i(r) < \alpha \cdot \hat{E}_{usbl}^{ave}(r) \\ P_{std} & ; \text{otherwise} \end{cases}, \quad (5)$$

where  $P_{std}$  and  $P_{low}$  are the standard beacon transmission power and that employed in the CHs intending to have small cluster size, respectively. The  $\alpha$  is a pre-defined control parameter for the  $\hat{E}_{usbl}^i(r)$ .

### D. Control procedure

As the control procedure in the proposed scheme, at the start of each round, a CH is selected from each small area during the first communication period assigned to each small area. In the CH selection, each EHN broadcasts its  $\hat{E}_{usbl}^i$  to other EHNs in its small area. The EHN receiving the  $\hat{E}_{usbl}^i$  from others compares its own  $\hat{E}_{usbl}^i$  with the received ones, and then if its  $\hat{E}_{usbl}^i$  is lower than one of the received ones, the EHN changes its operation state to sleep state until the next CH selection starts. If the EHN is still in active state at the end of the first communication period, the EHN is responsible for a CH during the current round.

After selecting the CHs at all small areas, the selected CHs transmit their  $\hat{E}_{usbl}^i$  to the SINK. The SINK calculates the mean value  $\hat{E}_{usbl}^{ave}$  over the received  $\hat{E}_{usbl}^i$ , and then broadcasts it to all the CHs. After then, the CHs select their beacon transmission power  $P_{bcn}^i$  based on the received mean value by using Eq.(5).

## IV. EVALUATION MODEL

The performance for the proposed scheme is evaluated through computer simulation. We assume that the network area is a square area with the side length of  $R$  [m], and is equally divided into  $n^2$  small square areas.  $N_{EHN}$  EHNs and  $N_{SN}$  SNs are deployed uniformly over every small square area. The CH is selected from only the EHNs, and the EHNs being out of the CH selection change their operation state to sleep state during the current round and re-charge their power through energy-harvesting function.

### A. Communication model

The network topology is two-layer clustered one, which consists of a SINK, CHs and SNs. The SNs generate data packets and transmit to the SINK through the CHs.

The communication protocol employs the IEEE802.15.4 MAC with beacon mode. In this MAC, the communications from SNs to CHs employ carrier sense multiple access/collision avoidance (CSMA/CA) during the contention access period (CAP). The dedicated CAP is assigned to each cluster (CH) without overlapping among the clusters on a time axis. Therefore, no interference between the clusters occurs. The packet loss in the communications from SNs to CHs occurs only due to packet collisions within the cluster. The hidden node problem is not taken into account. The communications from CHs to the SINK employ reservation based transmission during contention-free period (CFP), leading to loss-less communications.

The traffic model assumes network traffic load of  $G$  [packets/ $L_{\text{pkt}}$ ], which is defined as the number of the generated packets during the length of the data packets  $L_{\text{pkt}}$ . Each SN generates data packets according to a Poisson distribution with a mean traffic load  $G/(N_{\text{SN}} * n^2)$  [packets/ $L_{\text{pkt}}$ ]. The CHs receive data packets from subordinate SNs during their dedicated CAP. After receiving the data from the subordinate SNs, the CHs aggregate the received data and then transmit the aggregated data to the SINK during the following CFP. After transmitting the data, they change their operation state to sleep state until their next CAP.

### B. Power consumption model

This paper focuses on only EHNs and calculates the amount of power consumption at them. The power consumption is calculated assuming the parameters of CC2420 [6] shown in Table.1, which is one of typical communication transceivers conformable to the IEEE802.15.4 standard. The operation states for the EHNs are assumed to be the transmission state (Tx) transmitting data packets, the reception state (Rx) receiving data packets, the idle state (Idle) waiting for the data transmission or reception, and the sleep state (Sleep) turning off the wireless communication function. The amount of power consumption for the EHN  $i$  is calculated by:

$$E_{\text{Cons}}^i = \sum_{x=\text{Tx,Rx,Idle,Sleep}} P_x \cdot T_x^i, \quad (6)$$

where  $P_{\text{Tx}}$ ,  $P_{\text{Rx}}$ ,  $P_{\text{Idle}}$  and  $P_{\text{Sleep}}$  are the power consumption rate for each operation state, and  $T_{\text{Tx}}^i$ ,  $T_{\text{Rx}}^i$ ,  $T_{\text{Idle}}^i$  and  $T_{\text{Sleep}}^i$  are the time duration which the EHNs operate in each operation state, Tx, Rx, Idle or Sleep, respectively.

The EHNs equip initial battery power of  $E_{\text{ini}}$ , and the EHNs don't re-operate after depleting their battery power, even if they recover their battery power through energy-harvesting function.

### C. Energy harvesting model

This paper assumes the solar energy generation as energy-harvesting function. Each EHN constantly generates the power based on the energy-harvesting rate assigned to its belonging area.

To simulate the environments with geographical non-uniform energy-harvesting rate, two levels of the harvesting rates, high and low rates, are defined. The low harvesting rate  $E_{\text{harv}}^{\text{low}}$  is set for the area starting from one of the vertexes of the network area. The high harvesting rate  $E_{\text{harv}}^{\text{high}}$  is set for the area excepting that with the low harvesting rate.

Table.1 Parameters for power consumption

EHN state		power consumption [mW]
Transmission (Tx)	$P_{\text{Tx}}$	17.82
Reception (Rx)	$P_{\text{Rx}}$	33.84
Idle (Idle)	$P_{\text{Idle}}$	0.7668
Sleep (Sleep)	$P_{\text{Sleep}}$	0.036

Table.2 Simulation parameters

Side length of network area	$R$	200 [m]
Number of small areas	$n^2$	16
Number of EHNs in small area	$N_{\text{EHN}}$	5
Number of SNs in small area	$N_{\text{SN}}$	20
Beacon transmission interval	BI	3.93 [s]
Clustering duration		250 [BIs]
Low energy harvesting rate	$E_{\text{harv}}^{\text{low}}$	0.05 [mW]
High energy harvesting rate	$E_{\text{harv}}^{\text{high}}$	0.1 [mW]
Standard beacon transmission power	$P_{\text{std}}$	-10 [dBm]
Low beacon transmission power	$P_{\text{low}}$	-15 [dBm]
Initial battery power at EHNs	$E_{\text{ini}}$	0.5 [J]

## V. EVALUATION RESULTS

We evaluate the performance for the proposed scheme compared with the conventional one, the EHCAF, in terms of the data collection rate,  $P_{\text{DCR}}$ , and the network lifetime,  $T_{\text{life}}$ , which is defined as a time duration satisfying the  $P_{\text{DCR}}$  over 95%, when varying the control parameter  $\alpha$  for the proposed scheme, 0.5, 0.75 and 1.0.

### A. Data collection rate performance

Figure 4 shows the data collection rate  $P_{\text{DCR}}$  for the proposed and conventional schemes when an area ratio  $R_{\text{Low}}$  of the low energy harvesting rate area to entire network area is 40 [%] and network traffic load  $G$  is 0.05. The horizontal axis in Fig.4 shows the operating time elapsed from the start of the network operation. In any schemes, the data collection rate  $P_{\text{DCR}}$  keeps a high value constantly for a while after the network operation starts. However, the  $P_{\text{DCR}}$  is deteriorating slowly with the operating time. This is because that the CHs consume their power and begin to deplete their power for long operation time region. The power shortage forces some of the selected CHs to stop their operation in the middle of the round, leading to the  $P_{\text{DCR}}$  degradation. Comparing the proposed scheme with the EHCAF, the proposed one achieves higher  $P_{\text{DCR}}$  performance for the longer operation time region. This comes from the fact that the CHs locating at the low energy-harvesting area can achieve to prolong their operating by reducing their power consumption. For various values of the control parameter  $\alpha$ , the operating time achieving the high  $P_{\text{DCR}}$  gets longer with the  $\alpha$  in the proposed scheme. For the larger  $\alpha$ , more CHs can reduce their cluster size and decrease their power consumption. For this reason, these CHs can postpone the time of the power depletion. However, for the large  $\alpha$ , the packet collisions tend to occur due to the traffic increase at the CHs locating at high energy-harvesting area, and then the  $P_{\text{DCR}}$  deteriorates in the early period of the network operation.

B. Network lifetime performance

Figure 5 shows the network lifetime  $T_{life}$  performance for both schemes when varying the network traffic load  $G$ . Comparing the proposed scheme with the EHGAF, the proposed one achieves longer  $T_{life}$  performance. At the low  $G$ , the proposed scheme achieves large improvement, however, the improvement decreases with increasing  $G$ . This is because the  $P_{DCR}$  performance degrades at the clusters with expanding cluster size since the proposed scheme has more packet collisions at such clusters for large network traffic load.

Finally, Fig.6 shows the network lifetime  $T_{life}$  performance when varying the area ratio  $R_{Low}$ . Compared with the EHGAF, the proposed scheme also achieves longer  $T_{life}$  performance. At the low  $R_{Low}$ , the proposed scheme achieves large improvement, however, the improvement decreases with increasing  $R_{Low}$ . This is due to the fact that the cluster size control works less efficiently for large  $R_{Low}$  because most of the CHs locating at the low energy harvesting rate area can't change their cluster size even if they control their beacon transmission power.

VI. CONCLUSIONS

This paper has proposed the clustering control scheme, which shares communication load at CHs considering their energy-harvesting rate to prolong the network lifetime by depressing power depletion due to the non-uniform residual power arising in the conventional scheme under the geographical non-uniform energy-harvesting rate.

The proposed scheme divides the network area into some small areas, and selects one CH from each small area. Then the selected CHs control their cluster size so as to share the communication load adaptively to their residual power including the harvesting power. Thus, the proposed scheme prolongs the network lifetime since the CHs can reduce their power consumption when locating at the low energy-harvesting area.

From the performance evaluation through computer simulation, the proposed scheme can achieve the longer network lifetime with the higher data collection rate than the conventional one. For varying the area ratio of the low energy-harvesting area and the network traffic load, the proposed scheme can achieve the superior performance to the conventional one by controlling the cluster size.

For the future work, we will investigate enhanced cluster size control so as to achieve further longer network lifetime under non-uniform energy-harvesting rate conditions.

REFERENCES

[1] T. Rault, A. Bouabdallah, and Y. Challal, "Energy efficiency in wireless sensor networks: A top-down survey," *Computer Networks*, vol.67, no.4, pp.104-122, July, 2014.  
 [2] C. Bergonzini, D. Brunelli, L. Benini, "Algorithms for harvested energy prediction in batteryless wireless sensor networks," *Proc. of 3<sup>rd</sup> International Workshop on Advances in Sensors and Interfaces*, pp. 144-149, June, 2009.  
 [3] V. Raghunathan, A. Kansal, J. Hsu, J. Friedman, M. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems," *Proc. of Fourth International Symposium on Information Processing in Sensor Networks*, pp. 457-462, April, 2005.

[4] Yuta Ijiri, Yuichi Miyaji, Hideyuki Uehara, "Cluster-Based Routing Protocol Considering Unevenly Distributed Energy Harvesting Rate in Wireless Sensor Network," *Tokai-Section Joint Conference on Electrical, Electronics, Information, and Related Engineering, L1-1*, Sept. 2015. (In Japanese).  
 [5] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," *Proc. of MobiCom'01*, pp.70-84, July, 2001.  
 [6] Texas Instrument, RF transceiver CC2420 Data Sheet, vol.1, April, 2006.

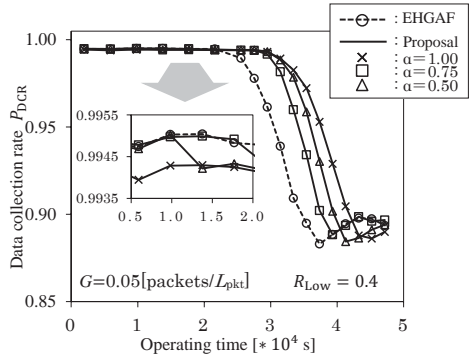


Fig.4 data collection rate  $P_{DCR}$

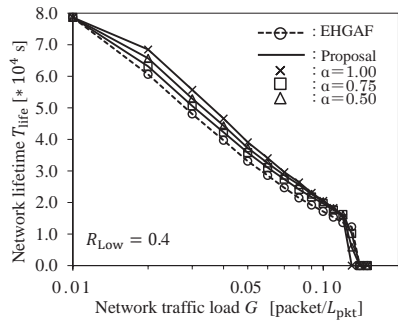


Fig.5 Network lifetime when varying traffic load  $G$

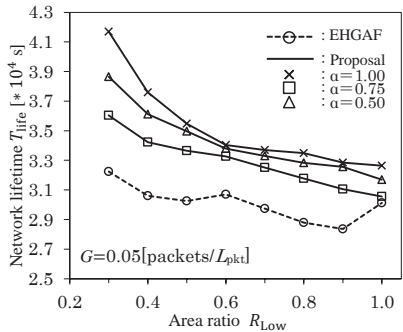


Fig.6 Network lifetime when varying area ratio  $R_{Low}$

# Denoising Images by Variational Mode Decomposition and Spectral Subtraction

Gavriloaia Bogdan, Vizireanu Radu, Fratu Octavian  
Telecommunications Dept.  
University Politehnica of Bucharest  
Bucharest, Romania  
bogdan.gavriloaia@gmail.com

Barca Cristian, Mara Constantin, Gheorghe Gavriloaia  
Electronics, Communications and Computers  
University of Pitesti:  
Pitesti, Romania

**Abstract**— Medical imaging has greatly improved the quality of medical investigations enabling modern medicine objectively assess the health of tissues and organs. Resonance magnetic imaging MRI provides a lot of data by using parallel processing techniques of signals acquired from many coils. Parallel processing of a large number of channel changes noise statistics and creates great difficulties in applying various methods of denoise. An algorithm of processing in two steps is described in the paper. In the first stage the signal is decomposed into several frequency spectra using a variational mode decomposition method. The signal from the frequency band with the largest frequencies is filtered in the second stage by using spectral subtraction method. The proposed algorithm ensures an appropriate tradeoff between spatial resolution and signal-to-noise ratio. The efficiency of the method is demonstrated through an example of reduction of noise level and increase the contrast of the pelvic MRI image.

**Keywords**—denoising; medical imaging; variational mode decomposition; spectral subtraction; magnetic resonance imaging;

## I. INTRODUCTION

One of the key differences between traditional and modern medicine refers to methods of treating patients. In the first case, the symptoms described by the patient or observed by physician are used. They have a high degree of subjectivity.

In the second case, anamnesis is accompanied by medical imaging, a technique mainly non-invasive, which can provide objective picture of the tissues or organs inside the patient rather than from symptoms. Medical diagnosis has become faster and more accurate. More than, medical imaging has used for diagnosis or as treatment planning, verification of administered treatment, surgical simulation, intra-operative navigation, radiotherapy planning, tracking the progress of disease, etc.

The energetic support of medical information can be sound field (ultrasound investigations), nuclear radiation (computed tomography), electromagnetic radiation (magnetic resonance imaging, MRI), thermal radiation (thermal imager), etc. These medical exams provide a large amount of data [1].

Tissues and organs investigated by MRI have low contrast and are affected by noise during the acquisition process. As a result, visual interpretation and algorithms used for segmentation and clustering will be strongly influenced [1, 2].

Attention of researchers, doctors and physicists, was focused on developing MRI technology, as a result of facilities offered by this medical investigation: high spatial resolution, acceptable signal-to-noise, and high speed acquisition. This technology presents a great disadvantage to the patient - the time of investigation is quite large, tens of minutes. Also, it is not possible to achieve simultaneous maximum values of these parameters. The trade-off should be achieved by clinical, scientific and financial factors [3-5].

The signal to noise ratio decreasing has a negative effect on post processing MR imaging: tissue classification and organ segmentation. Denoising images such as denoised output should be in the same dimensionality as the noisy input is an area of great interest because it is an important step for increasing image quality and performance of the tasks needed for imaging analysis. In literature have been proposed several methods, variational or statistical, but until now it is not possible to determine which method is most convenient for noise reducing and intensity bias, structure and edge preservation, automation and computational cost.

Partial differential equations are used to variational methods. The structures are simulated with different local simple image patterns. The bias introduced by Rician noise and manual tuning of some critical free parameters reduce extensively using of these methods [6, 7].

Buades has made the summary of the genuine denoising algorithms based on noise model [8]: Gaussian smoothing model [9] (where the smoothness is measured by the Dirichlet integral); anisotropic filtering model [5]; Rudin-Osher-Fatemi, ROF [10], total variation model and two recently proposed iterated total variation refinements [11, 12]; use of  $d\hat{f}$  and  $dct$  in a running window, neighborhood filters [13]; Wiener local empirical filter; translation invariant wavelet thresholding; unsupervised information theoretic, adaptive filtering; non local means (NL-means) algorithm, spectral subtraction (SS) [14, 15].

ROF method is based on the assumption that the initial image is composed by simple geometric description objects, smooth inside but with jumps across the boundaries. The functional space modeling these properties has a finite total variation. The recovered original image is a solution of the constrained minimization problem.

However, the ROF model favors piecewise constant solutions, causing unsatisfying “staircasing effects” [16]. Therefore, the ROF model has been improved by using different functional spaces. So, Dragomiretskiy has introduced the intrinsic mode functions in the frame of Variational Mode Decomposition. Actually, IMFs had been introduced before by Huang at the Empirical Mode Decomposition method, a method with limited mathematical theory [8].

In this study, we introduce an improved image denoising method by applying the spectral subtraction of a certain component of MRI acquisitions in  $k$ -space. This component is the highest oscillation mode obtained by applying the variational mode decomposition (VMD) method. It contains the greatest amount of noise.

The rest of the paper is organized as follows. Section II presents the proposed method, and discusses algorithm considerations, Section III shows simulation and experimental results illustrating denoising performance of the proposed algorithm, followed by the conclusions of the paper in Section IV.

## II. MATERIAL AND METHODS

Signals from real life are composed of many different oscillations (or modes), with complex waveforms and time-varying amplitudes and frequencies. They can be described by amplitude modulated-frequency-modulated (AM-FM) signals:

$$u_k(t) = A_k(t) \cos(\phi_k(t)) \quad (1)$$

where the phase  $\phi_k(t)$  is a non-decreasing function and the envelope  $A_k(t)$  is non-negative [9, 17]. The envelope and the instantaneous frequency vary much slower than the signal phase. Their properties can be used to assess health from signals generated by the human body.

The image from the output sensor,  $f$ , and can be recovered/denoised has two components,  $s=u+v$ , where  $u$  is the smooth component and  $v$  is noise, the remaining term (the case for the additive noise removal). The solution of the functional energy minimization:

$$F(u) = \frac{1}{2} \|s - u\|^2 + \lambda TV(u) \quad (2)$$

is unique [18].  $TV(u)$  is for total variation of  $u$  on the bounded variation space,  $BV$ , and  $\lambda$  is a regularization parameter. Because of the staircasing effects, different functional spaces have been used. Dragomiretskiy [19] has proposed to split  $BV$  into several frequency bans or modes like those described by (1) with  $\omega_k(t)$  as mean frequencies and  $BW$  as frequency band,  $BW=2(\Delta f + f_{FM} + f_{AM})$ , where  $\Delta f$  represents the maximum deviation frequency of the instantaneous frequency,  $f_{FM}$  corresponds to the rate of change, and  $f_{AM}$  is the highest frequency of the signal envelope from (1). The bandwidth of a mode is estimated through the H1 Gaussian smoothness of the demodulated signal.

In [19] are described the algorithms for  $u_k$  and  $\omega_k$  optimizing as a recursive process:

$$\hat{u}_k^{n+1}(\omega) = \frac{1}{1 + 2\alpha(\omega - \omega_k)} \left( \hat{s}(\omega) - \sum \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right)$$

$$\hat{\omega}_k^{n+1}(\omega) = \arg \min_{\omega_k} \{ \alpha \|j(\omega - \omega_k) [(1 + \text{sign}(\omega - \omega_k)) u_k(\omega)]\|^2 \}$$

As a result, the input signal is decomposed into a certain number of modes, each of them having limited bandwidth in the spectral domain. The process of mode bandwidth evaluation takes place in 3 steps: the evaluation of the unilateral frequency spectrum by using Hilbert transforms of the analytic signal; shift the mode's frequency spectrum to baseband; H1 smoothness (Dirichlet energy) of the demodulated signal. The algorithm and Matlab program are presented in [20].

Image reconstruction in MRI uses the signals acquired from many coils, parallel imaging techniques. The applicability of conventional denoising techniques is limited because there are different noise statistics that can produce edge and blurring artifacts [21-24]. The spectral subtraction method does not assume a model for the true signal. An adaptive estimation of noise is done at the beginning of acquisition process in  $k$ -space. The rows of  $k$ -space matrix are filled by the acquired signal  $s(t)$ :

$$s_r(t) + is_i(t) = u_r(t) + n_r(t) + iu_i(t) + in_i(t) \quad (3)$$

We will assume an additive noise model whereby the measured signal is practically the sum of a deterministic component  $u(t)$ , in addition to an independent random noise  $n(t)$ . Subscripts  $r$  and  $i$  denote real and imaginary components, respectively. Because in MRI technique signal and noise are uncorrelated, the power spectral density, PSD, of real component  $|S_r(f)|^2$  is given by the PSD of the noiseless signal  $|U_r(f)|^2$  and the PSD of the noise  $|N_r(f)|^2$  of the same real components:

$$|S_r(f)|^2 = |U_r(f)|^2 + |N_r(f)|^2 \quad (4)$$

PSD of noise is constant. By using a predefined weighting function of  $f$ ,  $Q(f)$  with real values in the interval  $[0, 1]$ , PSD of the output signal is:

$$|Y_r(f)|^2 = |U_r(f)|^2 - Q(f) |N_r(f)|^2 \quad (5)$$

Denoising of real component is obtained by taking inverse Fourier transform of (5):

$$Y_r(f) = |Y_r(f)| \exp(j \text{Phase } S_r(f)) \quad (6)$$



The same algorithm is applied for the imaginary parts of signals, for each row and line of  $k$ -matrix. Finally, the average of both denoised images is taken [25].

### III. MRI DENOISING

The algorithm described in Section II was implemented in order to investigate MR images. In Fig. 1 is depicted the axial section of pelvic area of human body having 384x300 pixels, and it was taken with MAGNETOM Skyra 3T MRI, Siemens. A region of interest ROI, 120x100 pixels, was selected in the middle of image and a central Line.

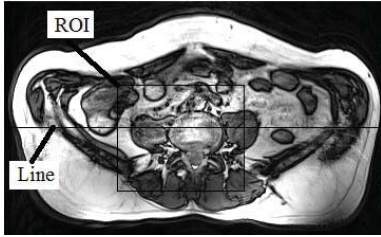


Fig. 1. MRI of pelvic area, axial section

The amplitude variation along Line is shown in Fig. 2 and a certain value of noise can be noticed. This signal was decomposed into five IMFs by using Variational Mode Decomposition method. Only three modes or IMFs, M1, M2 and M5 are shown in Fig 3, for a better view of their harmonic and spatial evolution.

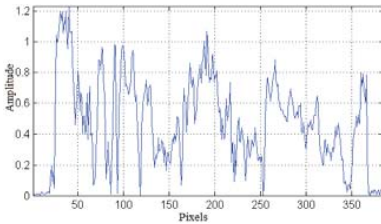


Fig. 2. Signal of central line from Fig. 1

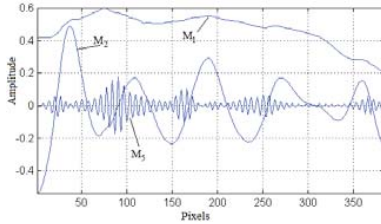


Fig. 3. Three IMFs of the central Line from Fig 1

The difference between the input signal and the all five IMFs represents the component having the most important contribution concerning to noise and spatial resolution. It is depicted in Fig. 4. Influence of noise on the components with low spatial frequencies is less significant and does not affect the processing of the image. Therefore, our investigation was focused only on the difference signal and not on the entire signal acquired from MRI.

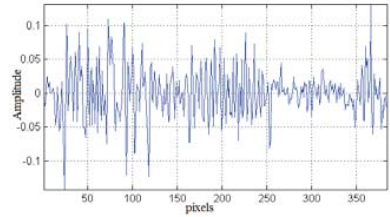


Fig. 4. Difference between the acquired signal and the five IMFs

Fast Fourier Transform was applied on the difference signal and power spectral density PSD is plotted in Fig. 5. The average noise power added to  $k$ -space to corrupt the image was chosen as average PSD values between pixels 160 and 180 from Fig. 5. Computer simulations were iterated 50 times for  $Q(f)$  reduction factor scaled to a range of [0, 1]. After applying (6), SNR values were extracted from the difference signal mean and standard deviation (SD):

$$SNR(x) = \frac{\text{mean}[Y_i(x)]_i}{SD[Y_i(x)]_i} \quad (7)$$

where  $Y_i(x)$  is the magnitude of the pixel at location  $(x)$  in the Line at the  $i$ th iteration [27]. The maximum SNR value of 8.27 was obtained for  $Q(f)=0.36$ .

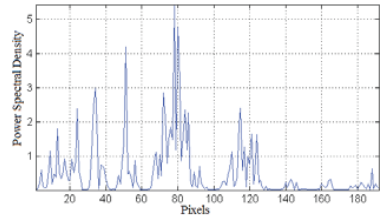


Fig. 5. PSD of the signal from Fig 4

In order to get the denoised Line signal, (5) and (6) were evaluated. The effect of denoising can be better noticed by analyzing of noisy and denoised ROI, Fig 1 and 6, respectively. Fig. 6 depicts a clearer ROI with a higher contrast. From mathematical point of view, the new area has a SNR of 26.4, instead of 15.7, improved factor  $R=1.68$ .

Different numbers of IMFs were tested, but when their number is greater than eight are not noticed significant improvements for SNR and image quality.

The improved factor  $R=1.42$  was realized by using only Variational Mode Decomposition with eight IMFs, and  $R=1.23$  in the case of spectral subtraction.

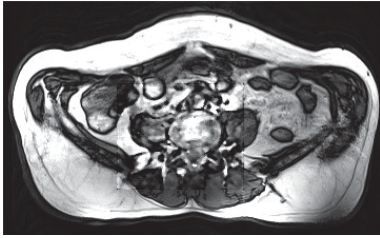


Fig. 6. MRI of pelvic area with ROI after denoisingMRI

#### IV. CONCLUSIONS

An improved algorithm for MR images has been proposed. It uses two denoising methods. The signal acquired directly from MRI is decomposed into a certain number of modes or intrinsic mode functions by using Variational Mode Decomposition method. In the second step, the difference between the acquired signal and the all modes added together is supposed to contain an important contribution to the image resolution, and it is filtered by spectral subtraction method. The method efficiency is argued by comparative evaluating of the improved factor for different methods

#### REFERENCES

##### ACKNOWLEDGMENT

The work has been funded by the EC through the FP7-IP project eWALL no. 610658, and by UEFISCDI through the FP7 eWALL support project no. 262/EU.

#### REFERENCES

[1] J. Peng, F.Dong, D. Kong, "Recent advances of variational model in medical imaging and applications to computer aided surgery" *Appl. Math. J. Chinese Univ.* 2012, 27(4): 379-411

[2] R. Eslami and M. Jacob, "Reduction of distortions in MRSI using a new signal model," *Proc. IEEE Intl. Symposium Biomed. Imag.*, pp. 438-441, 2009.

[3] J. Manjon, P. Coupe, L. Marti-Bonmati, D. Collins, M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels", *J Magn Resonance Imag.* Jan; 31(1): pp. 192-203, 2010.

[4] G. Gavrioloia, M-R Gavrioloia, "Acoustic resonant response of smooth or rough 3D small objects" *Journal Measurement*, Volume 45, Issue 3, pp. 627-630, 2012

[5] P. Perona, J. Malik, "Scale-space and edge-detection using anisotropic diffusion", *IEEE Trans Pattern Anal Mach Intell.* Jul; 12(7): pp. 629-639, 1990.

[6] Y. Kadah, "Adaptive denoising of event-related functional magnetic resonance imaging data using spectral subtraction", *IEEE Trans Biomed Eng.* Nov; 51(11):1944-1953, 2004.

[7] M. Erturk, P. Bottomley, A. El-Sharkawy, "Spectral subtraction denoising of MRI", *Proc. 6th Cairo Int. Biomed. Eng. Conf*; Cairo, Egypt. pp. 138-141, 2012.

[8] A. Buades, B. Coll, J. Morel, "A non-local algorithm for image denoising", *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*; 2, pp. 60-65, 2005.

[9] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," *Proc. IEEE Intl. Symposium Biomed. Imag.*, pp. 988-991, 2007.

[10] H. M. Nguyen, X. Peng, M. N. Do, and Z.-P. Liang, "Spatiotemporal filtering of MR spectroscopic imaging data by low-rank approximations," in *Proc. IEEE Intl. Symposium Biomed. Imag.*, 2011, pp. 857 - 860.

[11] X. Peng, H. Nguyen, J. Haldar, D. Hernando, X.-P. Wang, and Z.-P. Liang, "Correction of field inhomogeneity effects in limited k-space coverage MRSI data with anatomical constraint," *Proc. Intl. Soc. Mag. Reson. Med.*, pp. 883-886, 2010.

[12] G. Gavrioloia, M. Gherghe, M. Gavrioloia, "Anisotropic diffusion and wavelet filtering of ultrasound images" *Ultrasound in Medicine & Biology*, Volume 37, Issue 8, pp 151-154, 2011

[13] A. Bashir and D. A. Yablonskiy, "Natural linewidth chemical shift imaging (NL-CSI)," *Mag. Res. Imag.*, pp. 7-18, 2006.

[14] P. B. Barker and D. D. M. Lin, "In vivo proton MR spectroscopy of the human brain," *Prog. NMR Spect.*, vol. 49, pp. 99-128, 2006.

[15] U. Seeger and U. Klöse, "Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases," *Magn. Reson. Med.*, vol. 49, pp. 19-28, 2003.

[16] B. J. Soher, K. Young, A. Bernstein, Z. Aygula, and A. A. Maudsley, "GAVA: Spectral simulation for in vivo MRS applications," *J. Magn. Reson.*, vol. 185, pp. 291-299, 2007.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, pp. 2080-2095, 2007.

[18] W. Litvinov, T. ahman, X. Tai, "A modified TV-stokes model for image processing", *SIAM J Sci Comput*, 2011, 33(4): 1574-1597

[19] K Dragomiretskiy, D Zosso "Variational mode decomposition", - *Signal Processing*, IEEE Transactions on, 2014, Volume 62, Issue 3, pp. 531-544,

[20] J. Lellmann, C. Schnor, "Continuous multiclass labeling approaches and algorithms", *SIAM J Imaging Sci*, 2011, 4(4): 1049-1096.

[21] F Lecellier, J Fadili, S Jehan-Besson, GAubert, MRévenu, ESaloux. Region-based active contours with exponential family observations, *J Math Imaging Vision*, 2010, 36: 28-45.

[22] G. Gavrioloia, S. Halunga, R. Narita, "Coarse/acquisition GPS codes correlation properties and vulnerability to noise", *Conference, Telecommunications in Modern Satellite, Cable and Broadcasting Services, TELSIKS 2007, 8th International Conference on*, pp. 554-557, 2007

[23] Y. Wang and S.-J. Li, "Differentiation of metabolic concentrations between gray matter and white matter of human brain by in vivo 1H magnetic resonance spectroscopy," *Magn. Reson. Med.*, vol. 39, pp. 28-33, 1998.

[24] C. Choi, N. J. Coupland, P. P. Bhardwaj, S. Kalra, C. A. Casault, K. Reid, and P. S. Allen, "T2 measurement and quantification of glutamate in human brain in vivo," *Magn. Reson. Med.*, vol. 56, pp. 971-977, 2006.

[25] J. P. Haldar, T.-H. Wu, Q. Wang, C.-I. Chen, S.-K. Song, and Z.-P. Liang, "Further development in anatomically constrained MR image reconstruction: application to multimodal imaging of mouse stroke," *Proc. Intl. IEEE Eng. Med. Biol. Soc.*, pp. 422-425, 2008.

[26] HL Zhang, YMChen, J L Shi. "Nonparametric image segmentation using Renyi's statistical dependence measure", *J Math Imaging Vision*, 2012, 9: 1-11.

[27] S. Vaseghi, "Spectral subtraction", *Advanced Digital Signal Processing and Noise Reduction*, Second Edition, John Wiley & Sons Ltd, 84-86, 2000.

# Trend Predictive Model of Cardiovascular Complication for Type 2 Diabetes Mellitus with Hypertension Patients

Napa Rachata and Punnarumol Temdee

School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand  
E-mail: napa.rachata@lamduan.mfu.ac.th and punnarumol@mfu.ac.th

**Abstract**—Cardiovascular complication is the significant cause of dead for type 2 diabetes mellitus with hypertension patients. Unfortunately, there is no existing trend predictive model for cardiovascular complication prevention for type 2 diabetes mellitus with hypertension patients nowadays. For these reasons, this paper proposes a trend predictive model for preventing type 2 diabetes mellitus with hypertension patients from cardiovascular complication. The proposed model can demonstrate the evolution of patients' health status through the detected trends including positive and negative changing and not changing. Consequently, the patients can manage themselves appropriately throughout their healthcare journey. The lifestyle factors are used for modeling together with the clinical factors because of the changing of lifestyles will affect clinical factors and significantly impact the quality of life of the patients. More specifically, 15 factors are used for modeling accordingly to general fact and implicit knowledge of the human experts by using fuzzy logic. The simulation result shows that the constructed model provides 93% accuracy when being compared with those decided by the experts.

**Keywords**—*cardiovascular complication; fuzzy logic; hypertension; trend prediction; type 2 diabetes mellitus*

## I. INTRODUCTION

Diabetes mellitus, or diabetes, is a main type of chronic disease and is a deadly health problem worldwide. It has three main types including type 1 diabetes mellitus, type 2 diabetes mellitus, and gestational diabetes mellitus. Especially, the type 2 diabetes is the majority of people with diabetes around the world. The global prevalence of it has been rising each year and especially in the middle and low income countries. About 1.5 million dead people were attributable to diabetes in 2012 [1] and there were growing numbers of diabetes patients from 108 million people in 1980 to 422 million people in 2014 [1]. Additionally, it will be accountable for the 7<sup>th</sup> major cause of dead people in 2030 with World Health Organization (WHO) projects [2].

Hypertension is also a chronic disease and crisis public health issue in the world. Most people with hypertension do not know that they live with it because it rarely has symptoms. The consequence of hypertension causes more than 9 million deaths every year and one of three adult people normally has high blood pressure [3]. Moreover, the 40 percentages of people aging from 24 years old had high blood pressure in 2008 and

the 22 percentages of people aging from 18 years old had high blood pressure in 2014 [3].

The diabetes patients have a two-fold chance to have high blood pressure and have more chance to have serious complications. The main cause of dead diabetes patients are cardiovascular complication [4]. As a result, the dead patients with cardiovascular complication have been attributed to hypertension due to the type 2 diabetes mellitus by pushing the rate of cardiovascular faster [4, 5]. The hypertension is the key cause of cardiovascular diseases because half of all hypertension people die because of cardiovascular diseases [6]. Consequently, type 2 diabetes mellitus with hypertension patients will extremely have a higher cardiovascular complication risk.

These evidences represent that cardiovascular complication is the important role of the patients with type 2 diabetes mellitus and hypertension. For these reasons, this paper mainly focuses on preventing cardiovascular complication for type 2 diabetes mellitus with hypertension.

The paper is organized as follow. Section II explains the literature reviews. Section III proposes the research methodology. Section IV presents the simulation detail, results and discussion. Section V concludes the paper.

## II. LITERATURE REVIEWS

The predictive models for preventing diseases have been proposed widely including cardiovascular diseases problem [7-25]. These research works are mainly proposed for preventing people who do not have evidence of established cardiovascular diseases [7-18]. Moreover, there is less research work recently has been proposed for preventing patients from cardiovascular complication which focus on patients with type 2 diabetes mellitus in [19-24] and patients with type 2 diabetes mellitus and hypertension in [25] although there are raising numbers of patient over the world. However, there is no existing trend predictive model for cardiovascular complication prevention for type 2 diabetes mellitus with hypertension patients. Accordingly, this paper emphasizes on trend preventing cardiovascular complication for type 2 diabetes mellitus with hypertension patients. Moreover, the existing research works with cardiovascular prevention have focused on clinical data more than lifestyle

data for prediction. Based on the fact that the chronic diseases have a slow progression and take long treatment, so the lifestyle should also be focused together with the clinical factors because the changed lifestyle can affect the changing of clinical data and the quality of life of the patients. Therefore, the lifestyle factor and the clinical factor are used in this paper accordingly to general fact knowledge and implicit knowledge of the human experts. Consequently, instead of detecting general condition for risk prediction, this paper focuses on trend prediction which is affected from lifestyle changing of the patients.

All among existing predictive models, several methods have been developed previously to the prevention of cardiovascular diseases; nevertheless most of these methods have been used for dataset that cannot effectively deal with natural representation of implicit knowledge from human experts that generally uses uncertain terms [26]. Consequently, fuzzy logic is employed for the most recent works. Fuzzy logic is well known in representing expert knowledge [26] and dealing with dynamic behavior problem [27]. There are many research works using fuzzy logic in various areas including disease prediction [24], [26-31]. Hence, this paper uses fuzzy logic as the modeling method for modeling the implicit knowledge from human experts. Finally, the expected trend predictive model will be valuable for self-long healthcare so that the patients are able to have a better quality of life.

III. RESEARCH METHODOLOGY

The proposed research methodology is determined with three processes which are data gathering, model construction, and model validation as shown in Fig. 1.

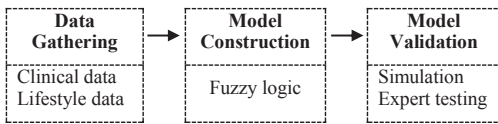


Fig. 1. Research methodology of proposed trend predictive model

A. Data Gathering

This first process is to gather clinical data and lifestyle data which are used to model construction. These data are gathered from Framingham [7] and implicit knowledge from human experts. The clinical data is Sex, Age, Body Mass Index (BMI), Total Cholesterol (TC), HDL Cholesterol, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Low Density Lipoprotein (LDL), HbA1C, Fasting Plasma Glucose (FPG), and Diabetes. Another data is lifestyle which includes Smoking behavior, Compliance, Basal Metabolic Rate (BMR), and Physical activity. All data are risk factors-related to cardiovascular complication for type 2 diabetes mellitus with hypertension patients.

B. Model Construction

This second process is to construct model based on Framingham [7] and implicit knowledge from human experts with using fuzzy logic. The model construction process consists of four main steps including fuzzification, fuzzy rule evaluation, aggregation, and defuzzification as shown in Fig. 2.

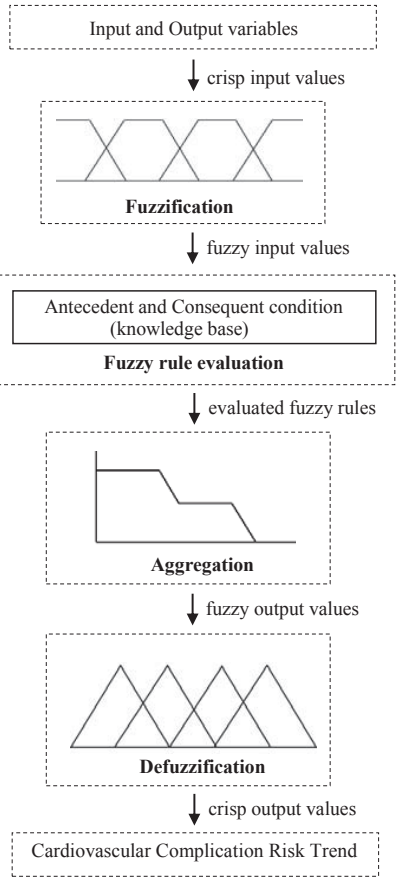


Fig. 2. Step of model construction process

To begin with, the input variables are defined from Framingham [7] and implicit knowledge from human experts which are 15 risk factors as mentioned before. One output variable is trend of cardiovascular complication risk including positive and negative changing and not changing trend of each cardiovascular complication risk as given in Table I.

TABLE I. TREND OF CARDIOVASCULAR COMPLICATION RISK

No.	Cardiovascular complication risk	Trend	Abbreviations
1	Very Low	Positive Not changing Negative Not changing Negative Changing	VL-PN VL-NN VL-NC
2	Low	Positive Not changing Negative Not changing Positive Changing Negative Changing	L-PN L-NN L-PC L-NC

No.	Cardiovascular complication risk	Trend	Abbreviations
3	Moderate	Positive Not changing	M-PN
		Negative Not changing	M-NN
		Positive Changing	M-PC
		Negative Changing	M-NC
4	High	Positive Not changing	H-PN
		Negative Not changing	H-NN
		Positive Changing	H-PC

The first step, these input and output variables are determined to fuzzy sets and membership functions. The fuzzy sets with their range values of input and output variables are represented in Table II.

TABLE II. FUZZY SETS OF INPUT AND OUTPUT VARIABLES

No.	Variables	Range values	Fuzzy sets		
1	Sex	0.5 1	Male Female		
2	Age (Male)	30 – 34	Range 1		
		35 – 39	Range 2		
		40 – 44	Range 3		
		45 – 49	Range 4		
		50 – 54	Range 5		
		55 – 59	Range 6		
		60 – 64	Range 7		
		65 – 69	Range 8		
		70 – 74	Range 9		
		≥ 75	Range 10		
2	Age (Female)	30 – 34	Range 1		
		35 – 39	Range 2		
		40 – 44	Range 3		
		45 – 49	Range 4		
		50 – 54	Range 5		
		55 – 59	Range 6		
		60 – 74	Range 7		
		≥ 75	Range 8		
		3	Body Mass Index (BMI) (Male / Female)	< 18.5	Underweight
				18.5 – 24.9	Healthy
25.0 – 29.9	Overweight				
> 29.9	Obese				
4	Total Cholesterol (TC) (Male / Female)	< 160	Protected		
		160 – 199	Normal		
		200 – 239	High		
		240 – 279	Very High		
		≥ 280	Extremely High		
5	HDL Cholesterol (Male / Female)	≥ 60	Protected		
		45 – 59	Normal		
		35 – 44	Very High		
6-7	Systolic Blood Pressure / Diastolic Blood Pressure (Male)	≤ 129 and ≤ 84	Normal		
		≤ 139 and ≤ 89	High		
		≤ 159 and ≤ 99	Very High		
		≤ 200 and ≤ 150	Extremely High		
6-7	Systolic Blood Pressure / Diastolic Blood Pressure (Female)	< 120 and < 80	Protected		
		≤ 139 and ≤ 89	Normal		
		≤ 159 and ≤ 99	Very High		
		≥ 200 and ≤ 150	Extremely High		
8	Low Density Lipoprotein (LDL) (Male / Female)	≤ 100	Protected		
		101 – 140	Normal		
		> 140	High		
9	HbA1C (Male / Female)	≤ 6.5	Protected		
		6.6 – 7.0	Normal		
		> 7.0	High		
10	Fasting Plasma Glucose (FPG) (Male / Female)	≤ 60	Very Low		
		61 – 80	Low		
		81 – 130	Protected		

No.	Variables	Range values	Fuzzy sets		
		131 – 140	High		
		> 140	Very High		
11	Diabetes (Male)	0.3	Normal		
		0.6	Very High		
11	Diabetes (Female)	0.3	Normal		
		1	Extremely High		
12	Smoking (Male / Female)	0.5 1	Normal Very High		
13	Compliance (Male / Female)	0.5 1	Regular Irregular		
14	Basal Metabolic Rate (BMR) (Male / Female)	0.3	Under		
		0.6 1	Normal Over		
15	Physical activity (Male / Female)	120 – 150	Protected		
		20 – 119 0 – 19	Low Sedentary		
16	Cardiovascular complication risk trend (Male)	0 – 8	VL-PN		
		4 – 10	VL-NN / VL-NC		
		8 – 13	L-PN / L-PC		
		10 – 16	L-NN / L-NC		
		13 – 20	M-PN / M-PC		
		16 – 25	M-NN / M-NC		
		20 – 45	H-PN / H-PC		
		25 – 100	H-NN		
		16	Cardiovascular complication risk trend (Female)	0 – 8	VL-PN
				4 – 10	VL-NN / VL-NC
8 – 13	L-PN / L-PC				
10 – 15	L-NN / L-NC				
13 – 18	M-PN / M-PC				
15 – 24	M-NN / M-NC				
20 – 27	H-PN / H-PC				
24 – 100	H-NN				

Moreover, the range values of each fuzzy set are designed to the membership functions. The input variables are designed with trapezoidal membership function. The trapezoidal membership function is presented by the following (1).

$$\mu_A(x) = \begin{cases} 0 & x \leq a \\ (x-a)/(b-a) & a < x < b \\ 1 & b \leq x \leq c \\ (d-x)/(d-c) & c < x < d \\ 0 & x > d \end{cases} \quad (1)$$

where  $A$  is the fuzzy set,  $x$  is member of fuzzy set,  $\mu_A$  is the membership function of fuzzy set, and  $a, b, c$  and  $d$  is parameter.

Another variable is designed with triangular membership function which can be calculated using the following (2).

$$\mu_A(x) = \begin{cases} 0 & x \leq a \\ (x-a)/(b-a) & a < x < b \\ 1 & x = b \\ (c-x)/(c-b) & b < x < c \\ 0 & x \geq c \end{cases} \quad (2)$$

where  $A$  is the fuzzy set,  $x$  is member of fuzzy set,  $\mu_A$  is the membership function of fuzzy set, and  $a, b$  and  $c$  is parameter.

Therefore, degree of membership of each input and output variables are produced according to the membership functions.

The secondly step is fuzzy rule evaluation determination which is based on the gathered general fact knowledge and implicit knowledge from the experts for getting the quality of antecedent and consequent condition. Table III shows some of fuzzy rules which are used in evaluation step.

TABLE III. SOME OF FUZZY RULES OF TREND OF CARDIOVASCULAR RISK

Rule	Description
1	If (Sex is Male) and (Age is Range 1) and (BMI is Overweight) and (TC is Normal) and (HDL is Normal) and (SBP is High) and (DBP is High) and (LDL is High) and (HbA1C is High) and (FPG is Very Low) and (Diabetes is Very High) and (Smoking is Very High) and (Compliance is Regular) and (BMR is Over) and (Physical activity is Low) <b>then</b> (Cardiovascular complication risk trend is VL-NN)
2	If (Sex is Male) and (Age is Range 7) and (BMI is Healthy) and (TC is Normal) and (HDL is Normal) and (SBP is High) and (DBP is High) and (LDL is High) and (HbA1C is High) and (FPG is High) and (Diabetes is Very High) and (Smoking is Normal) and (Compliance is Regular) and (BMR is Normal) and (Physical activity is Protected) <b>then</b> (Cardiovascular complication risk trend is L-PN)
3	If (Sex is Female) and (Age is Range 5) and (BMI is Obese) and (TC is High) and (HDL is Very High) and (SBP is High) and (DBP is High) and (LDL is High) and (HbA1C is High) and (FPG is High) and (Diabetes is Very High) and (Smoking is Normal) and (Compliance is Regular) and (BMR is Over) and (Physical activity is Low) <b>then</b> (Cardiovascular complication risk trend is M-NC)
4	If (Sex is Female) and (Age is Range 8) and (BMI is Overweight) and (TC is Very High) and (HDL is Very High) and (SBP is Very High) and (DBP is Very High) and (LDL is High) and (HbA1C is High) and (FPG is High) and (Diabetes is Very High) and (Smoking is Very High) and (Compliance is Regular) and (BMR is Over) and (Physical activity is Sedentary) <b>then</b> (Cardiovascular complication risk trend is H-NN)

After that, the evaluated fuzzy rules are aggregated according to fuzzy input values by using union operation. Therefore, this step produces membership functions of fuzzy rule-related trend of cardiovascular complication risk. The last step is the defuzzification step which is converting the result of fuzzy rules aggregation from a fuzzy output value into a crisp output value that can be used practically by using the centre of gravity (COG) technique. The COG defuzzification can be calculated using the following (3).

$$COG = \frac{\sum u(x)x}{\sum (x)} \quad (3)$$

where  $A$  is the fuzzy set,  $x$  is member of fuzzy set,  $\mu_A$  is the membership function of fuzzy set, and  $a$  and  $b$  is interval.

Then, the result of calculation produces the trend of cardiovascular complication risk for patients with type 2 diabetes mellitus with hypertension. Moreover, the mandani fuzzy inference system is used for formulating input/output mappings in this paper.

C. Model Validation

This process consists of simulation cases and decision from experts for accuracy testing with the model construction. Firstly, the model construction is validated with conducted

with 60 simulated cases for testing the proposed modeling. On the other hand, the constructed model is tested by comparing simulated results with experts. The result of model validation will be mentioned in the simulation results and discussion session.

IV. SIMULATION RESULTS AND DISCUSSION

The results of model validation are determined by simulation cases and decision with experts. Each of the case is checked with the condition simulation for detecting that the chances of not changing and changing of cardiovascular complication risk trend.

A. Simulation Results

The 60 simulated cases are used for testing the proposed predictive model. The examples of simulated cases are shown in Table IV.

TABLE IV. SOME OF SIMULATION RESULTS

Cases	Times	Results	Cardiovascular complication risk	Trend
1 (Male)	0	5.09	Very Low	Negative
	1	7.33	Very Low	Not changing
2 (Male)	0	6.54	Very Low	Positive
	1	5.15	Very Low	Not changing
3 (Male)	0	7.33	Very Low	Negative
	1	10.3	Low	Changing
4 (Male)	0	10.3	Low	Positive
	1	7.33	Very Low	
5 (Male)	0	18.4	Moderate	Negative
	1	18.7	Moderate	Not changing
6 (Male)	0	18.7	Moderate	Negative
	1	23.4	High	Changing
7 (Female)	0	7.33	Very Low	Positive
	1	5.15	Very Low	Not changing
8 (Female)	0	14.6	Low	Negative
	1	15.33	Moderate	Changing
9 (Female)	0	17.33	Moderate	Negative
	1	17.45	Moderate	Not changing
10 (Female)	0	19.6	Moderate	Negative
	1	22.2	High	Changing
11 (Female)	0	24.83	High	Negative
	1	25.7	High	Not changing
12 (Female)	0	27.4	High	Positive
	1	25.33	High	Not changing

The result of simulation with the 60 simulated cases gives 91.67% correct prediction that is 55 simulated cases and 8.33% incorrect prediction that is 5 simulated cases respectively.

B. Expert Testing Results

The model construction is tested by comparing the simulated results with experts. The expert testing results show that the constructed predictive model provides 93% accuracy when being compared to those decisions from the experts. Therefore, the constructed model has a high potential to be

used as the cardiovascular complication risk trend for these patients. However, there are the different accuracy results of simulated cases and decision from experts, although those accuracy results are complacency. Accordingly, the future of this paper will be focusing on adjusting the model based on testing the model with real patients and more experts for achieving higher prediction accuracy, including engagement with type 2 diabetes mellitus with hypertension patients to use the constructed model for self-monitoring along with their healthcare journey.

## V. CONCLUSION

This paper proposes a prevention of cardiovascular complication risk for type 2 diabetes mellitus with hypertension patients with detecting a trend predictive model by using a fuzzy logic. The proposed model uses 15 predictors including clinical and lifestyle risk factors based on the general fact knowledge and implicit knowledge of the experts. The simulation results show that the proposed trend predictive model provides 93% accuracy when compared to decide by the experts. Consequently, the proposed model has a high potential for predicting trend of cardiovascular complication risk for type 2 diabetes mellitus with hypertension patients. However, the constructed model still requires more adjustment based on more testing with real patients and experts.

## ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Ekkapob Pianpisises, a family medicine physician at Kasemrad Sriburin hospital for medical information support.

## REFERENCES

- [1] World Health Organization, Diabetes, 2015, [online]. Available: <http://www.who.int/mediacentre/factsheets/fs312/en/>
- [2] C.D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," *Plos med*, 3(11), e442, 2006.
- [3] World Health Organization, WHO global health days, 2013, [online]. Available: <http://www.who.int/campaigns/world-health-day/2013/en/>
- [4] J.R. Sowers, M. Epstein, and E.D. Frohlich, "Diabetes, hypertension, and cardiovascular disease: an update," *AHA Journals: Hypertension*, pp. 1053-1059, 2001.
- [5] N.R.C. Campbell et al., "Hypertension in people with type 2 diabetes Update on pharmacologic management," *CFP*, pp. 997-1002, September 2011.
- [6] World Health Organization, Q&As on hypertension, 2015, [online]. Available: <http://www.who.int/features/qa/82/en/>
- [7] P.W. Wilson et al., "Prediction of coronary heart disease using risk factor categories," *Circulation*, 97(18), pp. 1837-1847, 1998.
- [8] R.B. D'Agostino et al., "Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation," *Jama*, 286(2), pp. 180-187, 2001.
- [9] R. Conroy et al., "Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project," *European heart journal*, 24(11), pp. 987-1003, 2003.
- [10] World Health Organization, Prevention of cardiovascular disease: Guidelines for assessment and management of cardiovascular risk, WHO Document Production Services, Geneva: WHO, 2007.
- [11] J. Hippisley-Cox et al., "Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2," *Bmj*, 336(7659), pp. 1475-1482, 2008.
- [12] R.B. D'Agostino et al., "General cardiovascular risk profile for use in primary care the Framingham Heart Study," *Circulation*, 117(6), pp. 743-753, 2008.
- [13] M.J. Pencina, R.B. D'Agostino, M.G. Larson, J.M. Massaro, and R.S. Vasan, "Predicting the 30-year risk of cardiovascular disease the Framingham Heart Study," *Circulation*, 119(24), pp. 3078-3084, 2009.
- [14] C. Eswaran, R. Logeswaran, and A.R.A. Rahman, "Prediction models for early risk detection of cardiovascular event," *Journal of medical systems*, 36(2), pp. 521-531, 2012.
- [15] D. Otgontuya, S. Oum, B.S. Buckley, and R. Bonita, "Assessment of total cardiovascular risk using WHO/ISH risk prediction charts in three low and middle income countries in Asia," *BMC public health*, 13(1), 2013.
- [16] P.M. Ridker and N.R. Cook, "Statins: new American guidelines for prevention of cardiovascular disease," *The Lancet*, 382(9907), pp. 1762-1765, 2013.
- [17] P. Nordet et al., "Total cardiovascular risk assessment and management using two prediction tools, with and without blood cholesterol," *MEDICC review*, 15(4), pp. 36-40, 2013.
- [18] S. Selvarajah et al., "Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population," *International journal of cardiology*, 176(1), pp. 211-218, 2014.
- [19] J. Cederholm, K. Eeg-Olofsson, B. Eliasson, B. Zethelius, P.M. Nilsson, and S. Gudbjörnsdóttir, "Risk prediction of cardiovascular disease in type 2 diabetes A risk equation from the Swedish National Diabetes Register," *Diabetes care*, 31(10), pp. 2038-2043, 2008.
- [20] Emerging Risk Factors Collaboration, "Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies," *The Lancet*, 375(9733), pp. 2215-2222, 2010.
- [21] A. Willis, M. Davies, T. Yates, and K. Khunti, "Primary prevention of cardiovascular disease using validated risk scores: a systematic review," *Journal of the Royal Society of Medicine*, 105(8), pp. 348-356, 2012.
- [22] S. Van Dieren et al., "Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review," *Heart*, 98(5), pp. 360-369, 2012.
- [23] T. Robinson et al., "New Zealand Diabetes Cohort Study cardiovascular risk score for people with Type 2 diabetes: validation in the PREDICT cohort," *Journal of primary health care*, 4(3), pp. 181-188, 2012.
- [24] P. Radha and B. Srinivasan, "Hybrid prediction model for the risk of cardiovascular disease in type-2 diabetic patients," *International Journal*, 2(10), 2014.
- [25] World Health Organization, WHO/ISH risk prediction charts for 14 WHO epidemiological sub-regions, Geneva: WHO, 2007.
- [26] A.V. Senthil Kumar, Fuzzy expert systems for disease diagnosis, Medical Information Science Reference (an imprint of IGI Global): USA, pp. 2, 2015.
- [27] A. Ali and N. Mehli, "A fuzzy expert system for heart disease diagnosis," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 134-139, 2010.
- [28] G. Dudeky, A. Strzelewicz, M. Krasowska, A. Rybak, and R. Turczyn, "Fuzzy analysis of the cancer risk factor," *ACTA PHYSICA POLONICA B*, vol. 43, pp. 947-959, April 2012.
- [29] A.L. Alonso, O.A. Rosas-Jaimes, and J.A. Suarez-Cuenca, "Fuzzy logic assisted diagnosis for atherogenesis risk," *IFAC*, pp. 255-259, December 2013.
- [30] K.K. Oad, X. DeZhi, and P.K. Butt, "A fuzzy rule based approach to predict risk level of heart disease," *Global Journal of Computer Science and Technology*, 14(3-C), pp. 17-22, 2014.
- [31] J. Kim, J. Lee, and Y. Lee, "Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree," *Healthcare informatics research*, 21(3), pp. 167-174, 2015.

# Classification of Mental Tasks from EEG Signals using Spectral Analysis, PCA and SVM

Nikolay Neshov<sup>1</sup>, Agata Manolova<sup>2</sup>, Ivo Draganov<sup>3</sup>,  
 Faculty of Telecommunications  
 Technical University of Sofia  
 Sofia, Bulgaria  
 Email: {neshov<sup>1</sup>, amanolova<sup>2</sup>, idraganov<sup>3</sup>}@tu-sofia.bg

**Abstract**— Signals provided by the Electroencephalography (EEG) are widely used in brain-computer interface (BCI) applications. They can be further analyzed and used for thinking activity recognition. In this paper we proposed an algorithm that is able to recognize five mental tasks using 6 channel EEG data. The main idea is to separate the raw EEG signals into several frames and compute their spectrums. Next, a second-order derivative of Gaussian is applied to extract features and an optimum Gaussian kernel parameters grid search is performed with the help of cross-validation. The extracted features are further reduced by Principal Component Analysis. The processed data is utilized to train SVM classifier which is used for mental tasks recognition afterwards. The performance of the algorithm is estimated on publically available dataset. In terms of 5 folds cross-validation we obtained an average of 82.7% recognition rate (accuracy). Additional experiments were conducted using leave-one-out cross-validation where 67.2% correct classification was reported. Comparison to several state-of-the-art methods reveals the advantages of the proposed algorithm.

**Keywords**— *Electroencephalography (EEG); Brain Computer Interface (BCI); Fast Fourier Transform (FFT); Principal component analysis (PCA); Support vector machine (SVM)*

## I. INTRODUCTION

In recent years, the EEG-based brain-computer interface (BCI) has become one of the most promising areas of research in computer science and robotics thanks to neurorehabilitation. Neurorehabilitation is a relatively new field that combines series of therapies from the psychological to occupational, teaching or retraining patients on mobility skills, communication processes and other aspects of that person's daily routine [1]. The main goal of this complex medical process is the recovery of the patient from a nervous system injury and the minimization and/or compensation for related functional alterations or disabilities. There are several common medical conditions that are treated by this scientific field, for example: Cerebral palsy, Parkinson's disease, Brain injury, Spinal cord injuries, Multiple sclerosis, Stroke, Post-polio syndrome and Muscular Dystrophy. What these conditions have in common is the presence of damaged communication between the source of movement (brain, specific motor cortex) and the actuator of movement (muscles). The human nervous system is a highly complex, extremely versatile and dynamic neural network. So if the doctors, rehabilitators or the medical

personnel in general want to achieve any functional improvement, they need to be able to repair or substitute that part of the neural network that is not functional in part or in total anymore. Some of their goals may be to permanently replace with an appropriate device or artificial limb, or repair by retraining and re-learning, the impaired functionality. In either case one needs to study the relations between inputs and outputs of the system in order to achieve better patient rehabilitation.

A lot of research from different field experts is put into replacing or regaining the lost functionality with the use of emerging technologies such as robotics, computer vision, virtual reality and brain-computer interfaces for enhancing the disabled user's independence. The EEG is very important to understand the interaction of the different brain areas, study the effect of the abovementioned diseases on the neurological processes and build a BCI that can help paraplegic individuals [2]. The BCI will use the signals from neuronal activity in the brain to interface them with a computer. Thanks to EEG measurement many researchers are able to develop new technologies and therapies allowing the assessment of the resulting changes in the patient's brain [3] during their therapy sessions. Thus, the development of practical BCIs for disabled people using EEG signals should allow them to use all their existing mental and muscle functionalities as control possibilities in the system [4], [5]. This method had proved to be effective in helping patients with severe motor deficits to control remote devices such as computer cursor, artificial limbs or even moving a wheelchair. Various profiles can be built and different control strategies could be applied, depending on the condition of the patient.

The main contribution of this work is finding-out an appropriate feature extraction process for better analyzing EEG signals for mental task analysis widely used in BCI applications.

The rest of the paper is organized as follows: In the next section we provide a brief state-of-the-art review of some methods used for mental task recognition based on EEG signals. In Section 3 we present the content of the EEG Database [6] that we used to perform our investigations. In Section 4 we illustrate and explain the proposed algorithm. In Section 5 we give the experimental results. Finally, in Section 6, the paper ends with conclusions and discussions.

This work was supported in part by the contract DFNI I02/1 for research project: "Intelligent man-machine interface for assistive medical systems in improving the independent living of motor disabled users" of the Bulgarian Research Fund of the Ministry of Education and Science and the research project № 161PR0005-07, funded by RDS, TU-Sofia, Bulgaria.



## II. RELATED WORKS

Determining the user's activities is not only necessary for neurerehabilitation but according to [7] it is central for ubiquitous computing. The authors have developed an unobtrusive and lightweight single electrode BCI system that can be used to recognize with accuracy between 70-100 % different mental activities such as reading and relaxing with Bayesian networks. An interesting scenario is presented in [8] which deal with feature extraction and classification of horizontal mental task pattern on 1-D cursor movement from EEG signals. In this case three neural network classifiers are used: learning vector quantization, multilayer neural network and probabilistic neural network. Moustafa and Gad [9] use Linear Predictive Coding and Discrete Wavelet Transform for compression of EEG channels and feature extraction combined with Support Vector Machines (SVM) for classification of 5 mental tasks form [6] with an average recognition rate of about 85%. For the same database Hariharan et al. [10] propose the stockwell transform for feature extraction and then compare three classifiers: k-means nearest neighbors, linear discriminant analysis and SVM to test the strength of the proposed features.

The work cited in this paper works that share the same database contribute their own innovation in the field of mental tasks analysis for different applications either neurerehabilitation and BCI or ubiquitous computing. But it should be mentioned that direct comparison of their results is difficult due to the lack of uniformity in using the number of subjects, performing the types of experiments and presenting the results.

One of the main goals of our work is finding-out an appropriate feature extraction process for better analyzing EEG signals. A pool of methods for EEG feature extraction exists in the literature. The Auto-Regressive (AR) model, The Discrete Wavelet Transform (DWT), and the Fast Fourier Transform (FFT) are widely used for such purpose. The AR model is mainly used in signal processing especially for system identification [11]. Usually, in the AR model, the variable of interest is predicted by linear combination of past values of the variable. The feature vector is formed from the coefficients of the model and further used in BCI system [12]. However, the AR model cannot capture transient features from EEG signals [13] and the AR analysis suffers from speed, and hence not always applicable in real time applications. DWT represents efficiently EEG signals by decomposing each signal into frequency sub-bands. On the other hand, the window with varying size is the most significant parameter of this method since it ensures the suitable time frequency resolution in all frequency ranges [14]. Hence the choice of suitable window size might be a tricky for building high accurate system. In our work we use FFT as preferred method for feature extraction since it provides suitable frequency domain representation of the signal, many fast implementations exist, and is one of the most widely used methods in EEG analysis [15].

## III. DATABASE DESCRIPTION

In this study, we have used the publically available database, collected by Keirn and Aunon, from the BCI

laboratory of Colorado State University [6]. The database contains EEG signals recorded from 7 subjects, each of which performed at least five trials of five pre-defined mental tasks (each trial has duration of 10 seconds). The mental tasks are: 1.Baseline (relaxing as much as possible); 2.Multiplication (calculating multiplication mentally); 3.Letter-Composing (considering the contents of a letter); 4.Rotation (imagining rotation of a 3-D object); 5.Counting (imagining writing a number in order). The data were measured by 6 EEG channels and 1 EOG channel (to measure the movement of an eye). The electrodes are placed as shown in Fig. 1 and the measurements are made with reference to electrically linked mastoids, A1 and A2. The electrodes are connected to a bank of active filters whose band-pass bandwidths are set from 0.1 Hz to 100 Hz. The data is passed to 12-bit ADC operated at a sampling rate of 250 Hz. Thus, for a given task and a subject - 2,500 samples (250 Hz x 10 sec) per channel are recorded (for the whole trial). The experiments in this study are conducted for each subject independently. Thus, it is suitable to examine the algorithm performance utilizing equal amount of data per person. This allows objective comparison of the results among the subjects. Taking into account that subjects 2 and 7 completed only 5 trials, the analyses are performed over the EEG signals for the first 5 trials only for each participant.

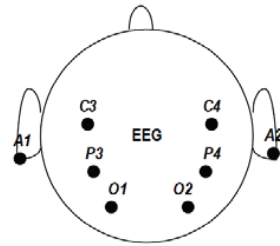


Fig. 1. Electrode placement

## IV. ALGORITHM DESCRIPTION

The EEG signal from each channel is first normalized to zero mean and variance one. Further it is divided into adjacent frames (each one has duration of 1 second which corresponds to a length of 250 samples). Thus, 10 frames are produced per channel. This process is illustrated on Fig. 2.

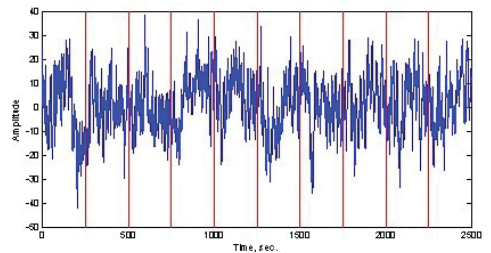


Fig. 2. Frames partitioning of the EEG signal recorded for one channel

The period of 1 second is chosen as reasonable for practical mental tasks recognition purposes. One can be asked to perform some thinking activity for at least one second before refocuses mind into another direction. Additional tests in our experimental setup showed that the maximum classification performance is reached using comparable time periods. The next step is Fast Fourier Transformation (FFT). The motivation behind the choice of FFT is inspired by the assumption that there is a relation between specific mental task and the distribution of magnitudes of specific group of frequencies of the EEG signals. Considering 250 samples for each frame and 6 channels data we have to compute 6 spectrums at a time. After applying the FFT the resulted spectrum is symmetrical around the zero frequency. Thus, only the half part (i.e. 125 samples) is further used in the next steps.

The third step is feature extraction. This is essential part for reaching high recognition accuracy. In order to describe the rate of change of spectrum shape we convolved the spectrum using second derivative Gaussian filter of size  $s$ . The choice of suitable kernel parameters (the filter size -  $s$  and the standard deviation -  $\sigma$ ) is very important for getting higher classification accuracy. We evaluated the classification performance for each person (in terms of 5 folds cross-validation accuracy) using different sets of parameters ( $s$ ,  $\sigma$ ). The analyses are presented later-on in Sec. V. The results produced by this step are six feature vectors per frame (i.e. one feature vector per channel). Each feature vector has 125 dimensional representations.

To form a dataset for analysis, we concatenated all 6 feature vectors in one which generates  $6 \times 125 = 750$  dimensional feature vector. Further, we used Principal Component Analysis (PCA) to reduce the vector's size preserving 99% of the energy. The reduced vector after PCA is then utilized for training Support Vector Machine (SVM) classifier using LibSVM [16]. In our study we use radial basis function (RBF) [17] where the optimum parameters are found by grid-search. To solve the multi-classification problem we used one-against-one strategy. Once the SVM model is built the algorithm can be used in classification/testing mode. In this mode the unknown/testing EEG signals are passed through each of the aforementioned steps.

## V. EXPERIMENTAL RESULTS

### A. Finding Optimal Filter Parameters

In our investigations we found out that the recognition rate is significantly depended by the kernel parameters we used in the feature extraction process. Our experiments also showed that the pair of kernel parameters (optimum size -  $s_{opt}$  and optimum standard deviation -  $\sigma_{opt}$ ) that gives the maximum cross-validation accuracy ( $Acc_{max}$ ) varies among the different persons from the database. Therefore we applied a grid search to find-out the optimal filter parameters for each subject individually. Especially, the size of the kernel  $s$  is in the range from 20 to 120 with an increment of 20. The parameter  $\sigma$  is between 10 and 60 and the increment is 10. The variation of the parameters, in the aforementioned ranges, leads to change in the accuracy of about 20%. This means that for real applications of the algorithm, the optimum parameters should

be found for a given user prior building the SVM model. In general (considering all subjects), it can be observed that the lowest accuracy performances are reported for the smallest standard deviations and the influence of variation of  $\sigma$  over the recognition accuracy is higher than that of  $s$ . The results from these investigations are presented in Table 1. The maximum accuracy is obtained for Subject 1 (92%) and the minimum – for Subject 3 (61.6%). The recognition rate is 82.7% on average. The optimum parameters found are further used for the experiments reported in the next subsection.

TABLE I. 5-FOLD CROSS-VALIDATION ACCURACY OBTAINED WITH THE OPTIMUM FILTER PARAMETERS

Subject	$s_{opt}$	$\sigma_{opt}$	Accuracy, %
1	120	50	92
2	160	40	86.4
3	20	30	61.6
4	40	20	88
5	40	30	82.8
6	100	30	82
7	140	60	86

### B. Performance Evaluation

In order to investigate the recognition performance of our algorithm in more realistic situations, we applied leave-one-trial-out cross-validation strategy for each subject individually. This means that at a time we used four of all five trials, available for a given subject, as a training data and the remaining one – as a testing data and calculate the recognition accuracy. Further we repeat the same procedure 4 times, each time considering different trial set for testing (and the remaining trails for training) and then the classification accuracy is evaluated based on the average over the five trials. This experimental setup makes it possible to perform a deep and realistic analysis of the algorithm performance since it represents as much as close to the real environment scenario where one can be asked to train the system in several trials and further utilizes it in recognition mode. We calculate the average recognition accuracy for all seven subjects using the first five trials. Since each trial contains 10 datasets (for a specific task) we have a total of 50 datasets per task. This gives: 50 datasets per task  $\times$  5 tasks = 250 datasets for training/testing per person.

In Table 2 we summarize the obtained results for each person. We compute the recognition accuracy in two manners: averaging the recognition rates by subjects and by tasks. It can be seen that for Subject 1, the Letter-Composing (LC) task is recognized correctly. However in the lowest accuracy is measured for Rotation and Counting. Considering the other six subjects we can see that the best accuracy is obtained for different tasks among the different persons.

Despite these disordering observations, we see promising results for the Subject 1. Probably this would mean that the success of recognition is hidden in the person's ability to fully concentrate the mind to certain mental task.

Obviously, the best recognition rate is obtained for Subject 1 (82.9%). The worst case is reached for Subject 3 with recognition rate of 47.1%. However, the standard deviation of accuracies does not exceed 13.4. Averaging the results among the different Subjects shows that the highest recognition rate is

obtained for the Letter-Composing task (73.2%) and the lowest one – for the Baseline task (59.3%). The last observation probably can be explained by the assumption that the full relaxing is difficult to be achieved, which causes misleading brain waves to produce wrong results.

TABLE II. AVERAGE ACCURACY (IN %) AND ITS STANDARD DEVIATION PER TASK AND SUBJECT OBTAINED BY LEAVE-ONE TRIAL OUT CROSS-VALIDATION (ROUNDED TO THE NEAREST INTEGER)

Task \ Subject	S1	S2	S3	S4	S5	S6	S7	Avg. accuracy for each task	St. dev. for each task
Baseline	80.0	63.8	39.0	66.7	58.5	52.1	59.3	59.9	12.7
Multiplication	88.4	78.2	52.1	81.3	54.0	70.2	61.1	69.3	14.1
Letter comp.	96.2	71.7	44.3	88.6	71.7	55.6	84.2	73.2	18.4
Rotation	73.6	95.2	48.8	58.2	72.1	66.7	71.7	69.5	14.5
Counting	76.6	59.2	51.3	67.8	40.0	82.0	72.4	64.2	14.9
Avg. accuracy for each subject	82.9	73.6	47.1	72.5	59.3	65.3	69.7	67.2	
St. dev. for each subject	9.2	14.1	5.5	12.2	13.4	12.0	10.1		

It should be noted that the authors of the most of the state-of-the-art papers cited in this paper do not separate their training/testing sets by trials but perform cross-validation randomly using training and testing samples from a specific trial. For example the algorithm presented in [15] was evaluated on the same brain data and the average recognition rate is only 30%. In another work [18] is presented a framework for classification of mental tasks from EEG signals using extreme learning machine and without further post-processing the optimum average classification rate is 56.07%.

VI. CONCLUSION

We proposed a method and algorithm for mental task recognition using EEG signals. In terms of 5-fold cross-validation accuracy we reached 82.7% correct classification. We conducted additional experiments by using leave-one-trial-out cross-validation strategy, where the average accuracy rate is 62.7%. The experimental results showed that our algorithm provided recognition accuracy higher than some of the conventional methods. We will enlarge our research work by using different feature extraction processes (e.g. the Autoregressive models) in combination with FFT that is expected to improve the recognition rate.

This improvement in the mental tasks classification can provide a better communication pathway between the brain and the machines, which in turn will help to develop of more reliable assistive devices such as brain wave controlled wheel chairs, prosthetic limbs and smart living environment for patients with brain injuries. We have started working on combining the EEG with EMG signals for estimating and classifying mental and muscle fatigue.

REFERENCES

[1] Perez-Marcos D., Solazzi M., Steptoe W., Oyekoya O., Frisoli A., Weyrich T., Steed A., Tecchia F., Slater M., Sanchez-Vives M. V., "A fully-immersive set-up for remote interaction and neurorehabilitation

based on virtual body ownership", *Frontiers in Neurology* (2012) 3:110. doi: 10.3389/fneur.2012.00110.

[2] Christian Flores Vega; Francisco Javier Ramirez Fernández, Recognition of mental task with the analysis of long-range temporal correlations on EEG brain oscillation, *Biosignals and Bioinformatics Conference (BRC)*, 2012 ISSNIP, DOI: 10.1109/BRC.2012.6222167.

[3] Golomb, M. R., McDonald, B. C., Warden, S. J., Yonkman, J., Saykin, A. J., Shirley, B., Huber, M., Rabin, B., Abdelbaky, M., Nwosu, M. E., Barkat-Masih, M., and Burdea, G. C. (2010). In-home virtual reality videogame telerehabilitation in adolescents with hemiplegic cerebral palsy. *Arch. Phys. Med. Rehabil.* 91, 1–8 e1.

[4] Lalitharatne, T. D., Teramoto, K., Hayashi, Y., Kiguchi K., Towards Hybrid EEG-EMG-Based Control Approaches to be Used in Bio-robotics Applications: Current Status, Challenges and Future Directions, *Paladyn, Journal of Behavioral Robotics*. Vol. 4, Issue 2, pp. 147–154, December 2013.

[5] Agata Manolova, Georgi Tsenov, Violeta Lazarova, Nikolay Neshov, Combined EEG and EMG fatigue measurement framework with application to hybrid brain-computer interface, *Communications and Networking (BlackSeaCom)*, International Black Sea Conference on, Sofia, Bulgaria, 2016, in print;

[6] Colorado State University, Brain-Computer Interfaces Laboratory: <http://www.cs.colostate.edu/eeg/>.

[7] AS Shirazi, M Hassib, N Henze, A Schmidt, K Kunze, What's on your mind?: mental task awareness using single electrode brain computer interfaces, *Proceedings of the 5th Augmented Human International Conference*, pp. 45, 2014.

[8] M. Serdar Basçil, Ahmet Y. Tesneli, Feyzullah Temurtas, Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface, *Australasian College of Physical Scientists and Engineers in Medicine*, 2015, DOI 10.1007/s13246-015-0345-6.

[9] Mohamed Moustafa, Azmy Gad, Classification of mental tasks using support vector machine based on linear predictive coding and new mother wavelet transform, *Biomedical Engineering and Computational Technologies (SIBIRCON)*, 2015 International Conference on, DOI: 10.1109/SIBIRCON.2015.7361873.

[10] M. Hariharan, Vikneswaran Vijean, R. Sindhu, P. Divakar, A. Saidatul, Szali Yaacob, Classification of mental tasks using stockwell transform, *Journal Computers and Electrical Engineering archive*, Vol. 40, Issue 5, July 2014, pp. 1741-1749.

[11] Y. Zhang, B. Liu, X. Ji, D. Huang, 'Classification of EEG Signals Based on Autoregressive Model and Wavelet Packet Decomposition' *Neural Processing Letters*, 2016.

[12] V. Lawhern, WD. Hairston, K. McDowell, M. Westerfield, K. Robbins, 'Detection and classification of subject-generated artifacts in EEG signals using autoregressive models', *J Neurosci Methods* 208, pp. 181–189, 2012.

[13] D. Hu, W. Li, X. Chen, 'Feature extraction of motor imagery EEG signals based on wavelet packet decomposition', *IEEE/ICME International Conference on ComplexMedical Engineering (CME)*, pp 694-697, 2011.

[14] D., Ubeyli, 'Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines. *Computers in Biology and Medicine'*, Vol. 38, No, 1, pp. 14–22, 2008.

[15] T. Kuremoto, Y. Baba, M. Obayashi, S. Mabu and K. Kobayashi, 'To extraction the feature of EEG signals for mental task recognition', *Society of Instrument and Control Engineers of Japan (SICE)*, 2015 54th Annual Conference of the, Hangzhou, pp. 264-269, 2015.

[16] C.-C. Chang, C.-J. Lin, 'LibSVM: a library for support vector machines', *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[17] B. Scholkopf and A. J. Smola, '*Learning with Kernels*', MIT Press, 2002.

[18] N. Liang, P. Saratchandran, G. Huang, N. Sundararajan, 'Classification of mental tasks from EEG signals using extreme learning machine', *International Journal of Neural Systems*, Vol. 16, No. 1, pp. 29–38, 2006.

# Increased Personalization and Intelligent Behavior of Future Workplaces Based on Applied CBR-Based Decisions Support Service

Anton N. Katov, Sofoklis Kyriazakos, Ramjee Prasad  
Center for Teleinfrastruktur (CTIF)  
Aalborg University, Denmark  
{annka,sk,prasad}@es.aau.dk

**Abstract**—Designing and building intelligent workplaces that effectively support people in their daily activities and increase their work productivity has been widely discussed over the last years. Such systems can be considered especially beneficial for elderly employees that struggle in their daily interactions with the modern technologies. The current paper describes an intelligent decision support system (DSS), developed for the needs of the Platform for Ergonomic and motivating, ICT-based Age-friendly Workplaces (PEARL) under the Ambient Assisted Living (AAL) Joint Programme. The proposed approach aims to provide a customized configuration plan to every new platform user during the registration process, based on his personal demographic, physical and mental characteristics, thus facilitating the initial platform deployment and improving the user experience. The DSS has been implemented as a RESTful web-service, which incorporates a case-based reasoning (CBR) matchmaker, coupled with a rule-based case adaptation mechanism.

**Keywords**— *ambient intelligence, AAL, DSS, CBR, rule-based systems*

## I. INTRODUCTION

The concept of pervasive computing, also referred to as ambient intelligence (AmI), embodies the idea of incorporating embedded technology into everyday objects, thus providing ubiquitous connectivity and computing capabilities using any device, in any location [1]. This enables unprecedented convergence of physical and virtual worlds that can offer better decision making, better monitoring, better control and improved efficiency in virtually every aspect of industry and everyday life. Intelligent workplaces are an integral part of the AmI concept, aiming to provide increased personalization and intelligent behavior, supporting the everyday tasks of the employees through the deployment of innovative ICT-based applications. Such solutions can be considered particularly beneficial for elderly employees, who have limited ICT literacy and struggle in their daily interaction with modern technologies.

The demographic trend of ageing of the population makes such problems even more relevant. In [2] it is indicated that the proportion of people aged 60 or over is expected to rise from 12% in 2013 to 21% in 2050 and therefore addressing the various challenges that older individuals might experience, such as impaired vision and hearing, limitation in the physical activity, decline in the cognitive functions, various chronic

decreases, becomes more and more relevant. Furthermore, ageing will introduce various socio-economic challenges to the health care systems and society as a whole. To tackle these problems ambient assisted living (AAL) paradigm has been introduced. AAL is defined by the European AAL Joint Programme (AAL JP) as an initiative that aims to “enhance the quality of life of older people through the use of Information and Communication Technologies (ICT)”.

While most of the current AAL solutions target the improvement of the living conditions of the elderly in their homes, Platform for Ergonomic and motivating, ICT-based Age-friendly Workplaces (PEARL) project is the first to address the problems that the elderly face at their workplaces in the years before retirement. The project has developed a holistic platform that gives access to a unique combination of ICT-based solutions, specifically developed to address the needs of older employees, as well a mechanism for ambient configuration of the surrounding environment, based on the currently performed task and the user preferences. In order to automate the platform’s deployment and to provide a customized configuration plan to every new platform user during the registration process, thus greatly improving the user experience, we have developed an integrated decision support system (DSS). The DSS undertakes the task of mapping a set of user profile, workspace and task characteristics to a set of platform configuration settings and adjusts the necessary configuration parameters, based on the specific characteristics of each user. The operation of the developed DSS is based on the Case-Based Reasoning (CBR) paradigm in combination with a Rule-Based engine for case adaptation. The system has been implemented and operates as a RESTful web service in order to ensure ease of integration with existing web-based platform solutions and particularly with the PEARL platform.

The current paper briefly outlines the underlying technologies and describes the system design and the implementation approach applied in the development of the intelligent decision support service for configurations customization. The paper is organized as follows: Section II introduces the case-based reasoning technique and the rule-based systems approach, which constitute the core intelligence of the proposed solution. Section III describes system design and the process of knowledge model definition for the CBR matchmaker and the rule-based case adaptation mechanism. Section IV presents the implementation details of the developed web service, Section V discusses the performance

evaluation of the developed system and finally Section VI concludes the paper.

## II. EXPERT SYSTEMS AND CASE-BASED REASONING

The proposed DSS solution consists of two major submodules – the CBR-based Matchmaker and the case adaptation Rule Engine. The Matchmaker is responsible for comparing the profile characteristics to the case base of existing users, calculating the best match and retrieving the corresponding configuration. The Rule engine then undertakes the retrieved configurations and adjusts them, based on the profile differences between the new and the retrieved users.

### A. Case-Based Reasoning

The CBR technique was selected as an appropriate solution because of its ability to operate in situations where the domain might have undefined parameters, a weak or unknown causal model or when its formalization requires a huge amount of rules. In [3] CBR is defined as reasoning by remembering, i.e. extracting and adapting already existing past solutions of a problem in the process of solving new problems. The fundamental component of every CBR system is the case base (CB), which stores the collections of all previously known or experienced problem situations in the form of cases. Each case defines contextually a known past experience in terms of a problem definition and the respective solution or the expected outcome. In the current design the problem definition consists of a set of user characteristics, i.e. user profile model, whereas the corresponding solution defines the preferred workspace configurations settings, i.e. configuration plan.

The general process of problem solving of a CBR system follows four distinct phases [4]. In the case retrieval phase the new problem is defined and assessed and the CB is searched for a matching or similar case in order to obtain a relevant solution. Here, during the retrieval phase the most similar existing profile will be selected and its corresponding configuration plan model will be returned as a solution. Once an initial solution is selected, it is adapted during the following adaptation phase in order to fit the specific requirements of the newly presented problem. The adaptation in the implemented DSS is based on a set of rules that account for the indicated differences and perform the necessary changes to the original solution. The applicability of the resulting solution is then verified in the evaluation phase. If the verification is successful, the newly formed problem-solution pair can be stored in the CB in the final phase of the CBR cycle. In case the result from the evaluation is unsatisfactory, the solution might be adapted further or additional solutions might be retrieved. In the current context the solution will be instantly applied by deploying the selected configuration settings and the user evaluation of the proposed solution will take place during the ongoing usage of the PEARL platform. The user will be provided with a user friendly interface to change the configuration settings (i.e. adapt the solution) at any point.

In addition to the cases which represent the specific problem situations, a CBR system may include also general knowledge about the specific problem domain, for which the

system has been designed. In [5] three types of problem domain knowledge are identified – vocabulary, adaptation knowledge and similarity measures. The vocabulary describes the feature parameters that define each case and are used to retrieve relevant cases. The vocabulary of the DSS matchmaker will comprise of the user profile and configuration plan model parameters. The adaptation knowledge contains information about the influence of each parameter in the form of explicit rules, which were defined and integrated as a rule-based expert system for case adaptation. And finally, the similarity measures encode the similarity model that is utilized in the case retrieval process.

### B. Rule-Based Case Adaptation System

In order enable the provision of fully customized configuration plans for every new user, rule-based semi-automatic case adaptation process was implemented in the proposed DSS. The implemented rule-based systems takes as an input the configuration plan parameters produced by the CBR matchmaker and returns an adapted solution, based on a set of rules that account for the differences between the new and the already existing users. Rules are a set of if-then statements that specify how to act or what conclusion to offer based on a given set of input data. There are two type of rules identified in [6]: knowledge rules that state the facts and their relationships and inference rules that define how two find a solution in the presence of a set of facts. The knowledge rules are stored in the knowledge base, whereas the inference rules are part of the inference engine.

Rule-based expert systems target narrow problem domains and deal with qualitative rather than quantitative problems. As it has been already mentioned above, providing a set of customized configuration parameters, based on a predefined list of user characteristics, can be characterized as a complex highly unstructured problem. Due to the high number of input and output parameters and the ambiguous interrelations between them, the explicit modeling of the problem domain by a set of rules was determined to be unfeasible. Therefore, for the needs of the current design we have implemented a case-based reasoning approach for the selection of the initial solution and we have integrated a rule-based system solely for the needs of adaptation of separate configuration parameters.

## III. SYSTEM DESIGN

The DSS service, presented in the current paper, has been developed as an integral part of the PEARL platform [7] and undertakes the tasks of mapping the user profile, workspace and task characteristics to a set of configurations parameters, adjusted to the needs of the particular user. The service is called during an initial user registration phase and aims to simplify the initial setup process and to improve the interaction with the platform by providing a customized configuration plan for each new user. In order to provide maximum flexibility during the initial platform configuration phase a semi-automated operation model is designed, where a set of customized configuration settings will be suggested and the user will be allowed to either accept them or further modify them, based on his/her preferences.

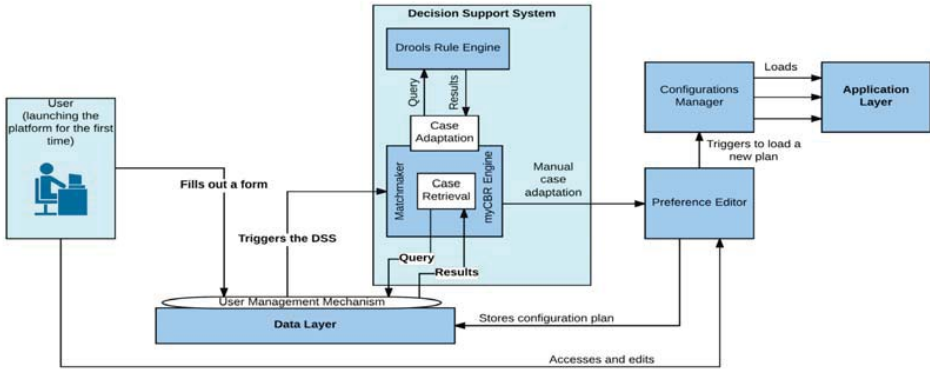


Fig. 1. Decision Support System High Level Design

### A. System Design and Workflow

The proposed solution is based on a Case-Based Reasoning (CBR) engine that incorporates a rule-based system module for case adaptation. The high level system design is presented in figure 1. The core of the DSS consists of two main submodules – the Matchmaker and the Rule Engine. The Matchmaker is powered by myCBR open source tool [8] and is responsible for matching a set of user preferences to a similar existing profile and retrieving the relevant configuration plan. In the process of selection, a set of similarity functions provided by myCBR engine are employed, which will be summarized in the following section. The Rule Engine is powered by Drools [9] and incorporates the case adaptation logic in the form of explicit rules, which are executed whenever a retrieved solution does not correspond to the requirement of the newly created user profile.

The workflow of the DSS is initiated by a new user who is willing to register to the platform and creates a new profile, filling out her/his personal characteristics. This information defines the problem input to the CBR Matchmaker, which is triggered upon saving the user’s data. The Matchmaker dynamically builds a case base by retrieving all fully completed profiles of existing platform users from the database, calculates their similarity scores and returns the best matching profile and its corresponding configuration settings plan as the suggested solution. At this point the Rule Engine is triggered to adapt the proposed configuration plan in order to cover potential differences between the new user and the profile retrieved by the CBR Matchmaker. The adapted configuration plan is returned in JavaScript Object Notation (JSON) format, parsed by the Preference Editor and offered to the user in the form of a prefilled user interface input form, specifically designed in accordance with the Web Content Accessibility Guidelines (WCAG) 2.0. At this point the user can either accept the proposed configuration plan or further modify it in accordance with her/his preferences. Once the

configuration settings are successfully completed/accepted by the user, a new case is formed, which can then be retrieved by the Matchmaker upon a following request. The user can also access and modify the configuration settings at any time via the Preference Editor menu. As the case base is dynamically created on each request, only the most relevant, currently active configurations settings will be used by the Matchmaker, which allows for a continuous adaptation of the CBR case base.

### B. Knowledge Model Definition

The knowledge model of the proposed solution consists of three main entities – vocabulary, similarity model and adaptation rules. The vocabulary defines the concepts and the underlying attributes of the CBR case base. It constitutes of a set of problem-solution pairs, where the problem part is defined by the available user characteristics, as specified by the User Profile Model, and the retrieved solution provides the Configuration Plan Model entity, available for this particular user. The User Profile Model contains information about the user’s demographic, cognitive, physical and professional characteristics that is used by the CBR matchmaker for the selection and customization of the platform features. The Configuration Plan Model defines not only the general characteristics of the workspace environment and the list of user’s common daily tasks, but also allows the end user to choose different workspace setups for the different tasks she/he is involved in.

The similarity model, used by the CBR engine in the process of case retrieval, was designed following the local-global principle [10], which proposes decomposition of the similarity function into local similarity function that is used for the evaluation of separate individual case attributes, and global similarity function that combines the local similarities and is used to compare cases on a higher level. The myCBR Workbench tool, which provides convenient GUIs for modeling knowledge-intensive similarity measures, has been

used to define the vocabulary attributes and the appropriate similarity functions [11]. Three main data types were used for the definition of the local similarity functions - Symbol, Float and Integer. The majority of the attributes of the User Profile Model has been defined as symbolic attributes undertaking a predefined set of values. For such attributes similarity tables were used, defining all pairwise combinations between the different values together with their similarities. For each of the numeric attributes symmetric polynomial similarity decrease functions were defined. The global similarity function was defined as a weighted sum function, defining the appropriate weights heuristically and further adjusting their values through a series of test retrievals to refine the retrieval results.

The final major entity of the DSS knowledge model is the rule base of the case adaptation Rule Engine, which contains a set of formalized rules that ensure the provision of a fully customized configuration plan model even in cases of closely matching user profile, retrieved by the CBR Matchmaker. We have operated under the assumption that even closely matching profiles can still differ in some key parameters and therefore some case adaptation logic is needed. The rules have been defined in the Drools native language and were expressed as separate WHEN-THEN structures [9]. Each rule consists of a rule name, optional attributes, condition and consequence. The rule name is a string value that describes the main purpose of the given rule and has to be unique in the given rule package. The attributes define the behavior parameters of the rule, such as dialect, salience, duration, etc. The left-hand side is the conditional parts of the rule. It starts with when keyword and consists of zero or more conditional elements. If no condition elements are defined, the rule will be defined as always true and will be always activated, when a new sessions is created. The conditional elements operate on one or more patterns that will be matched to the facts inserted in the working memory. Each pattern defines zero or more constraints and can be bound to a pattern binding variable. The constraints are expressions that return true or false. The right-hand side defines the consequence (or the action part) of the rule and contains the code that will be executed if the "WHEN" part is true. The plain purpose of the actions part is to modify working memory data by insert, delete or modify statements.

IV. DSS SERVICE IMPLEMENTATION AND PRINCIPLE OF OPERATION

In order provide a service that allows ease of integration with existing web-based platform solutions, the DSS was designed and implemented as a RESTful web service. Representational State Transfer (REST) architectural style allows web services to be designed to serve specific resources based on client request. In the current scenario the implemented web service takes as an input the ID of a newly registered user, retrieves her/his profile from the platforms database, matches it to the existing entries in the case base and returns the customized configuration plan in JSON format. One of the main advantages of the proposed approach is that it retrieves the user-related data directly from the platform's database and builds the required case bases dynamically on request, thus avoiding the use of outdated or irrelevant

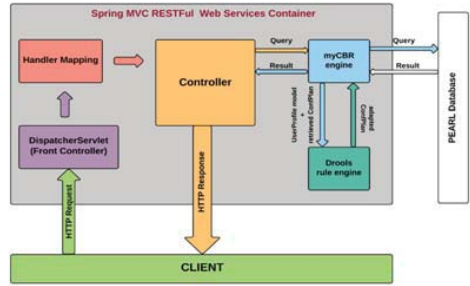


Fig. 2. DSS web-service architecture and principle of operation

In the current design Spring annotation-based Model-View-Controller (MVC) framework [12] was chosen for the implementation of the DSS RESTful web service due to its extensive documentation and library support. The basic architecture and principle of operation are illustrated in figure 3.

DSS web service is initially invoked by an HTTP GET request containing the new user ID sent by a client application. It is important to note that the new user profile data has to be stored already in the platform's database. The HTTP request is intercepted by the Dispatcher Servlet, provided by the Spring MVC framework, and directed to the registered handlers, which implement exception handling capabilities and map the request to the appropriate controller. The controller is a java object that implements a set of classes that are invoked to handle HTTP requests. The main controller class is responsible for invoking the CBR Matchmaker, powered by myCBR engine, retrieving the best configuration plan, triggering the rule engine and instantiating a new knowledge session to adapt the configuration plan, assembling the resource with configuration plan parameters in JSON format and returning an HTTP response to the client application. The two core components that implement the intelligence of the DSS service are the CBR Matchmaker and the adaptation Rule Engine.

The CBR Matchmaker uses the myCBR open source similarity retrieval tool core libraries, including methods for database access and case base construction and similarity-based case retrieval. The key method called by the controller takes an input an instance of the *UserProfileModel* class, converts the class variables to CBR concept attributes invokes the appropriate case retrieval method, prepares and executes the case base query and retrieves the id of the best matching user profile.

DSS adaptation rule engine is implemented with the use of the Drools business logic integration platform libraries, added to the project from a Maven repository. As every Drools-based project, the Rule Engine module consists of rules, facts and knowledge session. The facts are passed as input parameters to the engine's core method and include the existing instance of the *UserProfileModel* class, which holds the parameters of the newly created user profile, and an instance of the

*ConfigurationSettings* class, which contains the solution, retrieved by the matchmaker. The rules are defined in Drools native language and are stored in the corresponding *.drl* file. The engine's core method instantiates a Drools knowledge session, which defines the evaluation algorithm that determines how to match the rules against the current set of domain objects and fires all rules. The configuration settings are added to the knowledge sessions as a global variable, which allows the rule engine to directly modify certain configuration parameters, based on the predefined rules. Finally, the method returns an instance of the *ConfigurationSettings* class, which holds the adapted configuration settings. At this point the controller assembles this resource in JSON format and returns it in the body of the HTTP response.

## V. PRELIMINARY EVALUATION AND KNOWLEDGE REFINEMENT

The implemented DSS service testing was performed as part of the evaluation of the PEARL platform. The preliminary stage included two sets of structured lab trials with different groups of elderly employees. Comprehensive field trials have been planned towards the end of the project lifetime. The lab trials served as a basis for extending and refining the case base and the similarity measures in order to improve the system's performance. The process of knowledge refinement is illustrated in figure 3.

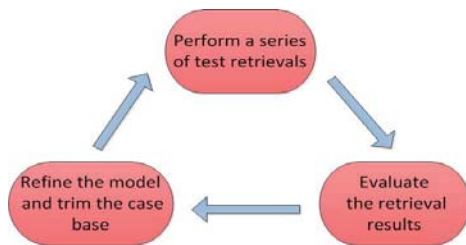


Fig. 3. CBR knowledge refinement process

In order to test the systems' performance and accuracy the case base was extracted from the platform's database and imported to myCBR Workbench tool. Twenty test profiles were created and entered in the system, varying the age, the physical and mental characteristics. The retrieval results were evaluated and the knowledge model was refined in several steps. The first step was to reexamine the *UserProfile* concept and look for attributes that might be considered irrelevant to the retrieved solution. Although it was decided to keep the original model's attribute unchanged, certain value ranges were reduced. Manipulating the case base, which was built during the lab trials, was the next step in the knowledge refinement process. We have looked for poorly defined or incomplete user profiles and we either removed, or edited them in order to offer tangible solutions. The last step was to redefine the knowledge model's similarity measures in order to produce more accurate case retrieval. The performance of

the refined system will be tested extensively during the upcoming field trials.

## VI. CONCLUSION

The paper presented an innovative design of an intelligent decision support system for configurations settings customization in the context of intelligent workplaces, designed for the needs of elderly employees. The implemented approach addressed the complex highly unstructured problem of defining a customized configurations plan, based on a set of personal characteristics by combining two complementing artificial intelligence techniques – case-based reasoning and rule-based expert systems. The proposed solution was developed as RESTful web service in order to ensure compliance with existing web-based platforms and to facilitate its integration in various scenarios, targeting increased personalization and intelligent behavior of future workplaces.

## Acknowledgment

The work, presented in this paper, has been carried out as part of the Platform for Ergonomic and motivating, ICT-based Age-friendly WorkpLaces (PEARL) project. The authors would like to acknowledge the contribution of all PEARL colleagues, as well as the European Commission for providing the Ambient Assisted Living (AAL) Joint Programme funding framework.

## References

- [1] F. Sadri, "Ambient intelligence: A survey," *ACM Comput. Surv.*, vol. 43, no. 4, pp. 36:1–36:66, Oct. 2011.
- [2] UN, "World population ageing," 2013, pp. 11–13.
- [3] J. Kolodner, "Case-Based Reasoning," San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [4] A. Aamodt, E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Communications*, IOS Press, Vol. 7: 1, pp. 39–59, 1994.
- [5] M.M. Richter, "On the notion of similarity in case-based reasoning," *Mathematical and Statistical Methods in Artificial Intelligence*, pp. 171–184. Heidelberg, Germany: Springer-Verlag, 1995.
- [6] E. Turban, J. Aronson, "Decision Support Systems and Intelligent Systems," 5th ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1997.
- [7] PEARL deliverable 2.2, "Use Cases, Scenarios and Integrated Functionalities," available at: [http://www.pearl-project.eu/data/dataimages/Upload/PEARL\\_D2-2\\_Scenarios\\_Functionalities\\_final.pdf](http://www.pearl-project.eu/data/dataimages/Upload/PEARL_D2-2_Scenarios_Functionalities_final.pdf), last accessed: 26/09/2016.
- [8] K. Bach, K.-D. Althoff, "Developing Case-Based Reasoning Applications Using myCBR 3," in *Case-Based Reasoning Research and Development*, B. D. Agudo and I. Watson, Eds. Springer Berlin Heidelberg, 2012, pp. 17–31.
- [9] Reference manual Drools 6.4.0.Final, available at <http://docs.jboss.org/drools/release/6.4.0.Final/drools-docs/html/index.html>, last accessed: 22/09/2016.
- [10] M. Richter, "Foundations of similarity and utility", *Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2007)*, AAAI Press, 2007.
- [11] C. S. S. Kerstin Bach, "Knowledge Modeling with the Open Source Tool myCBR," *CEUR Workshop Proceedings*, vol. 1289, 2014.
- [12] Spring Reference Documentation, "Section 22: Web MVC framework," available at: <http://docs.spring.io/spring/docs/current/spring-framework-reference/html/mvc.html>, last accessed: 23/09/2016.



# An overview of children activity monitoring framework for attention deficit hyperactive disorder analysis

Tew Hongthong  
School of Information Technology  
Mae Fah Luang University  
Chiang Rai, Thailand  
tew.hon@mfu.ac.th

**Abstract**— Attention deficit hyperactive disorder(ADHD) is a serious limitation in the child's learning. It begins at age 6 to 12 years old and has continued for more than six months. It has been seen in school ages and often make results in poor academic performance. It has effect with behaviour, emotion and learning. Research creates frameworks and applications that are used to collect information of the biological system and analyse the possibility of the occurrence of ADHD. We can recognise the symptoms early, we can cure the illness of a child faster. Also, education can use it to evaluate the effectiveness of learning in students.

**Keywords**—*Wearable sensors; Human Activity Monitoring; Human Behavior; Machine Learning; Attention Deficit Hyperactive Disorder;*

## I. Introduction

Recently, people might be familiar with internet, which is mostly overview of the connection between people and the Internet. The internet of things(IoT) is Internet technology to connect equipment such as mobile phones, cars, televisions, refrigerators and other devices by connecting and communicating with each other via the Internet. In the future, the technology will help people which controlled items both from home, office or from anywhere in the home such as temperature control, turn on/off the light[3]. The internet of things is using in an education such as mobile learning and cloud technology. Furthermore, researcher interested in the adoption of internet of things to apply in education. On the other hand, attention deficit hyperactive disorder(ADHD) is a serious limitation in the child's learning. The symptoms can begin at age 6 to 12 years old and have continued for more than six months<sup>2</sup>. It hasseen in school ages and

often make results in poor academic performance. Therefore, it has effect with behavior, emotion and learning. The idea of this research is adapt internet of things technology such as wearable device to help student who might faces with risk factor of attention deficit hyperactive disorder(ADHD).

This research will focus on application for monitoring information of biological system, activity or behavior of human to analysis symptom of attention deficit hyperactive disorder. In addition, researcher has studied many technologies include wearable sensors, human activity monitoring and human behavior. Attention deficit hyperactivity disorder is a mental issue of the neurodevelopmental type[1]. It is described by issues focusing, unnecessary movement, or trouble controlling conduct which is not fitting for a man's age. Human activity monitoring is a part of activity recognition that expects to perceive the activities and objectives of one or more specialists from a progression of perceptions on the operators' activities and the natural conditions. Human behavior alludes to the variety of each physical activity and recognizable feeling connected with people, and also mankind all in all. While particular attributes of one's identity and disposition might be more steady, different practices will change as one moves from birth through adulthood. Wearable technology is identified with both omnipresent processing and the history and advancement of wearable PCs. Wearables make innovation unavoidable by interlacing it into day by day life. Therefore, researcher design application and framework for analyze probability of attention deficit hyperactive disorder.

II. Overview

Furthermore, the study found internet of things can apply too many researches.

A. Internet. The internet of things(IoT) and wearable device

Attal et al. [2] introduced an audit of various characterization strategies used to see human exercises from wearable inertial sensor data. They explain activity recognition process: sensors placement, data pre-processing and data classification and manage classification techniques by k-NN, SVM, GMM, and RF. Figure 1 describes the different step of the action acknowledgement prepares utilising highlights.

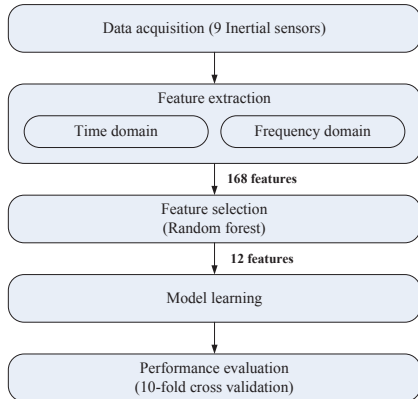


Figure 1. Step of the action acknowledgment prepare utilizing highlights extraction and determination

Uslu et al. [4] shown an activity description and discovery instrument for real-time activity monitoring using wearable sensors and hybrid classifiers. A single sensor single classifier model is displayed for the recognition of straightforward and composite activities. Next, the model is improved with different sensors and classifiers with the end goal of constant observing. The Multi-Sensor Multi Classifier (MSMC) model take two wearable TI Chronos watches with a built-in tri-axial accelerometer for information securing and an arrangement of naive Bayes, Susan Corner Detector(SCD) and Hidden Markov(HMM) classifiers for the recognition of moves between characterized activities continuously. Shinde et al. [5]

built application to recognizes circulatory strain, ECG, Heart rate and temperature. It is conveys to screen human exercises through wearable sensors. Thusly, data is watched and accumulated at adjaget and moreover remote user by internet of things (IoT). Figure 2 shows measurement of ECG that used Wilson Electrode System.

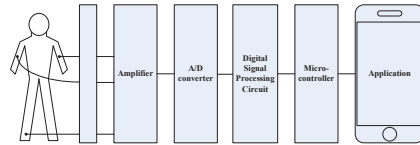


Figure 2. ECG diagram

B. Human activity monitoring

Uddin et al. [6] shown a generic framework to continuously monitor users' daily activities. The framework proposes light calculation errands on the wearable device to lessen the measure of information conveyed between the wearable, and its host. A 9-axis wristbands are being utilized to gather client's exercises. The gathered signs are liable to light weight preprocessing and division on the wearable gadget earlier sending to the host, were it experiences action identification calculations. Figure 3 explains warble sensing framework to do human activity recognition.

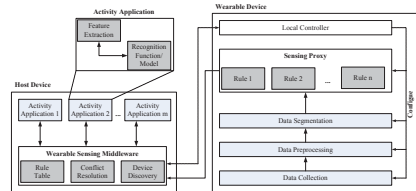
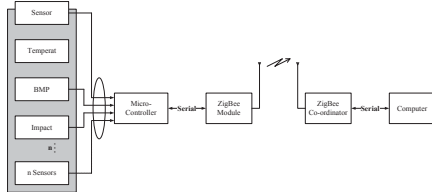


Figure 3. Wearable Sensing Framework

Mukhopadhyay[7] reported that the social insurance framework is experiencing a change in which ceaseless checking of occupants is conceivable even without hospitalization. The headway of detecting technologies, embedded frameworks, remote correspondence innovations, nanotechnologies, and scaling down make it conceivable to create keen frameworks to screen exercises of individuals consistently. Wearable sensors distinguish anomalous and/or unanticipated circumstances by checking physiological parameters alongside different

manifestations. Along these lines, essential can be given in times of desperate need. The research audits the most recent reported frameworks on movement checking of people in view of wearable sensors and issues to be tended to handle the difficulties. Figure 4 The block diagram representation of the Human Activity Monitoring (HAM) system



**Figure 4.** The Human Activity Monitoring (HAM) system

Selvaprabhu et al. [8] created a mobile medical application to recognize constant checking and deal with the human body blood glucoses levels. It guarantees to provide for amplifying the life time of human being. The system is trying gadget extremely productive to screen the glucoses levels. Sensor can wear anyplace in our human body as like as wrist watches. A sensor associates with advanced cells utilizing application. Once associated this product persistently to screen the blood glucoses levels in our body.

### C. Attention Deficit Hyperactive Disorder

Searight and Shinabarger [9] explained that Attention deficit hyperactivity disorder (AD/HD) is a condition ordinarily first clear in early youth. Indications incorporate shortfalls in consideration, focus, and fleeting memory. Behaviorally, kids with ADHD are excessively dynamic and not able to stay situated, exceptionally distractible, and imprudent. Worries about ADHD frequently at first emerge amid kindergarten or first grade, since these shortfalls essentially hinder scholastic execution and are problematic in a common classroom. Be that as it may, ADHD is progressively seen as a deep rooted condition with useful disability stretching out all through adulthood.

### III. Challenges

Technology and standards for the Internet of Things will be more developed management and security capabilities that are essential in the development of Internet of Things applications and will be an important component in the Device Mesh and Ambient User Experience. Researchers are interested in adopting a wearable device application tasks. However, we did not find research that brings wearable devices into the physical storage to analyse the possibility of having ADHD. At this point, we aim to combine internet of things technology with problems, to solve problems. Particularly, it will cause frameworks and applications that are used to collect information of biological system and analyse the possibility of the occurrence of ADHD. Accordingly, Attention deficit hyperactivity disorder symptoms are mostly behavioral expression by which to require observe symptoms. The symptoms do not appear to know the exact cause of a medical treatment, it might be late. According to the American Psychiatric Association [1], ADHD can be characterized by practices showed. People with ADHD display blends of the accompanying practices:

- Wiggling with hands or feet or squirming in their seat;
- Trouble staying situated when required to do as such;
- Trouble maintaining consideration and sitting tight for a turn in assignments, amusements, or gathering circumstances;
- Exclaiming answers to questions before the inquiries have been finished;
- Trouble finishing on directions and in sorting out assignments;
- Moving starting with one unfinished movement then onto the next;
- Neglecting to give close consideration regarding points of interest and evading rushed errors;
- Losing things essential for undertakings or exercises;
- Trouble in listening to others without being diverted or hindering;
- Wide ranges in emotional episodes; and
- Extraordinary trouble in deferring delight;

Children with ADHD demonstrate distinctive blends of these practices and commonly display conduct that is ordered into two primary classes: poor managed

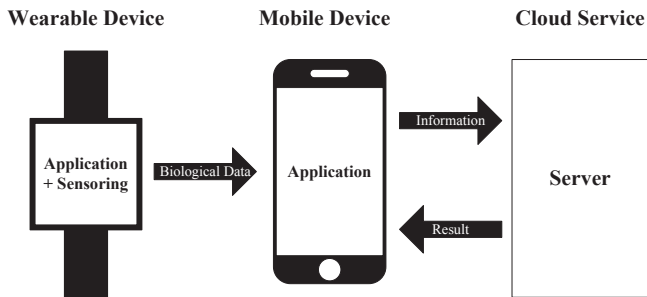


Figure 5. Children Activity Monitoring Framework

consideration and hyperactivity-lack of caution. As a result, researcher design the process of this research by create application to record information of biological such as heart rate, body movement and temperature. All information will record from wearable device, then save to cloud server. After application collect information complete, the system will analyze informations and compare with ADHD evaluation. Then, the system will report probability of ADHD symptom. In other words, The possibility of this research is based on several components, such as wearable devices which are used, the expression of a child's behaviour, applications that can be processed correctly, and other factors. Hence, If this research is successful, it will result in changes to the health education because the diagnosis of ADHD must be used to observe the symptoms of the disease by experts. Parents who are unsure of the symptoms of the disease, this can be a device and application of this test with their children. In education, it will help predict outcomes of child development, learning how powerful it is. Teachers can use the device and application to monitor how students are responding to the content or not. Figure 5 Children activity monitoring framework.

#### IV. Conclusion

In conclusion, the purpose of this research is to monitor information of biological system, activity or behavior of human to analysis symptom of attention deficit hyperactive disorder(ADHD). The internet of things technology can be used to benefit immensely. Researchers anticipate that this research will help in the education and health of children. If we can recognise the symptoms early, we can cure the illness of a child

faster. Also, education can use it to evaluate the effectiveness of learning in students.

#### References

- [1] [1] American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 5th edition. Arlington, VA., American Psychiatric Association, 2013.
- [2] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou and Y. Amirat. Physical Human Activity Recognition Using Wearable Sensors. *Sensors*. 2015; 15(12):31314-31338.
- [3] [3] GSMA, "Understanding the Internet of Things (IoT)", July 2014.
- [4] G. Uslu, H. I. Dursunoglu, O. Altun and S. Baydere. Human Activity Monitoring with Wearable Sensors and Hybrid Classifiers. *International Journal of Computer Information Systems and Industrial Management Applications*. ISSN 2150-7988 Volume 5 (2012) pp. 345-353
- [5] K. D. Shinde , S. B. Shinde and S.S. Savkare. A Review on system of IOT Based Human Activity Monitoring By Using Raspberry Pi. *International Journal of Advanced Research in Science Management and Technology (IJARSMT)*. Volume-2, Issue-6, June-2016
- [6] M. Uddin, A. Salem, I. Nam and T. Nadeem. Wearable Sensing Framework for Human Activity Monitoring. *The 2015 workshop on Wearable Systems and Applications*. pp. 21-26. 2015
- [7] S. C. Mukhopadhyay, Wearable sensors for human activity monitoring: A review, *IEEE Sensors Journal*. Volume: 15, Issue: 3, pp. 1321-1330, 2015.
- [8] S. Selvaprabhu, S. Ashok Senthil Kumar and A. Subash Chandar. Human health control monitor system using smart mobiles: Context changes dependent human behavior. 2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM). 30-31 March 2016
- [9] H Searight, C. Shinabarger. Attention Deficit Hyperactivity Disorder. In: Feldman MD, Christensen JF, Satterfield JM. eds. *Behavioral Medicine: A Guide for Clinical Practice*, 4e. New York, NY: McGraw-Hill; 2014.

# Empowerment of Autistic Children in Learning Experience of Life Skills and Society Adaptation through Digital Interactive Online Media

Banphot Nobaew, PhD  
 School of Information Technology  
 Mae Fah Luang University  
 Chiang Rai, Thailand  
 banphot@mfu.ac.th

Weeranut Intagan, PhD  
 Department of Mechanical Engineering  
 Silpakorn University  
 Nakhon Pathom, Thailand  
 weeranut\_n@hotmail.com

**Abstract**— This paper aims to propose the conceptual framework of developing the learning experience in autism children for society adaptation and real life skills by applying the digital interactive online media (DIOM) into the self-learning process. The framework considers the different settings of social scenarios from the critical real-life situation such as natural disaster, cooking practices and dangerous animal; these scenarios are integrated into a content of the digital media and applied to the informal learning course. This paper studies in different types of multimedia online application which are interactive media and online game; these multimedia applications have been applied to the content of life skills and society adaptation. The learning experience and interactivity pattern between autism children and digital media are the main focuses of this study. After practising with the interactive digital media the autism children will be tested with the real-life simulation and test further with the real-life situation.

**Keywords**— empowerment; autistic children; learning experience; life skills; society adaptation; digital interactive online media

## I. INTRODUCTION

A world population has confronted with many disease problems; autism prevalence is the most critical problem which causes human health and potential development. This disease can occur in all human sexes, races, regions and even different socioeconomic groups [1]. The Centers for Disease Control and Prevention reported that the world population has the amount of autism spectrum disorder (ASD) to 1 percent in 2014; it has strongly increased by 6-14 percent from 2000 to 2012 [2]. The autism prevalence can be found in 1 of every 40 female and 1 in every 70 male of the world population. National Research Studies show that the four countries with the highest rates of autism are Japan, United Kingdom, Sweden and Denmark; they are found 161, 94, 72 and 68 cases per 10,000 children studied respectively [3]. Autism rate in Africa is rarely found; Somalia had autism children less than 1 percent of the population in 1978, it usually occurred with the elite family or side effect from other diseases [4], [5], [6]; however, it increases to 4 percent of student [7]. This statistic of the

prevalence of autism children around the world in 2016 displays in the fig. 1.

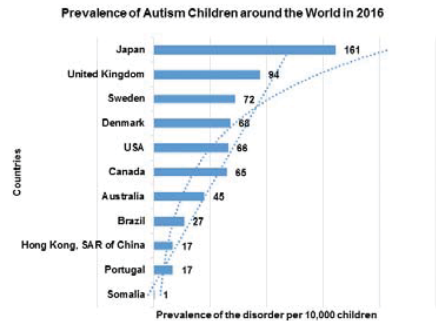


Fig. 1. Prevalence of Autism Children in 2016.

This disease causes human disability than other diseases: AIDS, diabetes and cancer. The people who have autism disease; they do not only get the effect on their disability body and mind but it also the effects of co-occurring conditions such as the medical care costs. Peacock and other [8] studied the medical care cost for the autism disease; they found that the medical care costs can come from the three distinctive effect combinations which are attention-deficit/hyperactivity disorder, intellectual disability and epilepsy disease. The average annual medical cost for autism person were \$10,709 in 2005. Peacock compared the annual cost of the three distinctive effect combinations; this result can be showed in the following diagram on the fig. 2.

According to the diagram fig. 2 the medical care cost per child for the autism children with intellectual disability and autism children with attention-deficit/hyperactivity disorder

have the highest value \$19,200 and \$11,900 respectively. It has a higher cost than children without ASD more than 10 times[8].

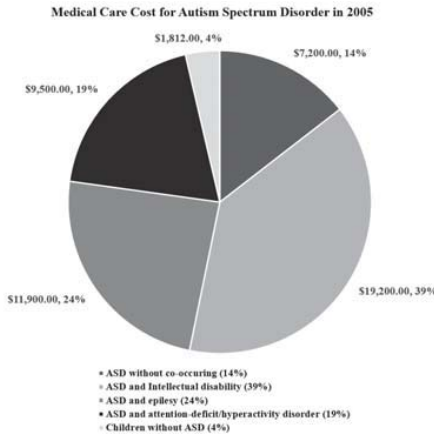


Fig. 2. Comparing the annual medical care cost per child.

The medical care cost for children with autism spectrum disorder has higher than children without ASD; moreover, ASD treatment needs more budget for intensive behavioural interventions. The total medical cost is about \$50,709 to \$70,709 per child per year [2]. In the UK. the costs of supporting for ASD children were estimated about \$3.29 billion each year [9]. However, if the autism disease can be found and the intervention for treatment can be adapted early this can be decreased the lifelong care cost by 2/3 of the total cost [10]. The increasing of early diagnosis and awareness of the developmental disorder can be applied to reduce the population of human disability. The autism diagnosis can be applied when children has age 2-year-olds for reliable detection and valid result [11]. The intervention process can be implemented into the learning process at the early stage of autism disease such as special learning course and content, suitable learning media and monitoring program. Many studies applied the multimedia into the intervention program to help the autism children. T. Tjus, M. Heimann, and K. E. Nelson [12] used the multimedia computer program for enhancing language and reading development in autism children; they found that the intervention can help the autism children to improve reading and language development skills. The autism children can get the benefit from the combining of a motivating multimedia program and positive interaction with a professional teacher. Besides, the interaction patterns between the teacher and learner must associate to the language level of the learner [13]. Hetzroni and Tannous [14] investigated the use of computer-based intervention for enhancing communication functions of autism children on daily life

activities; the result indicated that the intervention in which opens the opportunities for children to interact with everyday activities it will help them to gain knowledge and learn from the natural classroom environment. However, the concept of Hetzroni and Tannous still have a limitation of applying into real life situation since their study had been focused on the simulation of the classroom setting. Besides, the situation of the study conducted in the daily activities which are not much socialised with the communities. Nevertheless, when applying to the everyday life situation, the real community environment has different factors and more complex activities; it will cause a lot of problematic adaptation

Many studies are similarly found that using the computer-assisted instruction for the autism children can help them to develop their potential abilities [15], [16], [17], [18].

This study is developed based on the previous investigation by further extending to the practical real life situation. The previous investigation studied and focused on the utilising of interactive instructional application to enhance learning in autism child [19]; it applied the game-based learning framework to improve the communication skills and social adaptive skills in the moderated autism children. The study focused on the potential interactive media in order to increase the cognitive skill and measured the efficiency of tasks completing; the result showed the improvement skills of autism child both cooperative motor and problem-solving skills more than 50%, but they did not improve much in social adaptive skill in life situation. Additional result suggestion that the autism children can do the application tasks in 50%; however, many of them still need help from assistance during the tasks. According to the studied, the experimental investigation needs a professional assistance to help and guide the learner in the beginning of learning process and during the process. To achieve the increasing of life skill development, the extending study requires designing more practising in a structured setting of real life situation.

This paper purposes the conceptual framework of developing the learning experience in autism children for society adaptation and real life skills by applying the digital interactive online media (DIOM). The study aims to use the digital media to provoke the soft skills development such as speaking, thinking process and doing. In addition, the framework will be used as a supporting tool for parents and educator who work and involve the autism children.

## II. CONCEPTUAL FRAMEWORK

This study is inspired by two previous theoretical concepts: the theory of mind (ToM) which states to the idea that the autism person does not understand about the different between persons that people have different beliefs, ideas, points of view, and feelings[14]; and the theory of rare event learning (REL) which refers that it is very rare opportunity to combine all significant conditions to enhance the highest learning efficiency for disability person[12], [13]. These two theoretical concepts help to frame an idea of a situation-based learning by structuring and controlling the natural classroom setting. The digital media can be used to simulate the learning environment of daily life.

This study based on the assumption that the design interactive digital online media that are appropriate to a learning pattern of autism spectrum disorder can help and support foster self-learning independently. This continuity of self-learning can create the process of learning the pattern and leads to adaptive behaviour patterns of students to be able to adapt to the environment and society and live as normal people.

A. Conceptual Framework

Fig 1. illustrates conceptual framework of learning experience in life skills and society adaptation of autism children to support independently self-learning. The real life scenarios are the main concerning of learning outcome in this study.

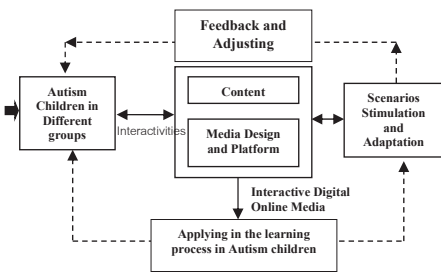


Fig. 3. Conceptual Framework of Learning Experience of Life Skills and Society Adaptation in Autism Children.

The framework of Fig. 3 will be used to make the intervention for ASD in order to recover disabilities; this can be applied to the self-learning process. This can be done in a formal classroom and a home school. The autism children can learn the contents from the interactive digital media both online and offline platforms. All contents are basically important issues for everyday living such as how to make an omelette, dish washing, how to behave when disaster strike and how to behave with dangerous animals. The contents are applied into the appropriate visual design in different media types such as interactive media and game. These media are suitable for children in the cities and the remote area when the connection unsteady. The content will be designed as real-life scenarios and later the autism learner will be tested with the real situation. The learning outcome will be assessed through the learning process in order to re-design and adjust the design of the contents

III. METHODOLOGIES

The fifteen children with autism will be selected as the investigation participants; the selection process will use the autism diagnosis: Childhood Autism Rating Scale. This process will proceed by the psychologist. The diagnostic assessment

process requires interviews, screening measures, parent surveys and personal data recording review. [20]

This study focuses on designing the digital interactive media for adaptive behaviour self-learning in autism children through the visual grammar design in the multimedia applications. The visual design of digital interactive media associated with the learning content and learning pattern including the nature of learners. The study aims to find the optimal design of interactive digital media in each media category relative to conventional learning such as content knowledge, elements of design: elements of design; visual presentation; sound and interactivity; and an appropriate period of presentation including the experience of learning. The methodological investigation can be described as the following fig. 4. diagram.

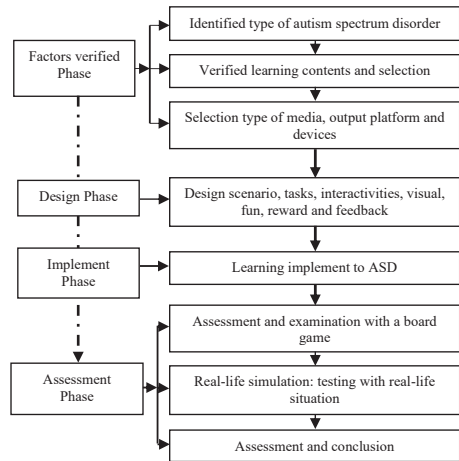


Fig. 4. Methodological diagram

The study consists of four phases: factor verified, design, implement and assessment phases. The first phase is to verify the type of autism children in term of learning potential and constraints. This will be used to design the levels of knowledge content: how much is difficult, self-learning timing, type of media and media platform. The second phase focuses on the content and visual design including tasks, interactivities between learner and the media, and feedback. The completed media design will be implemented to the learning process in the different group of ASD; the feedback data and the interaction will be gathered for assessment phase. During the media implementation, some media require being re-designed to suit the learner and learning style. The assessment will be applied in two ways: using the board game to make a simulation test; this be used to check the practical understanding and real applying in the simulation scenario. The board game can be various type of easy game like a puzzle

game or sorting game. The second assessment is let the autism children testing with the real-life situation, for instance, let them make the omelette.

Implementing this framework in the investigation it requires to apply various research tools for data collection and examination both statistical analysis and quality analysis. Quasi-experimental design is utilised for this research framework; the appraisal tools are testing, pre- and post-test, Paired t-tests, Multivariate Analysis of Variance (MANOVA). The research tools for data collection are Autism Diagnostic Interview (ADI), questionnaire, observation, video recording and recording software.

#### IV. RESULTS AND DISCUSSION

This conceptual framework of developing the learning experience in the autism children is developed based on the initial studies of the game-based learning for the autism children and kids without ASD; these previous studies are adopted and adapted from P. Pitchapa, B. Nobaew, P. Rungrueng, and R. Nobnop [19], and Nobaew[21]. To implement the conceptual framework of this paper into other media with autism children studies, a researcher should adopt and adapt a Randomised Treatment Trial design in the process of research development; additionally, the research design requests to repeat an intensive investigation on different groups of autism children and independent samples. The framework implementation requires investigating in a focus group and monitor in a length of time.

Although this conceptual framework is developed from the investigation of the case study of children with autism in Thailand, it can be applied to the other region by verifying the initial factors such as cultural factor, learning context, real life environment and diagnostic assessment. This framework can be also utilised as the initial guideline for learning empowerment in other disabilities.

#### V. CONCLUSION

The problem of increasing amount autism spectrum disorder (ASD) of the world population is growing rapidly; this is the main threat to human race development. The early diagnosis and intervention are necessary to implement in the first stage of applying this conceptual framework to the experimental study. This study is a part of the intervention process to help the abilities development of autism children to adapt to the society and social environment. The conceptual framework consists of four practical phase of implementation. The teachers are required in the learning process as well as the interactive online media which is necessary for transferring knowledge to the autism children. The learner monitoring and assessment have to be applied as in the implementation phase.

#### ACKNOWLEDGMENT

I would like to thanks Mae Fah Luang University who granted the budget for supporting this paper, and many thanks to all research staffs for the initial media studies in children with autism.

#### REFERENCES

- [1] D. L. Christensen *et al.*, "Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012," *MMWR Surveill. Summ.*, vol. 65, no. 3, pp. 1–23, Apr. 2016.
- [2] "Facts About ASDs," *CDC - Facts about Autism Spectrum Disorders - NCBDDD*. [Online]. Available: <http://www.cdc.gov/ncbddd/autism/facts.html>. [Accessed: 11-Oct-2016].
- [3] "Countries With The Highest Rates of Autism," *WorldAtlas*. [Online]. Available: <http://www.worldatlas.com/articles/countries-with-the-highest-rates-of-autism.html>. [Accessed: 11-Oct-2016].
- [4] V. Lotter, "Childhood autism in Africa," *J. Child Psychol. Psychiatry*, vol. 19, no. 3, pp. 231–244, 1978.
- [5] M. O. Bakare, P. O. Ebigo, and V. N. Ubochi, "Prevalence of autism spectrum disorder among Nigerian children with intellectual disability: a stopgap assessment," *J. Health Care Poor Underserved*, vol. 23, no. 2, pp. 513–518, 2012.
- [6] "Statistics," *UNICEF*. [Online]. Available: [http://www.unicef.org/infobycountry/somalia\\_statistics.html](http://www.unicef.org/infobycountry/somalia_statistics.html). [Accessed: 12-Oct-2016].
- [7] "Autism Statistics Alarm African Children," *Mercola.com*. [Online]. Available: <http://articles.mercola.com/sites/articles/archive/2008/09/20/autism-statistics-alarm-african-children.aspx>. [Accessed: 12-Oct-2016].
- [8] G. Peacock, D. Amendah, L. Ouyang, and S. D. Grosse, "Autism spectrum disorders and health care expenditures: the effects of co-occurring conditions," *J. Dev. Behav. Pediatr.*, vol. 33, no. 1, pp. 2–8, 2012.
- [9] M. Knapp, R. Romeo, and J. Beecham, "Economic cost of autism in the UK," *Autism*, vol. 13, no. 3, pp. 317–336, May 2009.
- [10] R. E. Mankoski, M. Collins, N. K. Ndozi, E. H. Mgalla, V. V. Sarwatt, and S. E. Folstein, "Etiologies of autism in a case-series from Tanzania," *J. Autism Dev. Disord.*, vol. 36, no. 8, pp. 1039–1051, 2006.
- [11] C. Lord, S. Risi, P. S. DiLavore, C. Shulman, A. Thurm, and A. Pickles, "Autism From 2 to 9 Years of Age," *Arch. Gen. Psychiatry*, vol. 63, no. 6, p. 694, Jun. 2006.
- [12] T. Tjus, M. Heimann, and K. E. Nelson, "Gains in Literacy through the Use of a Specially Developed Multimedia Computer Strategy: Positive Findings from 13 Children with Autism," *Autism*, vol. 2, no. 2, pp. 139–156, Jun. 1998.
- [13] T. Tjus, M. Heimann, and K. E. Nelson, "Interaction Patterns Between Children and their Teachers when Using a Specific Multimedia and Communication Strategy: Observations from Children with Autism and Mixed Intellectual Disabilities," *Autism*, vol. 5, no. 2, pp. 175–187, Jun. 2001.
- [14] O. E. Hetzroni and J. Tannous, "Effects of a computer-based intervention program on the communicative functions of children with autism," *J. Autism Dev. Disord.*, vol. 34, no. 2, pp. 95–113, 2004.
- [15] K. Higgins and R. Boone, "Creating Individualized Computer-Assisted Instruction for Students with Autism Using Multimedia Authoring Software," *Focus Autism Dev. Disabil.*, vol. 11, no. 2, pp. 69–78, Jan. 1996.
- [16] T. Hagiwara and B. Smith Myles, "A Multimedia Social Story Intervention: Teaching Skills to Children with Autism," *Focus Autism Dev. Disabil.*, vol. 14, no. 2, pp. 82–95, Jan. 1999.
- [17] T. R. Goldsmith and L. A. LeBlanc, "Use of technology in interventions for children with autism," *J. Early Intensive Behav. Interv.*, vol. 1, no. 2, p. 166, 2004.
- [18] M. Heimann, K. E. Nelson, T. Tjus, and C. Gillberg, "Increasing reading and communication skills in children with autism through an interactive multimedia computer program," *J. Autism Dev. Disord.*, vol. 25, no. 5, pp. 459–480, 1995.
- [19] P. Putthajaroen, B. Nobaew, P. Rungrueng, and R. Nobnop, "Interactive Instructional Application for Enhance learning in Autistic Childs," presented at the The International Conference on Digital Arts, Media and Technology (ICDAMT), Chiangrai, Thailand, 2016.



- [20] J. L. Matson, *Handbook of Assessment and Diagnosis of Autism Spectrum Disorder*. Springer, 2016.
- [21] B. Nobaew, "The comparative study of multimedia technological applications enhancing active self-learning in online course," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on*, 2016, pp. 1–5.

# Mobile Surveillance Application for Hybrid WiMAX-Wi-Fi Systems

C.S Luboby, M. E. Dlodlo, G. De Jager  
 Department of Electrical Engineering

University of Cape Town  
 Cape Town, South Africa

lbbsma001@myuct.ac.za, mqhele.dlodlo@uct.ac.za,  
 gerhard.dejager@uct.ac.za.

A. Zulu

Department of Electrical & Electronics Engineering

University of Zambia  
 Lusaka, Zambia

ackim.zulu@unza.zm

**Abstract**— This paper proposes and implements an algorithm for mobile surveillance application software ideal for a Smartphone operating in low bandwidth hybrid WiMAX-Wi-Fi and cellular network environments. Application software is a standalone program that solves a specific business or security need. Results show that our proposed algorithm and resulting implementation successfully compresses and transmits surveillance video/image to the server making it suitable for use in limited bandwidth environments.

**Keywords**— *WiMAX-Wi-Fi; Mobile Surveillance; Software applications; Smartphone*

## I. INTRODUCTION

A hybrid WiMAX-Wi-Fi video surveillance system [1] consists of a Base Station (BS) which is connected in a point-to-point configuration with the Customer Premises Equipment (CPE) which is also called the subscriber station (SS) [2]. The CPE has at least two wireless interfaces namely the Wi-Fi and WiMAX wireless interfaces. Some CPE have additional Ethernet interfaces for wired connections. The mobile devices with camera functionalities connect to the CPE using its Wi-Fi interface. The Wi-Fi interface is compatible with the common IEEE 802.11 Wi-Fi standards such as IEEE 802.11a/b and IEEE 802.11g. The base station then connects to the Internet, routers and video server as shown in Fig.1.

Application software is a standalone programme that solves a specific business or security need. The software development models are carefully chosen processes or methodologies for the development of the project in order to achieve project's objectives and goals. Usually, these models specify the various stages in the processes or methodologies including the order in which they should be carried out. Some examples of software development model include the waterfall, spiral, iterative and incremental software development models as discussed in [3] [4] [5].

we propose and implement an algorithm to achieve a mobile surveillance application software that can be installed in Smartphone operating in the WiMAX -Wi-Fi and cellular network low bandwidth environments.

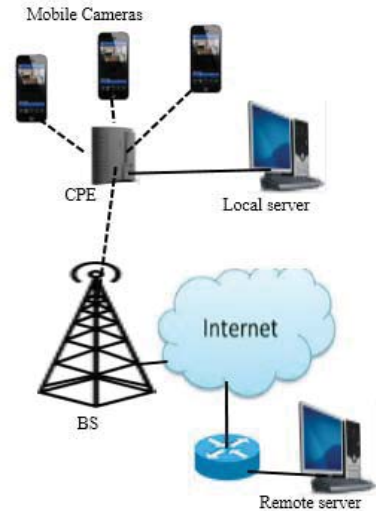


Fig.1: Mobile Hybrid WiMAX-Wi-Fi Video Surveillance System

The rest of paper is organised as follows: the next section discusses related work on mobile security application software, section III discusses existing software development process model and ends with our suggested software development process. Section IV, gives our proposed and implemented software algorithm. Section V explains the implementation methodology and assumptions made including working constraints. In section VI, results and discussion of the developed application software is given while the conclusion follows in section VII.

## II. RELATED WORK

Shinde *et al* [6] proposed a classic approach to security provision for a home based system robot and environment surveillance whereby any changes in the movement of the robot could be viewed via the camera mounted on the robot or

ceiling. Similarly, any intrusion attacks would be received to the human operator through SMS alerts or Internet based services on the Android phones. The Cisco Video Surveillance Operations Manager Mobile App allows one to view live video from a mobile device, such as an Android-based tablet or phone [7]. The Wire path IP Surveillance app allows one to view surveillance videos from encoders, IP cameras, and NVRs on an Android or iOS Smartphone or tablet [8]. A recent paper on video surveillance with intrusion detection via a mobile device was written by [8]. They set an electronic system designed to observe an area from a distance by means of electronic equipment with enhanced ability to automatically detect the presence of any moving body into the region being observed (watched). The system had extra capability to provide users with remote access to visual display on a mobile device via the internet video streaming [9].

In this work we show how mobile wireless surveillance application software was developed for the hybrid WiMAX-Wi-Fi and Cellular low bandwidth network environments.

III. EXISTING SOFTWARE DEVELOPMENT MODELS

Application software is a stand-alone program that solves a specific business or security need. Application software are tailored processes for businesses or technical information to ease the business. The software development models are carefully chosen processes or methodologies for the development of the project in order to achieve project’s objectives and goals. Usually, these models specify the various stages in the processes or methodologies including the order in which they should be carried out.

The testing techniques of the developed software are largely determined by the type of model selected. Several software development life cycle models have been developed; each with a specified achievable objective. They are classified as traditional and modern software development models. Traditional software development models include the Waterfall model, V model, Incremental model, RAD model, Agile model, Iterative model and the Spiral model. A description of some of these models is given below.

A. Waterfall Models

The Waterfall Model, also known as linear-sequential life cycle model was the first software development Process Model to be introduced [4]. Each phase in the waterfall model must be completed fully before the next phase can begin. This type of model is ideal for small projects with no uncertain requirements. Reviews are made at every stage to determine whether the project is still viable or should be discarded. Testing is performed at the completion of the project. Fig.2 shows the various stages of the waterfall models.

While this model has certain advantages such as simplicity and easy management due to its lean size, it also has weaknesses. The model is not flexible to changes when it reaches the testing stage and has no overlap stage [3] [10]. As such, this makes it risky and uncertain. The model is also not ideal for complex and object oriented projects.

This model has the advantage that additional software functionality can be included with time during the

development and implementation stages. The model also allows early software production during the life cycle of the software. However, the model needs expertise to do the risk analysis and success depends on positive gains in the risk results. It therefore does not work well for small projects.

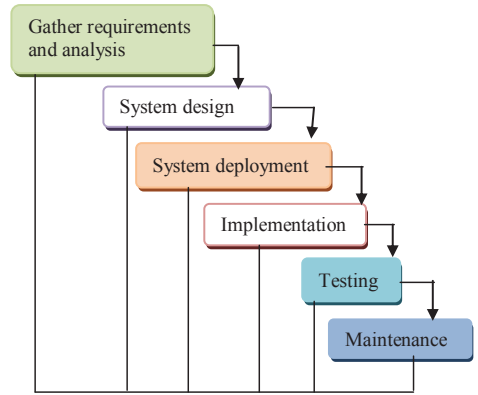


Fig.2: Waterfall Model

B. The Spiral Software Development Process Model

The spiral model places more emphasis on risk analysis [3] [10] and is similar to the incremental model. A software development process in the spiral model passes through four stages in iterations: Planning, Risk Analysis, Engineering and Evaluation [4]. In the planning stage, project requirements are assembled and risks involved assessed. Both business and system requirement specifications are determined. Each subsequent spiral builds on the baseline spiral. Fig.3 shows the spiral software development process model.

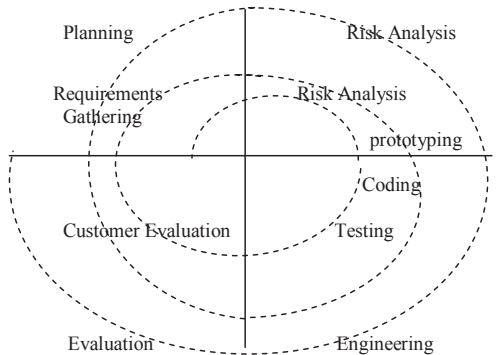


Fig.3: Spiral Software Development Process Model

Next, the model identifies risks and possible solutions. This is followed by prototype production after which alternate solutions are suggested for implementation [4]. The engineering stage is the stage at which the software is developed and tested while the evaluation stage allows for evaluation of the software and project and the spiral continues.

In designing the mobile wireless surveillance application software, we followed the three stage process model: modelling, construction and deployment.

The modelling stage involves the creation of models that allows the developer and the user to better understand the software requirements. It allows users and designers to formulate the problem and state software needs or requirements.

The construction stage involves code generation, compiling, testing and debugging. The central task in this process is coding. The codes must take into consideration the problem statement or requirements specifications identified in the modelling process. There are several errors detected during compiling of codes and testing. To detect these errors debugging is performed. Debugging is a process of problem analysis and resolution [10]. Once the problem has been identified and corrected, the code is modified to reflect the corrections made and the code is re-compiled.

The codes are then tested and when the code passes, the code is released to prepare for product deployment.

IV. MOBILE SURVEILLANCE ALGORITHM MODEL

Fig.4 shows the proposed and implemented mobile wireless surveillance model algorithm. A smart phone captures surveillance video/images using the camera that comes with the phone.

The video/image quality at this stage is high as the video is uncompressed. Due to the limited bandwidth in the wireless link, the video images are compressed and kept in the image gallery before uploading or sending them to the server for monitoring and viewing. One can choose to upload and send these images in its uncompressed format at the cost of added bandwidth and /or data units if a 3G or 4G mobile cellular network link is used.

The developed mobile wireless surveillance software provides a mechanism for selecting the video and/ or image to be sent to the server. Once the image has been sent to the server, it is viewed after the decompression process. The developed software allows the user to state the location of the captured image for easy identification during monitoring and viewing at the server end.

The location name also serves as the image/video name and helps to guide law enforcement officers in identifying the crime scene.

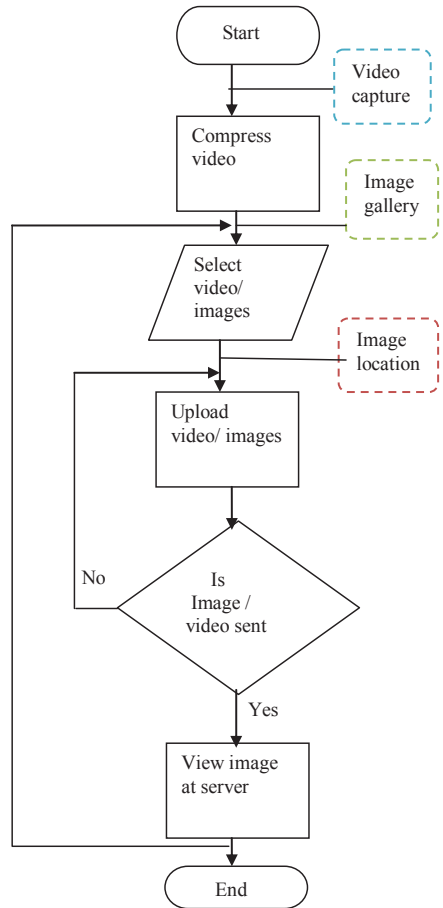


Fig.4: Mobile Wireless Surveillance Algorithm

V. IMPLEMENTATION METHODOLOGY

In developing the software application windows 7 enterprise, 64-bit operating systems, desktop computer with an Intel (R), core (TM) i5 CPU processor and 8GB Random Access Memory ( RAM) was used . The computer was installed with java Software Development Kennels (SDK) and Android Studio version 1.5.1 application development IDE. Android Studio is Google’s newest solution to many of our Android development woes [11]. The Android Studio IDE is specifically designed for Mobile phones, tablets and similar smart devices. It operates under the java programming language and the eXtensible Mark-up Language (XML). The XML is an open standard providing the means to share data

and information between computers and computer programs as unambiguously as possible [12]. Supported platforms include Windows, Mac OS X and Linux.

To test the codes and application, the USB debugging mode setting for the smart phone is activated under the developer options. A USB cable is connected to the Desk top personal computer and a real Android smart phone, Android version 4.4.2, model SM-G313H. The Smartphone is then activated for transmission on the WiMAX-Wi-Fi network and/or the mobile data option for transmission over the cellular network.

*A. Assumptions and Constraints*

In designing and developing the mobile security application software, the following constraints have been considered: firstly, most Smartphone have limited memory to store large files of video images. Secondly, existing smart phone encoders are of the H.264/AVC non scalable types which possess a further strain on camera memory and storage. The non-moving Images are encoded mainly using JPEG standard. Further since the application is intended to be used in low bandwidth wireless network environment, it becomes reasonable to tailor our application for JPEG images. We assume that the local and/ or remote server has adequate space to store the transmitted surveillance images and that both real and non-real time surveillance images are transmitted.

VI. RESULTS AND DISCUSSION

To test the codes and application, the USB debugging mode setting for the Smartphone is activated under the developer options. A USB cable is connected to the Desk top personal computer and a real Android Smartphone, Android version 4.4.2, model SM-G313H. The Wi-Fi network is then activated for transmission on the WiMAX- Wi-Fi network and/or the mobile data option for transmission over the cellular network.

The application is then run and a pop-up window shown in Fig.5 appears with an option to test on real or emulator device.

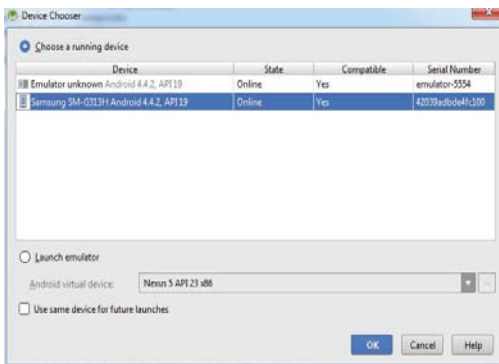


Fig.5: Testing the Codes and Mobile Surveillance Application

The real device is selected and the OK button is clicked. The choose image/video button is then pressed and an image/video selected from the image/video gallery. Finally, the upload button is pressed and the image is sent to a specific IPv4 IP address in the server.

When the image is uploaded a “successful upload” message pops up as shown in Fig.6 to indicate the image has been sent. The sent image can be sent to either a local or remote server/viewing PC or to both. In either server options, a specific video/image folder was created for the reception and verification of all surveillance images/videos transmitted by Smartphone. The java codes specify the destination IP address for the server(s) and this address is unknown by Smartphone users. Images are decompressed, stored and named according to the name given at transmission.

We then installed this application to ten other android Smartphone. To achieve this, application software is first converted into an APK format. The APK format is the executable file format for Smartphone devices. This format allows the surveillance application to be easily installed or uninstalled.

After installing the mobile wireless surveillance software to randomly selected smart phone users, the users were allowed to use the application by sending images to the server. The Smartphone is also used as a camera for capture of surveillance video/images.



Fig.6: The Application Showing a Successful Transmission of the Surveillance Image

VII. CONCLUSIONS

This paper has given related work on surveillance application software. The paper has also proposed and implemented an algorithm for low bandwidth mobile wireless surveillance application software. The developed and implemented algorithm allows compression of raw video and enables transmission of the compressed video in low bandwidth wireless network environment. The developed software has been tested on real Smartphone devices and on an active WiMAX-Wi-Fi and cellular network. It has performed as envisaged.

ACKNOWLEDGMENT

This work is supported by the University Science, Humanities and Engineering Partnerships in Africa (USHEPIA) Fund.

REFERENCES

[1] S. Banerji and R. S. Chowdhury, "Wi-Fi & WiMAX: A Comparative Study," *Indian J. Eng.*, vol. 2, no. 5, pp. 1–5, 2013.

[2] K. Dagar and P. Sharma, "Performance of Internet Protocol TV over WiMAX," vol. 4, no. 6, pp. 2507–2513, 2015.

[3] D. Jamwal, "Analysis of Software Development Models," *Int. J. Comput. Science Technol.*, vol. 1, no. 2, pp. 61–64, 2010.

[4] N. M. A. Munassar and A. Govardhan, "A Comparison Between Five Models Of Software Engineering," *Int. J. Comput. Sci. IJCS/vr*, vol. 7, no. 5, pp. 94–101, 2010.

[5] M. I. Daud, "Secure software development model: A guide for secure software life cycle," *Proc. Int. MultiConference Eng. Comput. Sci.*, vol. 1, 2010.

[6] M. Shinde, R. Shinde, N. Thavare, and A. Baviskar, "Security Based Home Surveillance System Using Android Application," *Int. J. Res. Eng. Technol.*, vol. 3, no. 4, pp. 814–815, 2014.

[7] Cisco Systems, "Cisco Video Surveillance Operations Manager Mobile App User Guide," pp. 1–10, 2013.

[8] Wirepath Surveillance, "WPS-IP Surveillance App Setup Guide." pp. 1–7, 2015.

[9] G. Ononiwu, G. Okorafor, J. Onojo, and R. Opara, "Low Cost Video Surveillance System With Intrusion Detection," *Int. J. Emerg. Technol. Res.*, vol. 1, no. 7, pp. 79–98, 2014.

[10] F. Tsui and O. Karam, "Software Process Models," *Essentials Softw. Eng.*, no. Cmm, pp. 98–101, 2010.

[11] C. Haseman and K. Grant, *Beginning Android Programming Beginning Android Programming*. 2014.

[12] E. L. Morgan, "Getting Started with XML: A Manual and Workshop," 2004.

# Maximizing LTE Network Performance by using Differential Scheduling and Carrier Aggregation

Mădălina Oproiu, Vlad Boldan

Technology Departament

Orange Romania

Bucharest, Romania

elena-madalina.oproiu@sdetib.pub.ro

Alexandru Vulpe, Ion Marghescu

Telecommunications Department

University Politehnica of Bucharest

Bucharest, Romania

alex.vulpe@radio.pub.ro

**Abstract**— Nowadays, mobile operators develop and optimize their networks for meeting user requirements and offering high data rates in order to support advanced services, applications and most importantly streaming. One of the main goals is to provide more throughput on the transmission path, but it is also important to provide a high cell capacity, so that rates are high enough for all users, anywhere. That is why in these situations choosing the right scheduling technique is a must. Moreover, using 3GPP Rel. 10 LTE-Advanced with the carrier aggregation technique, the network can be further optimized by using different scheduling techniques. The paper presents measurements made to observe the behavior of two users that simultaneously use carrier aggregation with two component carriers and differentiated scheduling with the Proportional Fair algorithm. To avoid the situation where using basic scheduling, a CA-enabled user would have wasted the resources for the secondary cell as long as there were non-CA users on it, leading to a waste of resource blocks, we studied and analyzed in our tests the differentiated scheduling algorithm. Tests were done in the Orange Romania Laboratory (Network Operational Skill Center) in the beginning of 2016.

**Keywords**— Carrier Aggregation, Differentiated Scheduling, PCell, SCell, Throughput, Bandwidth, Resource Block

## I. INTRODUCTION

Out of the many problems that wireless cellular networks face nowadays, scheduling and resource allocation is an ongoing research challenge. The emergence of LTE-Advanced with the Carrier Aggregation (CA) technique has given birth to new research challenges for scheduling and resource allocation. This is because the existence of multiple Component Carriers (CCs) gives another degree of freedom in formulating scheduling and resource allocation algorithms [1,2].

The LTE system adopts shared-channel transmissions in which time-frequency resources are dynamically shared by user equipment (UEs). eNodeBs perform scheduling to allocate these resources for uplink (UL) and downlink (DL) transmissions.

Dynamic packet scheduling and, related to it, link adaption are functions which ensure a high spectral efficiency, at the same time, supplying the necessary QoS level within the cell.

The Packet Scheduler takes scheduling decisions every Transmission Time Interval, by allocating Resource Blocks (RBs) to the users, but also by controlling transmission

parameters, including the Modulation and Coding Scheme (MCS). Therefore the minimum scheduling unit consists of 12 subcarriers (spanning 180 kHz in the frequency domain) and 1 sub-frame (1 ms in the time domain) [3].

The target of the scheduling algorithm is to maximize the cell capacity, while, at the same time, ensuring that the QoS requirements for EPS bearers are fulfilled and that there are also resources available for bearers with no QoS requirements. The scheduling decisions are taken for each user, although there might be several data flows undergoing. In order to differentiate among them, each data flow is identified by Logical Channel ID (LCID), and the MAC layer decides the quantity of data to be scheduled for each LCID [4,5].

The paper is organized as follows: Section 2 presents scheduling under carrier aggregation and differences compared to traditional scheduling in LTE; Section 3 presents the test topology used in the laboratory and the measurements made. Results are shown and discussed in Section 4 and, finally, Section 5 draws the conclusions.

## II. SCHEDULING UNDER CARRIER AGGREGATION

This section describes how scheduling works when Carrier Aggregation (CA) is enabled.

For GBR (Guaranteed Bit Rate) services, scheduling with CA enabled is almost the same as scheduling with CA disabled. The goal of scheduling is still to meet the QoS requirements of GBR services.

For non-GBR services, there are two scheduling methods: basic scheduling and differentiated scheduling. The scheduling method must be consistent between serving cells to stave off data transmission exceptions.

### A. Basic scheduling

In basic scheduling, the data rate of a CA-enabled UE is the total data rate on the CCs of the UE. Under the same channel condition and sufficient bandwidth, the average data rate of a CA-enabled UE is approximately the same as the average data rate of a non-CA UE that has services with the same QoS class identifier (QCI).

Due to inconsistent channel quality for the CA-enabled UE between the PCell and SCells, basic scheduling fails to achieve an identical data rate for a CA UE and a non-CA UE when planning to allocate the same number of physical resource blocks (PRBs) to the two UEs [6].

- If the difference in channel quality between the PCell and SCells is large (for example, when the CA-enabled UE is located at the edge of a cell), there is a large difference in the number of allocated PRBs between the CA UE and the non-CA UE.
- If the channel quality is almost the same between the PCell and SCells, the number of PRBs allocated to the CA-enabled UE is close to that of PRBs allocated to the non-CA UE [6].

**B. Differentiated scheduling**

In differentiated scheduling, the data rate of a CA-enabled UE for priority calculation is defined as the data rate only on the current CC of the UE. On each CC, the CA-enabled UE is allocated the same number of RBs as a non-CA UE on the same CC. Therefore, the number of RBs allocated to the CA UE is the sum of the average number of RBs allocated to a non-CA UE in each of the serving cells. The average data rate of the CA UE served by N CCs will be almost N times the average data rate of a non-CA UE, when the spectral efficiencies of the CCs are close, UEs are evenly distributed on the CCs, and channel conditions are the same between the CCs [6].

In differentiated scheduling, a CA UE is treated as a single UE on each CC. Therefore, a CA-enabled UE can receive more PRBs and experience higher throughput than a non-CA UE and the amount of radio resources available for non-CA UEs decreases.

**C. Scheduling algorithms**

The eNodeBs that are used in the Orange Romania network and throughout this paper support four scheduling algorithms:

- **Max C/I-** takes into account channel quality only when allocating radio resources and schedules users with the best channel quality at that moment. This algorithm maximizes cell capacity but cannot ensure that UEs in the same cell experience the same data rate because UEs do not experience the same channel conditions in a cell. Moreover, if a UE is constantly experiencing poor channel conditions, it cannot be scheduled, leading to “starving” the UE. As a result, user experience is poor.
- **Round Robin (RR)** is a simple scheduling technique in which mobile devices are served without taking into account channel state information. Users are allocated resources without discrimination and by default. Because it does not take into account the current or previous radio channel conditions, there will be a variation of the throughput from one user to another, due to the variation of the radio channel. This type of scheduling is not suited for situations involving users with significant differences in experienced radio conditions. Compared with Max C/I, RR ensures scheduling fairness between UEs in the sense that each user receives the same amount of resources in a cell. However, it cannot maximize the system throughput.
- **Proportional Fair (PF)** algorithm is a compromise between Max C/I and RR. PF scheduling, in its simplest form, allocates resources to the user that maximizes a

function for each PRB. The function is called *Proportional Fair index* and is defined as the ratio between the instantaneous user throughput and the historical average user throughput. However, the algorithm, in this form, does not take the QoS requirements into account and therefore cannot ensure satisfactory user experience, but does lead to fairness in resource allocation

- **Enhanced PF (EPF)** algorithm is an enhancement to the PF algorithm. Compared with PF, EPF considers user experience and ensures that QoS requirements are fulfilled.

For example, an operator can select Max C/I to achieve high system throughput for a peak throughput test. We must also add the fact that the default method that is commonly used by MNOs is basic scheduling. The measurements that were made and presented in the next sections use differentiated scheduling and the PF algorithm.

**III. MEASUREMENT SETUP**

Measurements were made for evaluating the DL throughput performance for two users that simultaneously use the carrier aggregation feature with two carriers. The two aggregated carrier frequencies are located in the ORANGE frequency bands of 2600 MHz (BW 20 MHz) and 1800 MHz (BW 20 MHz). The tests were done in Orange Romania laboratory, in a controlled environment. We used the test topology depicted in Fig. 1.

To achieve tests on the downlink path, a laptop connected remotely to an FTP server was used. From the laptop UDP packets were sent to a smartphone. The iPerf program was used for sending and receiving packets, and measuring throughput and jitter [7]. We monitored the transmission, on the radio interface, from Operational and Maintenance Center (OMC) that is connected remotely to U2000 Network Management System.

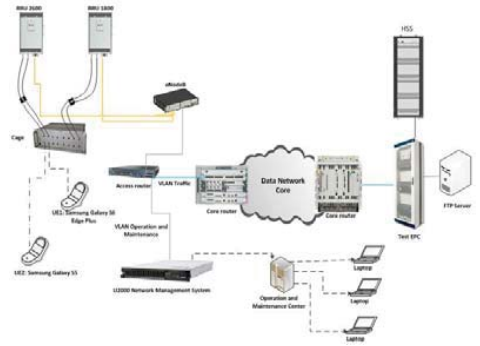


Fig. 1. Laboratory test topology

The smartphone used for tests were a Samsung S6 Edge+, which is a Cat 9 device and a Samsung Galaxy S5 which is a Cat 6 device. For performing the measurements we placed the two devices in a measurement cage (Faraday cage). Also, in the



cage, two antennas were placed: one for 2600 MHz frequency and the other for 1800 MHz frequency. These antennas spread the signal received from two Radio Remote Units: RRU 2600 MHz and RRU 1800 MHz. The cage is connected through 4 feeders to the RRUs: 2 feeders to the 2600 MHz RRU and 2 feeders to the 1800 MHz RRU. On each of these feeders we put fixed 40 dB attenuators for the 2600 MHz frequency band and 30 dB for the 1800 MHz frequency band. Fig. 1 illustrates the measurement setup.

Each eNodeB consists of a baseband unit (BBU) and a remote radio unit RRU which are connected by Common Public Radio Interface (CPRI) interface. This interface transmits RF information from the BBU to the RRU in digital format and then, at the RRU, this information is converted to RF electromagnetic waves. A BBU unit performs operation and maintenance management, signal processing and ports resource management for the whole eNodeB and provides ports for data communications between eNodeB and MME/S-GW. RRU units perform modulation, demodulation, processing, combining and splitting of broadband and radio signals.

The link between the eNodeB and the access router is via optical fiber (there may also be a microwave radio link). An access router aggregates usually 5 to 10 sites. The eNodeB is connected via the access router to the U2000 Network Management System. U2000 is a powerful LTE network management system which is responsible through the VLAN for operation and maintenance of the eNodeB. This VLAN is set by each mobile operator. The main functions of U2000 are: configuration management, performance management, fault management, security management, log management, topology management, software management, and system management [8]. The laptop is connected to the OMC and the activity of the eNodeB can be monitored. Throughput traces were made in U2000 to observe the throughput of the equipment that uses carrier aggregation. The results presented in section IV were also obtained via the software.

All data transmission, in the lab topology, is full-IP. Under laboratory conditions the losses can be approximated with the data loss given by the equation of free space losses.

#### IV. MEASUREMENTS AND RESULTS

Our scenario was made to observe the behavior of two users that use differentiated scheduling under carrier aggregation. We had at the beginning two users (UE1 and UE2) that were placed in the center of the 2600 MHz cell which is also the center of the 1800 MHz cell. We used for measurements two cells that are concentric. We calculated the relation between the diameters of the two cells using the equation of free space losses (under laboratory conditions we can approximate the losses with the data loss given by this model) in a previous paper [9]:

$$FSPL = 20 \log_{10} \frac{4 \cdot \pi \cdot d \cdot f}{c} \text{ (dB)} \quad (1)$$

$$d_{1800\text{MHz}} = 1.449 \cdot d_{2600\text{MHz}} \quad (4)$$

This relation is important to determinate the coverage and we can design the moving route for UE2. So, the moving user first leaves the small cell (2600 MHz cell) and then the 1800

MHz cell. In these conditions, we could make measurements for differentiated downlink scheduling under carrier aggregation with 2 CCs.

UE1 is placed in the center of the 2600 MHz cell and is fixed there for the entire scenario duration. UE2 moves from the center to the edge of the 1800 MHz cell. This scenario was made to observe how two users that use two aggregated carriers make use of the allocated resources. UE1 is moving to observe what is happening with the resources of the two cells when it comes across the states presented in Fig. 2.

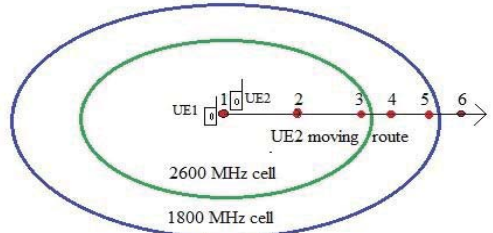


Fig. 2. The route and states of the two users

Table 1 shows the frequencies used, the DL bandwidth, the cell ID of the three cells and their usage.

In Table 2 and Table 3 the values for throughput and RSRP (Reference Signal Received Power) are summarized for the two users, on our route. These values are provided from the throughput traces, made in U2000, that are showed in details in graphs from Fig. 3 to Fig. 8.

TABLE I. DETAILS ABOUT THE CELLS USED

Freq [MHz]	Downlink EARFCN	Downlink bandwidth [MHz]	Cell ID	Cell usage	
				UE1	UE2
1800	1400	20	210	SCell1	SCell1 in state 3, 4, 5
2600	2850	20	211	PCell	PCell in state 1, 2

TABLE II. THROUGHPUT AND RSRP FOR UE1 AND UE2

Time	State1		State2		State3	
	11:15		11:24		11:33	
	RSRP [dBm]	Thr [Mbps]	RSRP [dBm]	Thr [Mbps]	RSRP [dBm]	Thr [Mbps]
UE1	-58	148	-60	140	-60	143
UE2	-53	146	-86	116	-80	61

TABLE III. THROUGHPUT AND RSRP FOR UE1 AND UE2

Time	State4		State5		State6	
	11:42		11:57		12:10	
	RSRP [dBm]	Thr [Mbps]	RSRP [dBm]	Thr [Mbps]	RSRP [dBm]	Thr [Mbps]
UE1	-60	219	-60	218	-60	294
UE2	-94	14	-106	4	-110	-

In state 1 both users are placed in the center of the 2600 MHz cell. We used for the two frequency bands a bandwidth of 20 MHz, so we have 200 available RBs [6]. In the cell center

each RB has approximately 1.5 Mbps of capacity. When the user moves to the cell edge the value decreases. Each user in state 1 gets allocated 100 RBs. UE1 has a throughput of 148 Mbps and UE2 has a throughput of 146 Mbps.

In state 2 UE2 moves from the cell center towards the edge of 2600 MHz cell. UE2 has a throughput of 116 Mbps, its PCell is the 2600 MHz one and UE1 has a throughput of 140 Mbps.



Fig. 3. DL throughput for UE2 in state 2



Fig. 4. DL throughput for UE1 in state 2

In state 3 UE2 arrives close to the cell edge of the 2600 MHz cell. The RSRP on this cell for UE2 is -90 dB, and for 1800 MHz cell it is -80, so a more powerful signal. In this state the cell of 1800 MHz will become the PCell for UE2. The throughput for UE1 is 143 Mbps and for UE2 61 Mbps.



Fig. 5. DL throughput for UE2 in state 3

In state 4 UE2 exits from the 2600 MHz cell. The RSRP on the 2600 MHz cell is -110 dBm, and we can see that this value is smaller than the value for which the user left the cell. In this state UE2 is camped only on 1800 MHz cell and it is now a non-CA user. Its throughput is 14 Mbps. We can notice that, because of the differentiated scheduling, UE1 receives in this state 100 % of the resources for the 2600 MHz cell and 50 % of the resources for 1800 MHz. It remains in the cell center, so its throughput now has a value of 218 Mbps (147 Mbps on PCell and 71 Mbps on SCell). If we use basic scheduling, UE1 cannot use the resources for the SCell as long as we have non-CA users on this cell and its throughput remains 147 Mbps. So,

the resource blocks in this situation are wasted. This was the reason the differentiated scheduling algorithm was studied and used in our tests, for optimizing the network.

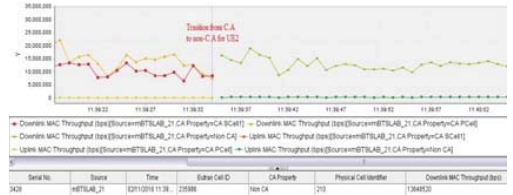


Fig. 6. DL throughput for UE2 in state 4

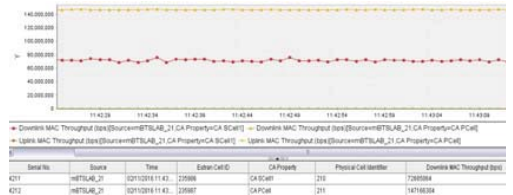


Fig. 7. DL throughput for UE1 in state 4

In state 5, UE2 arrives very close to the edge of the 1800 MHz cell. We performed the measurements in this state to see that even if UE2 has a very small value for the throughput (4 Mbps), UE1 doesn't have a higher throughput than in state 5. While there are users in the cell, even at the cell edge, it can receive only 50 % of the resources.

In state 6 UE2 leaves the 1800 MHz cell, and now UE2 receives all the resources and has a throughput of 294 Mbps.

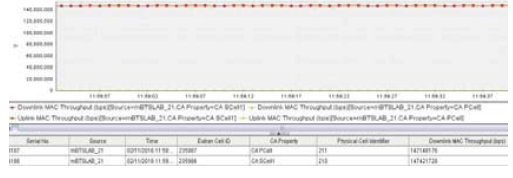


Fig. 8. DL throughput for UE1 in state 6

V. CONCLUSIONS

As we can observe for our tests, the usage of the differentiated scheduling method gives priority to CA UEs over the non-CA UEs. From the measurements described in section IV we saw that a CA-enabled UE can receive more RBs and experience higher throughput than a non-CA UE. However, the amount of radio resources available for non-CA UEs decreases. This is a good method of avoiding the waste of the resources when the RSRP and throughput for the non-CA users decrease (they are placed at the cell edge).

It is up to each Mobile Network Operator to decide which method is more beneficial for its network. This decision depends largely on the number of the non-CA and CA users in the network, but can also be influenced by other factors such as

number of users with GBR services and scheduling algorithm used.

In the future, the authors want to study traffic scheduling algorithms for evaluating basic and differentiated downlink scheduling under carrier aggregation with 3 CCs.

ACKNOWLEDGMENT

This work was supported by the department Huawei Network Operational Skill Center from Orange Romania and it has been funded by Orange Educational Program Engineering from Orange Romania.

REFERENCES

[1] A. Vulpe, O. Fratu, A. Mihovska and R. Prasad, "A multi-carrier scheduling algorithm for LTE-advanced," *Wireless Personal Multimedia Communications (WPMC), 2013 16th International Symposium on*, Atlantic City, NJ, 2013, pp. 1-5

[2] A. Vulpe, S. Halunga and O. Fratu, "QoS-aware downlink scheduling in multi-carrier communication systems," *Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic*

*Systems (VITAE), 2014 4th International Conference on*, Aalborg, 2014, pp. 1-5.

[3] eRAN Scheduling Feature Parameter Description, HUAWEI TECHNOLOGIES CO., LTD., august 2015

[4] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Oxford, UK: Academic Press, 2011

[5] 3GPP TS 36.321, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification (Release 8)," 2010

[6] Carrier Aggregation Feature Parameter Description, HUAWEI TECHNOLOGIES CO., LTD., august 2015

[7] <https://iperf.fr/>

[8] <http://e.huawei.com/en/products/wireless/elte-access/network-element/m2000>

[9] M. Oproiu, A. Vulpe and I. Marghescu, "Effects of enabling carrier aggregation in a mobile operator network," *Telecommunications Forum Telfor (TELFOR), 2015 23rd*, Belgrade, 2015, pp. 91-94

# Robust C-RAN Precoder Design for Wireless Fronthaul with Imperfect Channel State Information

Dong Wang, Ying Wang, Ruijin Sun, Xiangyang Zhang  
 State Key Laboratory of Networking and Switching Technology  
 Beijing University of Posts and Telecommunications, Beijing 100876, P.R. China  
 Email: wangying@bupt.edu.cn

**Abstract**—Cloud Radio Access Network (C-RAN) architecture with optical fiber fronthaul has been confirmed as a promising solution to achieve high capacity and low latency signal transmission, which has been a key technology and trend of the evolving fifth generation (5G) cellular networks. However, with the fiber fronthaul increasing, the complexity and cost of the C-RAN fronthaul networks will grow exponentially. Accordingly, the hybrid fronthaul network of wireless and optical will be the direction of C-RAN architecture design in the future. In this paper, we study the wireless fronthaul C-RAN system in downlink and propose a robust precoder design. The channel state information (CSI) at the baseband unit (BBU) pool and remote radio head (RRH) cluster is assumed to be imperfect, where the additive channel state information error is modeled as Gaussian distributed. Based on this model, we propose a robust C-RAN precoder design that minimizes the total transmit power under a signal-to-interference-plus-noise ratio (SINR) constraint at each user terminal. The original goal is to establish SINR constrained power allocation formulations in the form of convex conic optimization problem. The analysis results reveal that the original problem formulation is non-convex, in general. We develop a novel conservative approximation scheme for handling the non-convex constraint. Furthermore, we solve the optimization problem by transforming it into a semidefinite program with relaxation, which can be efficiently solved. Simulation results show the advantage of using the proposed power-conserving robust precoding algorithm.

**Index Terms**—C-RAN, Wireless Fronthaul, Imperfect Channel State Information, Robust Precoder.

## I. INTRODUCTION

Cloud radio access network (C-RAN) has gained considerable amount of interest as an emerging technology for the evolving fifth generation (5G) cellular networks since it can significantly increase data throughput and reduce energy consumption [1], [2]. In C-RAN systems, the traditional role of the base stations (BSs) is replaced with three parts: the baseband units (BBUs) clustered as a BBU pool in a centralized cloud server, the distributed installed remote radio heads (RRHs) and the fronthaul between RRHs and BBUs. With this structure, the signals received by cost-effective and power-efficient RRHs can be filtered and compressed from user equipment (UE) and then forwarded to the advanced network coordinative BBU pool through the fronthaul. There are substantial advantages employing C-RAN systems, such as easier network management, software-defined networking, network function virtualization, low energy consumption, and efficient mitigation of interference through large-scale cooperative processing [3].

The fronthaul plays a key role in C-RAN systems, because the large volumes of data transmission must be achieved in real time. Therefore, it is easy to consider that the fiber fronthaul solution is a great candidate scheme for high capacity and low latency fronthaul links, which has been proved by many researchers and deployed by some companies [3]. However, with the fiber fronthaul increasing, the complexity of the C-RAN fronthaul network will be growing exponentially, which eventually results in the cost-wasted on capital expenditure (CAPEX) and operating expenditure (OPEX) [4]. In order to address these problems, wireless fronthaul solution has been proposed and proved its feasibility [5]. Hence, in order to achieve low CAPEX/OPEX cost and flexibility in network deployment, we should turn our attention to C-RAN with wireless fronthaul links [6].

Although high spectral efficiency (SE) and energy efficiency (EE) have been provided, C-RAN systems still take heavy power consumption [7]. Recently, numerous research projects are involved in new energy consumption model and resource optimization in C-RAN systems. The algorithm proposed in [8] optimizes the BBU resource allocation in such a way that BBUs are put to sleep and active states by switching them OFF and ON respectively according to their resource usage. By passive optical networks (PONs) exploiting power over fiber (PoF), joint control method of RRH sleep and transmission power are proposed [9], which can be installed with low installation cost and is capable of providing communication services without external power supply for RRH. However, those methods did not take into account imperfect channel state information (iCSI), which is very intractable in C-RAN systems with wireless fronthaul.

In this paper, instead of simply considering total transmission power consumption with perfect CSI, we study the C-RAN system precoder to achieve power efficient transmission, where iCSI is assumed at the BBU pool and RRH cluster. This problem is fundamentally different from conventional amplify-and-forward (AF) relay networks [6], [10]. Our work is focused on a novel robust BBU pool and RRH cluster precoder design in order to minimize the C-RAN system transmit power under the signal-to-interference-plus-noise ratio (SINR) constraint for each operator. Furthermore, simulation results show that the proposed C-RAN joint precoding method can reduce the transmission power and outage probability at the same time while satisfying the individual SINR constraint.

The remainder of this article is organized as follows: The C-RAN with wireless fronthaul system model is given in

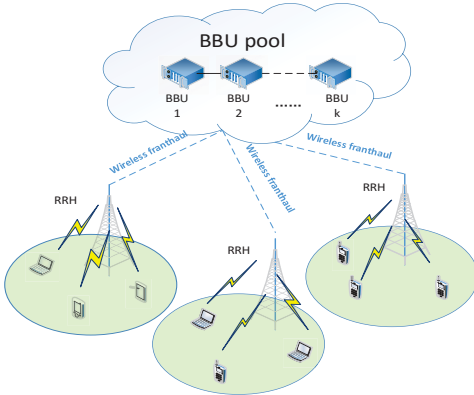


Fig. 1. system model of C-RAN architecture with wireless fronthaul

Section II. Section III formulates the C-RAN system precoding optimization problem with power efficiency as the objective function. The proposed robust C-RAN precoder design to solve this optimization problem is discussed in Section IV. Finally, simulation results and conclusions are presented in Sections V and VI, respectively.

*Notations:* We use boldface letters to denote vectors and matrices. The  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^*$ ,  $\text{tr}(\mathbf{A})$ , and  $\text{vec}(\mathbf{A})$  denote the transpose, conjugate transpose, conjugate, trace of the matrix  $\mathbf{A}$ , and the operator transforming the matrix  $\mathbf{A}$  into a vector, respectively. The  $\otimes$  denotes the Kronecker product.  $\mathbb{N}(0, \sigma^2)$  denotes the distribution of a complex Gaussian random variable with zero mean and variance  $\sigma^2$ .

## II. SYSTEM MODEL

We consider a downlink C-RAN system consisting of one BBU pool, one RRH cluster with  $R$  RRHs and  $K$  users in each RRH as illustrated in Fig. 1. The BBU pool employs  $N$  transmit antennas, each RRH is equipped with  $M_R$  antennas and each user has a single antenna. The direct links between BBU and users are not used since it is assumed that there are some functions still need to be completed in RRH.

The transmission process consists of two steps. During the first step, BBU pool selects a symbol  $s_{rk}$  for  $r = 1 \dots R$ ,  $k = 1 \dots K$ , precodes with vector  $\mathbf{c}_{rk} \in \mathbb{C}^{N \times 1}$  and sends it to the RRHs through the wireless fronthaul. Then each RRH processes the received signals. The BBU pool transmit signal can be presented as

$$\mathbf{s} = \sum_{r=1}^R \sum_{k=1}^K \mathbf{c}_{rk} s_{rk}. \quad (1)$$

Assuming the slow-flat fading scenario. Then, the baseband signal received at the  $r$ th RRH, is given by

$$\mathbf{x}_r = \mathbf{h}_r \left( \sum_{k=1}^K \mathbf{c}_{rk} s_{rk} \right) + \sum_{\substack{p=1 \\ p \neq r}}^R \mathbf{h}_r \left( \sum_{k=1}^K \mathbf{c}_{pk} s_{pk} \right) + \mathbf{z}_r, \quad (2)$$

where  $\mathbf{h}_r \in \mathbb{C}^{M_R \times N}$  denotes the wireless fronthaul channel between BBU pool and the  $r$ th RRH.  $\mathbf{z}_r \sim \mathbb{N}(0, \sigma_1^2 \mathbf{I}_{M_R})$

is supposed to be an additive white Gaussian noise at the  $r$ th RRH. RRH filters its received signal and multiplies by a complex weight matrix  $\mathbf{w}_r \in \mathbb{C}^{M_R \times M_R}$ , then transmits it to the destinations. The transmitted signal by the  $r$ th RRH can be written as follows

$$\mathbf{t}_r = \mathbf{w}_r \mathbf{x}_r. \quad (3)$$

In the second step, the RRH transmits their processed signals to the multiple destinations. After taking these two steps of transmission, the received signal [11] at the  $k$ th receiver can be expressed as follows

$$\begin{aligned} y_{rk} &= \mathbf{g}_{rk}^T \mathbf{t}_r + n_{rk}, \quad \forall k = 1 \dots K \\ &= \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{rk} s_{rk} + \mathbf{g}_{rk}^T \sum_{\substack{i=1 \\ i \neq k}}^K \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{ri} s_{ri} \\ &\quad + \mathbf{g}_{rk}^T \sum_{\substack{p=1 \\ p \neq r}}^R \sum_{i=1}^K \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{pi} s_{pi} + \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{z}_r + n_{rk}, \end{aligned} \quad (4)$$

where  $\mathbf{g}_{rk}^T \in \mathbb{C}^{1 \times M_R}$  denotes the wireless link channel between the  $r$ th RRH and the  $k$ th UE. We assume unit signal power, that is,  $E[|s_{rk}|^2] = 1$  ( $k = 1 \dots K$ ). The noise at UE  $n_{rk} \sim \mathbb{N}(0, \sigma_2^2)$  for  $r = 1 \dots R$ ,  $k = 1 \dots K$  is independent, identically distributed complex additive white Gaussian noise samples. The first, the second, the third and the fourth terms in second line of (4) are the desired signal, interference from the same RRH, interference from different RRHs, and noise components, respectively.

## III. PROBLEM FORMULATION

By denoting  $\gamma_{rk}$  as the specified SINR threshold, the SINR at the  $k$ th destination UE can be written as

$$\text{SINR}_{rk} = \frac{P_s^{rk}}{P_{I_1}^{rk} + P_{I_2}^{rk} + P_N^{rk}}. \quad (5)$$

Here

$$\begin{aligned} P_s^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{rk} s_{rk} \right|^2 \right\}, \\ P_{I_1}^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \sum_{i \neq k} \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{ri} s_{ri} \right|^2 \right\}, \\ P_{I_2}^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \sum_{p \neq r} \sum_i \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{pi} s_{pi} \right|^2 \right\}, \\ P_N^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{z}_r + n_{rk} \right|^2 \right\}, \end{aligned}$$

$P_s^{rk}$ ,  $P_{I_1}^{rk}$ ,  $P_{I_2}^{rk}$  and  $P_N^{rk}$  are the desired signal, interference from the same RRH, interference from different RRHs, and noise power at the  $k$ th destination UE, respectively.

The problem is to minimize the total BBU pool and RRHs transmit power in C-RAN system  $P_{total} = P_{BBU} + \sum_{r=1}^R P_{RRH_r}$  under the SINR constraint at each UE, which is formulated as

$$\begin{aligned} \min_{\mathbf{c}_r, \mathbf{w}_r} \quad & P_{total} \\ \text{s.t.} \quad & \text{SINR}_{rk} \geq \gamma_{rk}, \\ & r = 1 \dots R, k = 1 \dots K. \end{aligned} \quad (6)$$

## IV. ROBUST C-RAN PRECODER DESIGN

As stated in [12], We apply the stochastic error model in this work, where the probability distribution of the CSI error is Gaussian. This model is applicable when the channel estimation error is more important than the quantization error. The channel uncertainties are modeled as follows

$$\begin{aligned} \mathbf{h}_r &= \hat{\mathbf{h}}_r + \mathbf{e}_r, \quad r = 1 \dots R, \\ \mathbf{g}_{rk}^T &= \hat{\mathbf{g}}_{rk}^T + \mathbf{f}_{rk}^T, \quad k = 1 \dots K, \end{aligned} \quad (7)$$

where the vector  $\mathbf{h}_r$  and  $\mathbf{g}_{rk}^T$  represent the estimate of the channel coefficients between the BBU and the  $r$ th RRH and that between the  $r$ th RRH and the  $k$ th UE, respectively. The actual values of channel coefficients are severally denoted by  $\hat{\mathbf{h}}_r$  and  $\hat{\mathbf{g}}_{rk}^T$ . The corresponding CSI errors are  $\mathbf{e}_r$  and  $\mathbf{f}_{rk}^T$ , respectively. The error vectors are defined as

$$\begin{aligned} \mathbf{e}_r &\sim \mathcal{N}(0, \sigma_e^2 \mathbf{I}), \quad r = 1 \dots R, \\ \mathbf{f}_{rk}^T &\sim \mathcal{N}(0, \sigma_f^2 \mathbf{I}), \quad k = 1 \dots K. \end{aligned} \quad (8)$$

Furthermore, we assume that  $\mathbf{e}_r$  and  $\mathbf{f}_{rk}^T$  are uncorrelated with  $\hat{\mathbf{h}}_r$  and  $\hat{\mathbf{g}}_{rk}^T$ , respectively.

The BBU pool transmit power is expressed as

$$P_{BBU} = \sum_{r=1}^R \sum_{k=1}^K \|\mathbf{c}_{rk} s_{rk}\|^2 = RK \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H). \quad (9)$$

The  $r$ th RRH transmit power is expressed as

$$\begin{aligned} P_{RRH_r} &= E \left\{ \text{tr}(\mathbf{t}_r \mathbf{t}_r^H) \right\} \\ &= K \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) E \left\{ \text{tr}(\mathbf{w}_r \mathbf{h}_r \mathbf{h}_r^H \mathbf{w}_r^H) \right. \\ &\quad \left. + \eta^2 (R-1) \text{tr}(\mathbf{w}_r \mathbf{h}_r \mathbf{h}_r^H \mathbf{w}_r^H) \right\} \\ &\quad + M_R \sigma_1^2 \text{tr}(\mathbf{w}_r \mathbf{w}_r^H), \end{aligned} \quad (10)$$

where  $\eta^2$  denotes the joint reduction based on RRH filter and BBU precoding interference suppression. By inserting (7) and (8) into (10) referencing [13],  $P_{RRH_r}$  is further written as

$$\begin{aligned} P_{RRH_r} &= K \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \underbrace{(1 + \eta^2 (R-1))}_{P_r} E \left\{ \text{tr}(\mathbf{w}_r \right. \\ &\quad \left. (\hat{\mathbf{h}}_r + \mathbf{e}_r) (\hat{\mathbf{h}}_r + \mathbf{e}_r)^H \mathbf{w}_r^H) \right\} + M_R \sigma_1^2 \text{tr}(\mathbf{w}_r \mathbf{w}_r^H) \\ &= K P_r \left( \text{tr}(\mathbf{w}_r \hat{\mathbf{h}}_r \hat{\mathbf{h}}_r^H \mathbf{w}_r^H) + \text{tr}(\mathbf{w}_r \mathbf{e}_r \mathbf{e}_r^H \mathbf{w}_r^H) \right) \\ &\quad + M_R \sigma_1^2 \text{tr}(\mathbf{w}_r \mathbf{w}_r^H) \\ &= K P_r \text{tr}(\mathbf{w}_r \hat{\mathbf{h}}_r \hat{\mathbf{h}}_r^H \mathbf{w}_r^H) \\ &\quad + (K P_r \sigma_e^2 + M_R \sigma_1^2) \text{tr}(\mathbf{w}_r \mathbf{w}_r^H) \\ &= K P_r \text{tr} \left( \underbrace{(\hat{\mathbf{h}}_r^T \otimes \mathbf{I}_{M_R})}_{\mathbf{D}} \underbrace{\text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r^H)}_{\mathbf{f}_w} (\hat{\mathbf{h}}_r^T \otimes \mathbf{I}_{M_R})^H \right) \\ &\quad + (K P_r \sigma_e^2 + M_R \sigma_1^2) \text{tr}(\text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r^H)) \\ &= K P_r \text{tr}(\mathbf{f}_w^H \mathbf{D}^H \mathbf{D} \mathbf{f}_w) + (K P_r \sigma_e^2 + M_R \sigma_1^2) \text{tr}(\mathbf{f}_w^H \mathbf{f}_w) \\ &= \mathbf{f}_w^H (K P_r \mathbf{D}^H \mathbf{D} + (K P_r \sigma_e^2 + M_R \sigma_1^2) \mathbf{I}_{M_R^2}) \mathbf{f}_w. \end{aligned} \quad (11)$$

In the following, we express the SINR of  $k$ th UE as a function of  $\mathbf{f}_w$ . The power of the desired signal in (7) is calculated as

$$\begin{aligned} P_s^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{rk} s_{rk} \right|^2 \right\} \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) E \left\{ \text{tr}(\mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{h}_r \mathbf{h}_r^H \mathbf{w}_r^H \mathbf{g}_{rk}^*) \right\} \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) E \left\{ \text{tr} \left( (\hat{\mathbf{g}}_{rk}^T + \mathbf{f}_{rk}^T) \mathbf{w}_r \right. \right. \\ &\quad \left. \left. (\hat{\mathbf{h}}_r + \mathbf{e}_r) (\hat{\mathbf{h}}_r + \mathbf{e}_r)^H \mathbf{w}_r^H (\hat{\mathbf{g}}_{rk}^* + \mathbf{f}_{rk}^*) \right) \right\}. \end{aligned} \quad (12)$$

Due to the assumption that  $\mathbf{g}_{rk}^T$  and  $\hat{\mathbf{h}}_r$  are uncorrelated with  $\mathbf{f}_{rk}^T$  and  $\mathbf{e}_r$ , we could further write (12) as follows

$$\begin{aligned} P_s^{rk} &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \text{tr} \left( \hat{\mathbf{g}}_{rk}^T \mathbf{w}_r \hat{\mathbf{h}}_r \hat{\mathbf{h}}_r^H \mathbf{w}_r^H \hat{\mathbf{g}}_{rk}^* + \sigma_e^2 \hat{\mathbf{g}}_{rk}^T \mathbf{w}_r \mathbf{w}_r^H \hat{\mathbf{g}}_{rk}^* \right. \\ &\quad \left. + \sigma_f^2 \mathbf{w}_r \hat{\mathbf{h}}_r \hat{\mathbf{h}}_r^H \mathbf{w}_r^H + \sigma_e^2 \sigma_f^2 \mathbf{w}_r \mathbf{w}_r^H \right) \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \text{tr} \left( \underbrace{(\hat{\mathbf{h}}_r^T \otimes \hat{\mathbf{g}}_{rk}^T)}_{\mathbf{A}} \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H \right. \\ &\quad \left. + \sigma_e^2 \underbrace{(\mathbf{I}_{M_R} \otimes \hat{\mathbf{g}}_{rk}^T)}_{\mathbf{B}} \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H (\mathbf{I}_{M_R} \otimes \hat{\mathbf{g}}_{rk}^T)^H \right. \\ &\quad \left. + \sigma_f^2 (\hat{\mathbf{h}}_r^T \otimes \mathbf{I}_{M_R}) \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H (\hat{\mathbf{h}}_r^T \otimes \mathbf{I}_{M_R})^H \right. \\ &\quad \left. + \sigma_e^2 \sigma_f^2 \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H \right) \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \mathbf{f}_w^H (\mathbf{A}^H \mathbf{A} + \sigma_e^2 \mathbf{B}^H \mathbf{B} \\ &\quad + \sigma_f^2 \mathbf{D}^H \mathbf{D} + \sigma_e^2 \sigma_f^2 \mathbf{I}_{M_R^2}) \mathbf{f}_w. \end{aligned} \quad (13)$$

Similarly, as in (12) and (13), the power of the RRH internal interference is obtained as

$$\begin{aligned} P_{I_1}^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \sum_{i \neq k} \mathbf{w}_r \mathbf{h}_r \mathbf{c}_{ri} s_{ri} \right|^2 \right\} \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \sum_{i=1}^K E \left\{ \text{tr}(\mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{h}_r \mathbf{h}_r^H \mathbf{w}_r^H \mathbf{g}_{rk}^*) \right\} \\ &= \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \sum_{i=1}^K \mathbf{f}_w^H (\mathbf{A}^H \mathbf{A} + \sigma_e^2 \mathbf{B}^H \mathbf{B} \\ &\quad + \sigma_f^2 \mathbf{D}^H \mathbf{D} + \sigma_e^2 \sigma_f^2 \mathbf{I}_{M_R^2}) \mathbf{f}_w. \end{aligned} \quad (14)$$

The interference power from different RRHs is obtained as

$$\begin{aligned} P_{I_2}^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \sum_{p \neq r} \sum_i \mathbf{w}_p \mathbf{h}_r \mathbf{c}_{pi} s_{pi} \right|^2 \right\} \\ &= \eta^2 \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \sum_{p \neq r} \sum_i E \left\{ \text{tr}(\mathbf{g}_{rk}^T \mathbf{w}_p \mathbf{h}_r \mathbf{h}_r^H \mathbf{w}_p^H \mathbf{g}_{rk}^*) \right\} \\ &= \eta^2 \text{tr}(\mathbf{c}_{rk} \mathbf{c}_{rk}^H) \sum_{p \neq r} \sum_i \mathbf{f}_w^H (\mathbf{A}^H \mathbf{A} + \sigma_e^2 \mathbf{B}^H \mathbf{B} \\ &\quad + \sigma_f^2 \mathbf{D}^H \mathbf{D} + \sigma_e^2 \sigma_f^2 \mathbf{I}_{M_R^2}) \mathbf{f}_w. \end{aligned} \quad (15)$$

Concerning the power of the effective noise is obtained as

$$\begin{aligned}
 P_N^{rk} &= E \left\{ \left| \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{z}_r + n_k \right|^2 \right\} \\
 &= M_R \sigma_1^2 E \left\{ \text{tr} \left( \mathbf{g}_{rk}^T \mathbf{w}_r \mathbf{w}_r^H \mathbf{g}_{rk} \right) \right\} + \sigma_2^2 \\
 &= M_R \sigma_1^2 \text{tr} \left( \hat{\mathbf{g}}_{rk}^T \mathbf{w}_r \mathbf{w}_r^H \hat{\mathbf{g}}_{rk} + \sigma_f^2 \mathbf{w}_r \mathbf{w}_r^H \right) + \sigma_2^2 \\
 &= M_R \sigma_1^2 \text{tr} \left( \left( \mathbf{I}_{M_R} \otimes \hat{\mathbf{g}}_{rk}^T \right) \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H \left( \mathbf{I}_{M_R} \otimes \hat{\mathbf{g}}_{rk} \right)^H \right. \\
 &\quad \left. + \sigma_f^2 \text{vec}(\mathbf{w}_r) \text{vec}(\mathbf{w}_r)^H \right) + \sigma_2^2 \\
 &= M_R \sigma_1^2 \mathbf{f}_w^H \left( \mathbf{B}^H \mathbf{B} + \sigma_f^2 \mathbf{I}_{M_R^2} \right) \mathbf{f}_w + \sigma_2^2.
 \end{aligned} \tag{16}$$

Combining (13), (14), (15) and (16), the problem formulation (6) can be written as

$$\begin{aligned}
 \min_{\mathbf{C}, \mathbf{W}_r} P_{total} \\
 \text{s.t. } & \text{tr} \left[ \left( \text{tr}(\mathbf{C}) (1 - K \gamma_{rk} (1 + \eta^2 (R - 1))) \right. \right. \\
 & \quad \left. \left. \left( \mathbf{A}^H \mathbf{A} + \sigma_c^2 \mathbf{B}^H \mathbf{B} + \sigma_f^2 \mathbf{D}^H \mathbf{D} + \sigma_c^2 \sigma_f^2 \mathbf{I}_{M_R^2} \right) \right. \right. \\
 & \quad \left. \left. - \gamma_{rk} M_R \sigma_1^2 \left( \mathbf{B}^H \mathbf{B} + \sigma_f^2 \mathbf{I}_{M_R^2} \right) \mathbf{W}_r \right] \geq \gamma_{rk} \sigma_2^2, \\
 & r = 1 \dots R, \quad k = 1 \dots K.
 \end{aligned} \tag{17}$$

By defining

$$\begin{aligned}
 \mathbf{C} &\triangleq \mathbf{c}_{rk} \mathbf{c}_{rk}^H, \mathbf{C} \succeq 0, \text{ and } \text{rank}(\mathbf{C}) = 1, \\
 \mathbf{W}_r &\triangleq \mathbf{f}_w \mathbf{f}_w^H, \mathbf{W}_r \succeq 0, \text{ and } \text{rank}(\mathbf{W}_r) = 1,
 \end{aligned} \tag{18}$$

where  $\mathbf{C} \succeq 0$ ,  $\mathbf{W}_r \succeq 0$  mean  $\mathbf{C}$ ,  $\mathbf{W}_r$  are symmetric positive semi-definite matrices. Although we have fixed the BBU transmit vector, the objective function of problem (17) is still complicated and non-convex. It is of importance to transform this problem into a tractable form. Motivated by the user SINR constraint idea [14], we introduce RRH SINR constraint  $\gamma_r$  to rewrite the problem (6) as the following:

$$\begin{aligned}
 \min_{\mathbf{c}_r, \mathbf{W}_r} P_{total} \\
 \text{s.t. } & \text{SINR}_{rk} \geq \gamma_{rk}, \\
 & \text{SINR}_r \geq \gamma_r, \\
 & k = 1 \dots K, r = 1 \dots R.
 \end{aligned} \tag{19}$$

Then, we can minimize BBU pool transmit power in C-RAN system at first,

$$\begin{aligned}
 \min_{\mathbf{c}_r} P_{BBU} \\
 \text{s.t. } & \text{SINR}_r \geq \gamma_r, r = 1 \dots R.
 \end{aligned} \tag{20}$$

The power of the desired signal in (2) is calculated as

$$\begin{aligned}
 P_r &= E \left\{ \left| \mathbf{h}_r \mathbf{c}_r \mathbf{s}_r \right|^2 \right\} \\
 &= \sum_{k=1}^K E \left\{ \mathbf{c}_{rk}^H \left( \hat{\mathbf{h}}_r + \mathbf{e}_r \right)^H \left( \hat{\mathbf{h}}_r + \mathbf{e}_r \right) \mathbf{c}_{rk} \right\} \\
 &= \sum_{k=1}^K \text{tr} \left( \mathbf{c}_{rk}^H \hat{\mathbf{h}}_r^H \hat{\mathbf{h}}_r \mathbf{c}_{rk} + \sigma_1^2 \mathbf{c}_{rk}^H \mathbf{c}_{rk} \right) \\
 &= \sum_{k=1}^K \left\{ \underbrace{\left( \mathbf{I}_{M_R} \otimes \hat{\mathbf{h}}_r \right)}_{\mathbf{A}_1} \underbrace{\text{vec}(\mathbf{c}_{rk})}_{\mathbf{f}_c} \text{vec}(\mathbf{c}_{rk})^H \left( \mathbf{I}_{M_R} \otimes \hat{\mathbf{h}}_r \right)^H \right. \\
 &\quad \left. + \sigma_1^2 \text{vec}(\mathbf{c}_{rk}) \text{vec}(\mathbf{c}_{rk})^H \right\} \\
 &= \sum_{k=1}^K \mathbf{f}_c^H \left( \mathbf{A}_1^H \mathbf{A}_1 + \sigma_1^2 \mathbf{I}_{M_R^2} \right) \mathbf{f}_c.
 \end{aligned} \tag{21}$$

Similarly, the power of the interference is obtained as

$$\begin{aligned}
 P_I &= \sum_{\substack{p=1 \\ p \neq r}}^R E \left\{ \left| \mathbf{h}_r \mathbf{c}_p \mathbf{s}_p \right|^2 \right\} \\
 &= \eta^2 \sum_{p=1}^R \sum_{\substack{k=1 \\ k \neq r}}^K \mathbf{f}_c^H \left( \mathbf{A}_1^H \mathbf{A}_1 + \sigma_1^2 \mathbf{I}_{M_R^2} \right) \mathbf{f}_c.
 \end{aligned} \tag{22}$$

The power of the noise received by  $r$ th RRH is  $P_N = M_R \sigma_1^2$ . Combining (9), (18), (21) and (22), the optimization problem (20) can be reformulated as following

$$\begin{aligned}
 \min_{\mathbf{C}} P_{BBU} \\
 \text{s.t. } & K \text{tr} \left( \left( 1 - \eta^2 \gamma_r (R - 1) \right) \left( \mathbf{A}_1^H \mathbf{A}_1 \right. \right. \\
 & \quad \left. \left. + \sigma_1^2 \mathbf{I}_{M_R^2} \right) \mathbf{C} \right) \geq \gamma_r M_R \sigma_1^2, \\
 & r = 1 \dots R, \text{ rank}(\mathbf{C}) = 1.
 \end{aligned} \tag{23}$$

By relaxing the non-convex constraint (18) in (23), the original problem turns out to be convex in  $\mathbf{C}$ , and can be solved effectively by semi-definite programming (SDP) relaxation, using the convex optimization toolbox CVX.

Note that  $\mathbf{C}^{opt}$  may be not rank-one in general, after SDP relaxation. Hence, this will make a trouble for us because the optimal weight vector  $\mathbf{c}_{rk}$  can be recovered from  $\mathbf{C}^{opt}$  straightforwardly, using the principal eigenvector corresponding to the only non-zero eigenvalue. In [10], an approximation iterative solution to the original problem is obtained by applying the randomization method. In the randomization iterative method, the singular value decomposition (SVD) of  $\mathbf{C}^{opt}$  is first computed  $\mathbf{C}^{opt} = \mathbf{U} \Sigma \mathbf{V}^H$ . To initialize,  $\mathbf{c}_{rk}^{(i)}$  is set to  $\mathbf{c}_{rk}^{(i)} = \mathbf{U} \Sigma^{\frac{1}{2}} \mathbf{x}$ , where  $\mathbf{x} \in \mathbb{C}^{M_R^2 \times 1}$  is a randomly generated zero mean circular symmetric complex Gaussian vector and  $i$  is the iteration index. We need to set a scaling factor  $\alpha_r^{(i)} = \frac{\gamma_r M_R \sigma_1^2}{K \text{tr} \left( \left( 1 - \eta^2 \gamma_r (R - 1) \right) \left( \mathbf{A}_1^H \mathbf{A}_1 + \sigma_1^2 \mathbf{I}_{M_R^2} \right) \mathbf{C} \right)}$  for  $r$ th RRH to fulfill its SINR requirement exactly. The iteration continues until all the number of iterations  $L = 100$  is completed.

After resolved the optimization problem (23), we can complete the integral robust C-RAN precoder design in the same

way, by inserting the result into (17). Detailed steps to achieve the proposed robust C-RAN precoder design are stated in Algorithm 1.

---

**Algorithm 1** proposed robust C-RAN precoder design
 

---

- 1: Initialize:  $R, K, N, M_R$  for C-RAN;
  - 2: **for**  $r = 1 : R$  **do**
  - 3:   Set  $\gamma_r$ ;
  - 4:   Solve problem in (20) by CVX to obtain  $P_{BBU}$  and  $\mathbf{C}^{opt}$ ;
  - 5:   **for**  $k = 1 : K$  **do**
  - 6:     Set  $\gamma_{rk}$ ;
  - 7:     Insert  $\mathbf{C}^{opt}$  to problem (17). Solve the problem by CVX and obtain  $P_{RRH_r}$  and  $\mathbf{W}_r^{opt}$ ;
  - 8:   **end for**
  - 9: **end for**
  - 10: Calculate  $P_{total}$
  - 11: Using an approximation iterative solution to solve SDP relaxation problem, obtain  $\mathbf{c}_{r,k}^{opt}$  and  $\mathbf{w}_r^{opt}$ ;
  - 12: **return** robust precoder design  $\mathbf{c}_{r,k}^{opt}$  and  $\mathbf{w}_r^{opt}$ .
- 

## V. SIMULATION RESULTS

In our simulation, we consider a perfect CSI/imperfect CSI multi-user C-RAN system with wireless fronthaul. There are  $R = 3$  RRHs and each RRH connects  $K = 3$  users. The BBU pool is equipped with  $N = 4$  transmit antennas, each RRH is equipped with  $M_r = 2$  antennas and each user is equipped with single antenna. The channel realization is generated by zero-mean and unit-variance independent and identically distributed complex Gaussian matrix. The variance of AWGN noise per receive antenna is assumed to be the same for all RRHs and users  $\sigma_1^2 = \sigma_2^2 = 10^{-3}$ . In addition, the symbols transmitted to multiple users are assumed to have the unit power. We set the wireless fronthaul and the wireless link variances of uncertainty as  $\sigma_c^2 = \sigma_l^2 = 10^{-2}$ , the joint reduction based on RRH filter and BBU precoding interference suppression as  $\eta^2 = 0.01$ .

In Fig. 2, we compare the BBU pool transmit power of the proposed robust precoding algorithm in wireless fronthaul C-RAN scenario (iCSI) with non-robust precoding (CSI/iCSI). It shows that the proposed robust precoding design in imperfect CSI wireless fronthaul achieve lower power consumption than non-robust precoding design. However, both robust and non-robust precoding design in imperfect CSI scenario need more transmit power than perfect CSI. In other words, if we do not consider the feature of imperfect CSI, RRH might not obtain the deserved SINR at the same BBU transmit power. It also shows that the saved power portion achieved by the proposed robust algorithm increase as the target SINR growing.

Fig. 3 gives the consumed transmit power in C-RAN system  $P_{total}$  versus the SINR threshold  $\gamma_{rk}$  for the robust and non-robust methods. It is observed that the total transmit power for both cases increase with an increasing SINR threshold. This is because more power has to be payed in order to meet higher SINR requirement. By using the robust method, the consumed total transmit power can be reduced significantly compare to the non-robust method at the same SINR requirement.

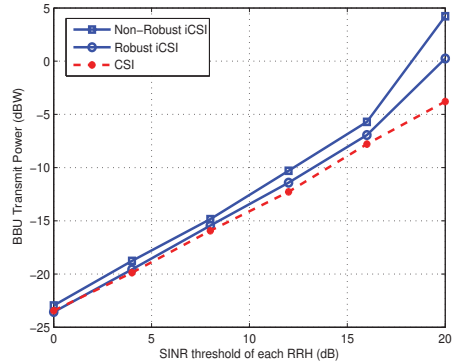


Fig. 2. Transmit power of BBU pool versus the RRH target SINR constraint  $\gamma_r$ , with three RRHs served by one BBU pool.

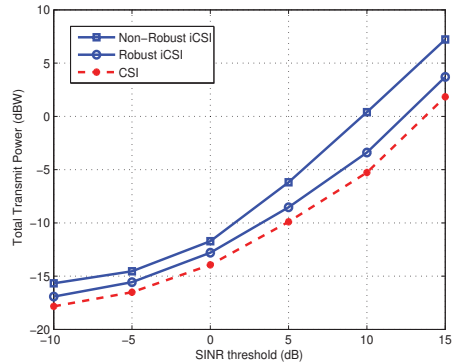


Fig. 3. Total transmit power of C-RAN system versus the user target SINR constraint  $\gamma_{rk}$  at  $\gamma_r = 15dB$ .

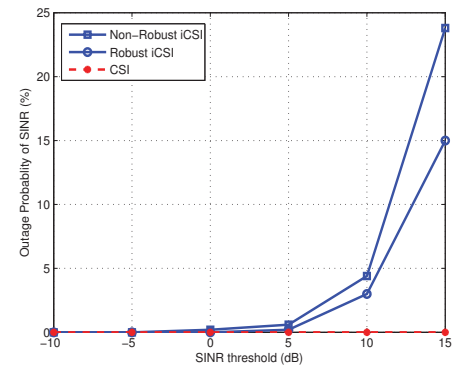


Fig. 4. Outage probability of SINR versus the user target SINR constraint  $\gamma_{rk}$  at  $\gamma_r = 15dB$ .



In Fig. 4, we plot the average outage probability for entire users in C-RAN system as a function of the user target SINR threshold  $\gamma_{r,k}$ . It should be noted that the reason of different outage probabilities here is not only the fading of the channel itself by convention, but also the probabilities that the received SINR at the users are below the threshold due to the wireless fronthaul and wireless link channel estimate errors. It can be observed from Fig. 4 that both robust and non-robust have higher outage probability than perfect CSI in imperfect CSI scenario, especially when the SINR threshold is increasing. Above all, it shows that the outage probability for the proposed robust precoding method is lower than non-robust method, whereas the proposed robust precoder design consumed less transmit power through all the SINR ranges.

## VI. CONCLUSION

In this paper, we propose a robust two-stage precoder design in the BBU pool and RRH cluster to achieve a power efficient transmission for the wireless fronthaul C-RAN system with imperfect CSI. We study the actual case that the wireless fronthaul and the wireless link are imperfect CSI, where robust C-RAN precoder design can save the power consumption further. Based on the analyzed C-RAN system with imperfect CSI, robust BBU pool and RRHs precoding schemes have been proposed to save the power consumption while satisfying each user SINR constraint. As discussed in this paper, the original objective function of problem is complicated and non-convex. It is of importance to transform this problem into a tractable form. So we introduce the RRH SINR constraint, which make the objective function convex and resolved smoothly. The optimization problem can be transformed into a semidefinite program with relaxation. Then the relaxed problem can be efficiently solved by the convex optimization toolbox CVX. We compared this robust precoding algorithm with the non-robust handling scheme, which ignores the influence of channel estimation error for precoder design in imperfect CSI scenario. Simulation results show that the benefits of reducing power consumption and outage probability can be achieved from using the proposed robust precoding algorithm design.

## ACKNOWLEDGMENT

This work is supported by National 863 Project (2014AA01A705), National Nature Science Foundation of China (61431003, 61421061).

## REFERENCES

- [1] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 3, pp. 11–21, 2015.
- [2] Y. Wang, J. Xu, and L. Jiang, "Challenges of system-level simulations and performance evaluation for 5g wireless networks," *IEEE Access*, vol. 4, pp. 1553–1561, 2014.
- [3] China Mobile Research Institute, "C-RAN: the road towards green RAN," *White Paper, ver2.5*, vol. 2, 2011.
- [4] A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1077–1083, 2015.
- [5] G. P. Fettweis, "HetNet Wireless Fronthaul The Challenge Missed," *White Paper*, vol. 2, 2014.
- [6] T. Vu, T. Nguyen, and T. Quek, "Power Optimization with BLER Constraint for Wireless Fronthauls in C-RAN," *IEEE Journals and Magazines*, vol. 20, no. 3, pp. 602–605, 2016.
- [7] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, 2015.
- [8] M. Khan, R. Alhumaima, and H. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in C-RAN," *IEEE European Conference on Networks and Communications (EuCNC)*, pp. 169–174, 2015.
- [9] K. Miyanabe, K. Suto, Z. Fadlullah, H. Nishiyama, N. Kato, H. Ujikawa, and K.-I. Suzuki, "A cloud radio access network with power over fiber toward 5G networks: QoE-guaranteed design and operation," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 58–64, 2015.
- [10] C. Wang, X. Dong, and Y. Shi, "Robust relay design for two-way multi-antenna relay systems with imperfect CSI," *Journal of Communications and Networks*, vol. 16, no. 1, pp. 45–55, 2014.
- [11] X. Wang, Y. Wang, and S. Ma, "Upper bound on uplink sum rate for large-scale multiuser mimo systems with mrc receivers," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2154–2157, 2015.
- [12] J. Li and M. Haardt, "Robust MIMO relay precoder design for multiple operator one-way relaying with imperfect channel state information," *VDE Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, pp. 1–5, 2013.
- [13] M. Amin Maleki Sadr, M. Ahmadian-Attari, and B. Mahboobi, "Low-complexity robust relay optimisation for multiple peer-to-peer beamforming: a safe tractable approximation approach," *IET Communications*, vol. 9, no. 16, pp. 1968–1979, 2015.
- [14] H. Du and P.-J. Chung, "A probabilistic approach for robust leakage-based MU-MIMO downlink beamforming with imperfect channel state information," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1239–1247, 2012.

# Multiple-access performance and interleaver collision in IDMA systems

<sup>1</sup>Olanrewaju B. Wojuola, <sup>2</sup>Stanley H. Mneney and <sup>1</sup>Viranjay M. Srivastava

<sup>1</sup>School of Engineering,  
University of KwaZulu-Natal, Durban, South Africa.

<sup>2</sup>Department of Electronics and Communication Engineering  
St. Joseph University in Tanzania, Dar es Salaam, Tanzania  
wojuolao@ukzn.ac.za, shmneney@gmail.com and viranjay@ieee.org

**Abstract**— Interleave-division multiple-access (IDMA) is a communication technique believed to be a potential candidate for future wireless systems. The technique employs interleavers for user-separation, and is known to have the capacity for minimizing transmission bandwidth and computational requirements. In recent times, significant work has been done on certain key performance indices of interleavers and IDMA systems, but not much reported work exists on multiple-access performance and interleaver collision in IDMA systems. We seek to fill this gap. The outcome of this work shows that high level of collision exists among interleavers when employed as user-separating elements, and this results in significant multiple-access interference. There is therefore need for improvement in collision level of interleavers, in order to pave way for IDMA as a promising candidate for future wireless systems.

**Keywords**—Interleave-division multiple-access (IDMA), code-division multiple-access (CDMA), interleaver collision, multiple-access interference, bit-error-rate.

## I. INTRODUCTION

Interleave division multiple access (IDMA) is a relatively recent technique that was first proposed in literature around the turn of the century[1-4]. It has been attracting much attention because it is generally believed to be a promising candidate for future wireless technology, and as an alternative to code-division multiple-access (CDMA).

IDMA involves the use of interleavers as user-separating elements. Each user is assigned a unique interleaver which is used for encoding and distinguishing the user's signal from those of others. This can be compared to CDMA technique where spreading codes(or sequences) are used for identifying users. In principle, IDMA can be considered as a special case of CDMA having a spreading factor of unity, and which employs user-specific interleavers for user-identification [5].

In IDMA, signal spreading and matrix multiplication are avoided, thereby minimising transmission bandwidth and computational requirements [4-8]. These are important benefits of IDMA. The treatise of Li, Qinghua and Jun[4], and KusumeBauch, and Utschick [5, 9] are examples of comprehensive works on the merits of IDMA. Useful models and important results on the performance of IDMA systems have been published by these and other notable authors.

Significant amount of work has been done on IDMA systems, but literature survey shows that not enough attention has been given to investigating collision level among interleavers. This has direct bearing on the system bit-error-rate (BER) performance. While important parameters such as memory requirements, bandwidth consumption and ease of interleaver generation have been addressed, issues associated with the collision level among interleavers and the system BER performance are yet to be properly investigated. This paper seeks to investigate these important factors.

It is useful to note that the objective of this work is not to present IDMA as a superstar, which is commonly the case in existing literature. Rather, this work attempts to give a balanced view of merits and limitations of IDMA. Our paper highlights certain important factors that have been ignored in literature when evaluating performance of IDMA systems. This helps to shed a better light on potentials and limitations of IDMA.

The rest of this paper is organised as follows. Basic structure of IDMA systems is presented in Section II. This is followed by an outline of approach used for this work in Section III. Simulation results on multiple-access performance of interleaver and IDMA system are presented in Section IV. Results are discussed in Section V; and lastly, a summary of important conclusions are given in Section VI.

## II. SYSTEM MODEL

We consider a general model of CDMA and IDMA transmitter as shown by Fig. 1. Data stream  $b_k$  of a user  $k$  is first encoded by a forward-error-correction (FEC) encoder, followed by interleaving by the interleaver  $\Pi_k$  for alleviating burst errors resulting from fading effects. The interleaved signals undergo spreading operation using a spreading sequence. Following this, the spread signal is mapped onto complex symbol constellation, and then placed onto a RF carrier for onward transmission through a communication channel.

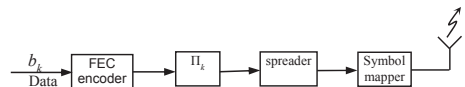


Fig. 1 General model of CDMA and IDMA transmitter

Consider a direct-sequence CDMA system. Let  $b_k(t)$  (with a bit time  $T$ ) be the data for the  $k^{\text{th}}$  user, and  $C_k(t) = \sum_{i=1}^M c_k(t - iT_c)$  be the unique code for the user. Assume that there are  $K$  users, so that there are  $K$  unique codes. Let  $d_k$  be the output of the FEC encoder, which is interleaved to give  $d_k^{(\pi_k)}$ . After signal spreading, the coded output for the user is given by

$$y_k(t) = d_k^{(\pi_k)} \cdot C_k(t) = d_k^{(\pi_k)} \sum_{i=1}^M c_k(t - iT_c), \quad (1)$$

where  $T_c$ , the chip time, is much less than the bit time  $T$ . The multiplication process in (1) has the implication that the spectrum of the bit which is proportional to  $1/T$  is now much larger and is proportional to  $1/T_c$ . Thus the encoding in (1) spreads (enlarges) the spectrum of the signal. The output of the spreader is mapped onto a complex constellation symbols  $y_k^j(t)$ , which gives the signal  $\{y_k^j(t)\}$ , where

$$y_k^j(t) = d_k^{(\pi_k, j)} \sum_{i=1}^M c_k(t - iT_c). \quad (2)$$

This is placed onto a RF carrier for onward transmission through the communication channel. In general, let the channel impulse response  $h_k(t)$  for user  $k$  and  $L$  independent paths be modelled as:

$$h_k(t) = \sum_{l=1}^L \beta_{kl} e^{j\gamma_{kl}} \delta(t - \tau_{kl}), \quad (3)$$

where  $\beta_{kl}$  is path gain,  $\tau_{kl}$  is propagation delay,  $\gamma_{kl}$  is phase shift, and  $l$  is path index. At the receiving end, the received signal  $r_k(t)$  for the user is obtained by convolving the impulse response with the transmitted signal, and adding noise:

$$r_k = \int_{-\infty}^{\infty} y_k^j(\tau) h_k(t - \tau) d\tau + n(t), \quad (4)$$

where  $n(t)$  is receiver noise.

In IDMA system, signal spreading is avoided. Fig. 2 shows the basic structure of IDMA systems having  $K$  users. Each user is assigned a unique interleaver. Here interleaving is used not only for alleviating burst errors resulting from channel fading, but also for user-separation. The interleaved signals from the various users are transmitted through the multiple-access channel where they mix up with one another. Every signal constitutes unwanted interference to the other signals. To recover each signal, the signal is processed iteratively at the receiver using a combination of an elementary signal estimator, a de-interleaver and a FEC decoder, to give the recovered output  $\hat{b}_k(t)$ .

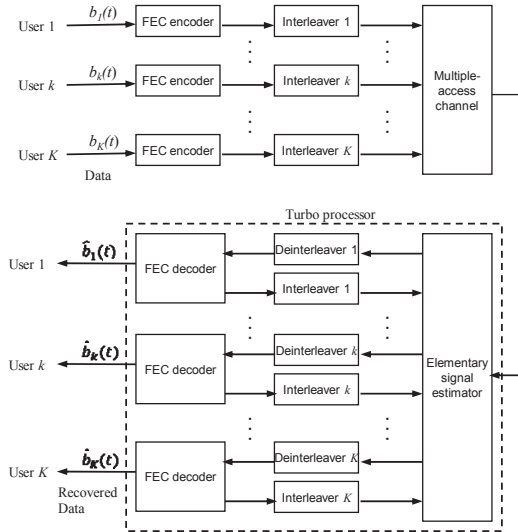


Fig.2. Basic structure of IDMA systems

### III. METHODOLOGY

Software simulations were carried out for the transmission of random binary-phase-shift keying (BPSK) symbols through an additive white Gaussian noise (AWGN) channel of zero mean

and unit variance. Employing AWGN channel is beneficial as it provides lower bounds for the system performance. In addition, the Gaussian channel provides a good environment for investigating multiple-access performance and collisions of interleavers when employed as user-separating elements.

For the IDMA system, pseudo-random interleavers were used as identifying elements. Random interleavers were generated by independently repeating a random interleaver generation procedure  $K$  times, achieved by drawing pseudo-random numbers from a uniform distribution.

The simulations were performed for 1,000 blocks of data samples having a block size of 512 symbols, interleaved with 512-long random interleavers, with Hamming (7,4) code used for error-correction. The simulation was carried out for a single user in the first instance, and then for multiple users. Effects of multiple-access interference were observed.

For the purpose of comparison, the simulation was also carried out for a CDMA system. For the CDMA system, Gold codes having a length of 511 chips, which is comparable to the length of the interleaver, were used as spreading sequence. Performance of uncoded transmission was also simulated. By *uncoded*, we mean transmission not involving the use of interleaver, spreading sequence nor error-correction code. For every simulation, BER was calculated and plotted as a function of signal-to-noise ratio (SNR). Perfect synchronisation was assumed.

We are aware that convolutional code, and not Hamming code is usually used in literature for IDMA systems, and that iterative multi-user detection (MUD) is often employed at the receiver. One of the aims of this work is to study inherent (intrinsic) collision existing among interleavers, and bringing in mitigating techniques such as iterative MUD will obscure this. Furthermore, this work seeks to compare multiple-access performance of IDMA and CDMA systems. This being the case, fair comparison requires that similar conditions, parameters and techniques be used for the two systems.

IV. RESULTS

A. Performance for a single user

We start our consideration by looking at the system performance for a single user, as shown by Fig. 3. As there is no interfering user here, multiple-access interference does not arise. The right-most curve on this figure is that of uncoded data transmission. The figure shows the close agreement between results of analysis and simulation for the uncoded data transmission. For reference purposes, this curve will be retained on all results to be presented in this paper.

Next to the graph for the uncoded data transmission is the performance for IDMA system for a single user. The results show that the deployment of IDMA gives some performance gain. Compared to the uncoded transmission, the results show that at a BER of  $10^{-4}$ , IDMA provides a coding gain of about 2.82 dB.

The leftmost curve of Fig. 3 shows the performance of a CDMA system for a single user. Clearly, the figure shows that CDMA gives better BER performance. At a BER of  $10^{-4}$ , CDMA provides a coding gain of about 27.17 dB, which is much higher than that obtained from IDMA. The results also show that CDMA operates below noise level (i.e. below 0 dB). This is an important property that enables the use of CDMA for covert transmission. For the purpose of comparison, the single-user

graphs for both CDMA and IDMA shall be retained on subsequent results to be presented in this paper.

B. Performance for two users

Next we shall examine the performance for a user in the presence of other interfering users, from a few to tens of users. The treatment begins with two simultaneous users, the result of which is shown by Fig. 4. Comparing the performance for a single user to that of two users, it can be seen that the presence of an interfering user makes BER to become worse. This is expected, as a result of multiple-access interference (MAI). At a BER of  $10^{-4}$ , CDMA experienced about 3.07 dB loss in performance resulting from the presence of the interfering user.

Now turning to IDMA, comparing the performance for a single user to that of two users (Fig. 4), it can be seen that IDMA has severe loss of performance when an interfering user is present. For the two-user system, the BER for IDMA can be seen to degrade rapidly, culminating in the appearance of error floor. This is an indication of existence of severe MAI and high level of collision among the interleavers.

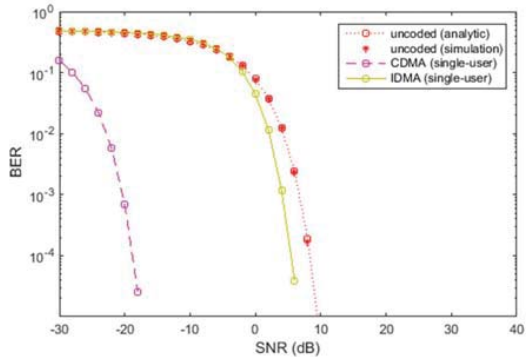


Fig.3 Performance for a single user

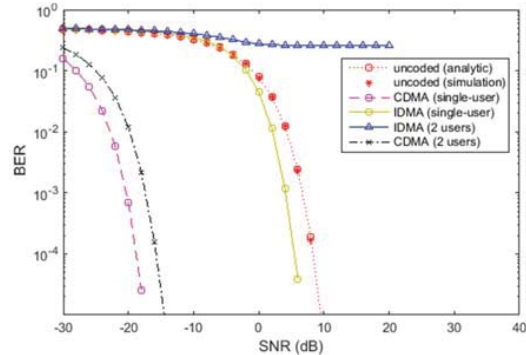


Fig.4 Performance for two users

C. Performance for three to five users

Multiple-access performance of IDMA and CDMA was further investigated for higher number of users. Fig. 5 shows the results for three to five users. For CDMA, this figure shows that performance is slightly worse for three, four and five users in order. This trend is expected, due to increasing MAI resulting from higher number of interferers. Using the performance for a single-user as reference, at a BER of  $10^{-4}$ , CDMA experiences a loss of 4.91, 5.94 and 6.80 dB in performance for three, four and five users respectively. However, it can be seen that CDMA maintains its waterfall BER characteristic behaviour.

Looking at Fig. 5(a)-(c) shows that situation is different for IDMA. It can be seen that BER performance of IDMA is significantly poorer than that of CDMA. The performance of IDMA worsened rapidly with increasing number of users. The visible error floor persisted, worsening to a value of about 0.34 for five users.

D. Performance for higher number of users

The system performance was further tested for higher system loading. Fig. 6 shows the results for 10 and 50 users. Again as expected, the results show that the system performance worsened with increasing number of interferers. However, as with previous results, it can be seen that CDMA is well-behaved when loaded, having a reasonable BER when heavily loaded. Even with 50 users, no error floor emerged for CDMA. In contrast, Fig. 6 shows that for as little as ten users, the BER performance of IDMA has basically flattened out to an error floor of about 0.5.

V. DISCUSSION

We are aware that existing works showed that IDMA outperformed CDMA, and that the performance advantage increased as the number of users increased, particularly when the number of users was greater than length of spreading sequence. These are useful results, but a situation where the number of users exceeds the length of spreading sequence is not appropriate when comparing the two systems, because a CDMA system is not supposed to be operated under such a condition. Furthermore, it has been observed that in published works comparing the two systems, the length of spreading sequences used for CDMA system is usually much smaller than the size of interleavers used for IDMA system. For example in certain cases, the length of spreading sequences used for CDMA is 16, while the size of interleavers used for IDMA system runs up to thousands (e.g. 2048) [4, 5, 7, 10]. This situation is unfavourable for CDMA system. Fair comparison requires that size of elements for user-separation in both systems should be comparable. That is to say, interleaver size for IDMA system should be comparable to the length of spreading sequence for CDMA system. This is essential before the true position of BER performance of IDMA system can be established. In this paper, we put this important factor into consideration, and these are part of what made a difference in the results presented in this paper.

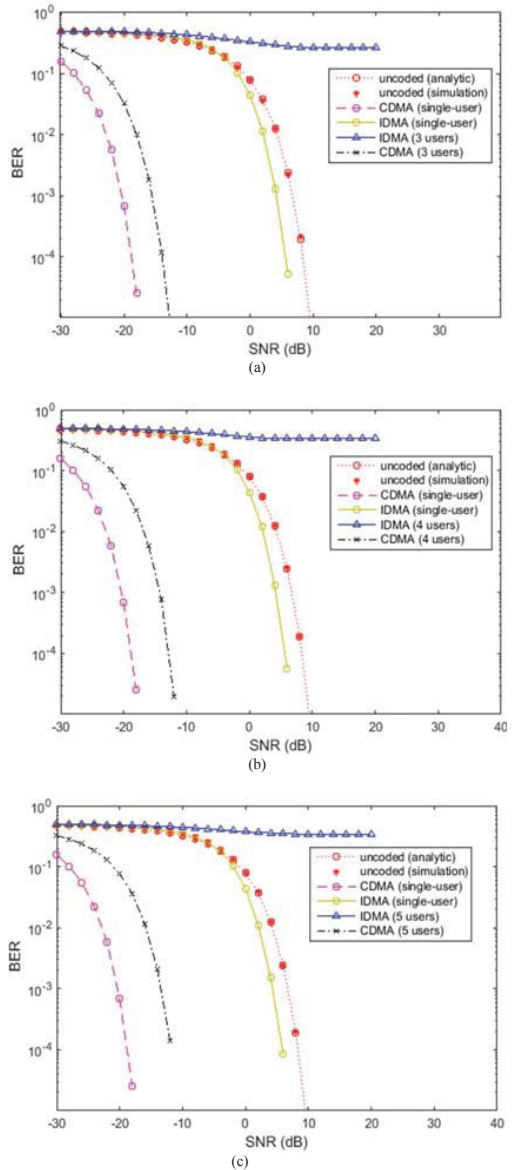


Fig.5 Performance for (a) three, (b) four and (c) five users

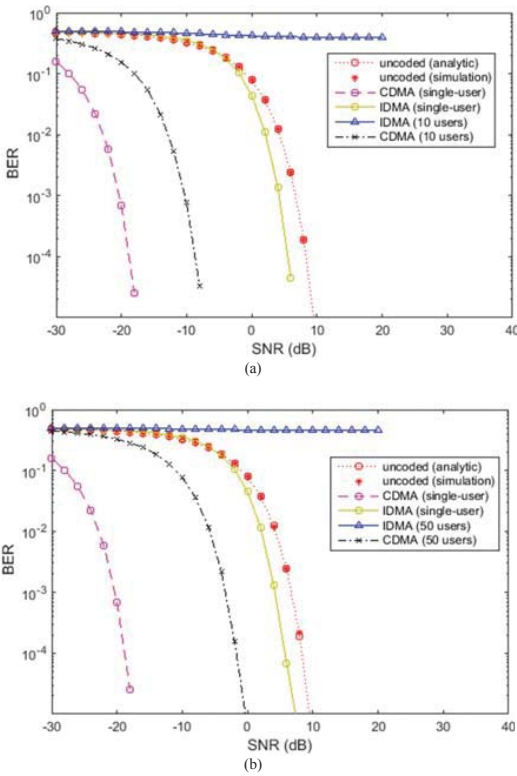


Fig.6 Performance for (a) 10 and (b) 50 users

VI. CONCLUSION

This work investigated interleaver collision and multiple-access performance of IDMA systems. IDMA is known to possess important benefits including less computational requirements and reduced transmission bandwidth. These are important benefits from practical point of view, but they are not without a price. Unfortunately in literature, there is general silence on the trade-offs associated with the benefits of IDMA. This paper showed that the complexity reduction of IDMA has associated loss in system performance. Although IDMA system is less computational intensive, results of this work showed that it

has worse bit-error-rate performance. There was rapid emergence of error-floor when number of users in the IDMA system was increased. This behaviour is indicative of high level of collision existing among the IDMA interleavers.

Interleavers are known to be simple but powerful tools for combatting burst errors in fading channels, but they are not optimal when used as user-separating elements in multiple-access applications. To enhance the success of IDMA as a technique for future telecommunication systems, there is need for improvement in the level of collision existing among IDMA interleavers.

REFERENCES

- [1] W. K. Leung, L. Lihai, and P. Li, "Interleaving-based multiple access and iterative chip-by-chip multiuser detection," *IEICE transactions on communications*, vol. 86, pp. 3634-3637, 2003.
- [2] L. Ping, "Interleave-division multiple access and chip-by-chip iterative multi-user detection," *IEEE Communications Magazine*, vol. 43, pp. S19-S23, 2005.
- [3] P. A. Hoehner and H. Schoeneich, "Interleave-division multiple access from a multiuser theory point of view," in *4th International Symposium on Turbo Codes & Related Topics; 6th International ITG-Conference on Source and Channel Coding (TURBO-CODING)*, 2006, pp. 1-5.
- [4] P. Li, L. Lihai, W. Keying, and W. K. Leung, "Interleave division multiple-access," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 938-947, 2006.
- [5] K. Kusume, G. Bauch, and W. Utschick, "IDMA vs. CDMA: Analysis and Comparison of Two Multiple Access Schemes," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 78-87, 2012.
- [6] M. K. Shukla, A. Gupta, and R. Bhatia, "A Survey on Various Interleavers in Iterative IDMA Communication System," in *International Conference on Special Functions and their Applications in Science and Engineering*, 2011.
- [7] P. Li, G. Qinghua, and T. Jun, "The OFDM-IDMA approach to wireless communication systems," *IEEE Wireless Communications*, vol. 14, pp. 18-24, 2007.
- [8] R. Gupta, B. Kanaujia, R. Chauhan, and M. Shukla, "Prime number based interleaver for multiuser iterative IDMA systems," in *International Conference on Computational Intelligence and Communication Networks (CICN)*, 2010, pp. 603-607.
- [9] K. Kusume, G. Bauch, and W. Utschick, "IDMA Vs. CDMA: Detectors, Performance and Complexity," in *IEEE Global Telecommunications Conference (GLOBECOM '09)* 2009, pp. 1-8.
- [10] K. Kusume, G. Bauch, and W. Utschick, "IDMA Vs. CDMA: Detectors, Performance and Complexity," in *IEEE Global Telecommunications Conference (GLOBECOM 2009)* 2009, pp. 1-8.

## A Combined method of LDPC-Polar code for image transmission

TanapornPayommai and Kosin Chamnongthai

Department of Electronic and Telecommunication Engineering, Faculty of Engineering

King Mongkut's University of Technology Thonburi

126 Pracha-uthit Rd., Bangmod, Thungkhru, Bangkok 10140, Thailand

E-mail: [kosin.cha@kmutt.ac.th](mailto:kosin.cha@kmutt.ac.th), [Tanaporn@vru.ac.th](mailto:Tanaporn@vru.ac.th)

**Abstract**— this paper proposed a method of Low density parity check-Polar (LDPC-Polar) coding as an efficient channel coding technique to protect data for image transmission. The performance of LDPC-Polar code on the Additive White Gaussian Noise (AWGN) channel is studied. The construction of LDPC-Polar code is designed with two-difference-code lengths in the same code rate. As simulation results, the information sent by using a combined LDPC-Polar code can outperform another code. The performance is evaluated by bit error rate (BER) and the proposed method can adequately maintain the quality of the received image and high peak signal-to-noise ratio (PSNR) at low SNR.

### I. INTRODUCTION

Image transmission is an important for wireless network. Error control code is the significant for transmission. However, wireless channel is usually sensitive to time-varying and multi-path fading, which causes the transmitted images to exhibit a high BER. Loss of either the completeness or quality of the images can occur at a receiver if the images are compressed. The BER is influenced by receiver imperfection, and a high bit error makes the reconstructed image unacceptable. In future communication systems, a lower BER will be required to provide high-quality multimedia service; hence, more powerful channel coding is needed. In the past few years, research on channel coding has attracted considerable interest for the purpose of improving performance at a low cost. The introduction of redundant bits into the input data can help remove errors during transmission over the channel.

Low-density parity-check (LDPC) codes was introduced by Gallager [1] in 1962 and it was rediscovered by Mackay and Neal [2] as one of many kinds of linear block codes that have been vastly studied in communication system. In [3], Mackey and Davey explored whether Gallager codes are useful for high rates ( $R > 2/3$ ) and small block lengths ( $N < 5000$ ) and showed that they could outperform Reed Solomon code. Ryan and Yongqing [4,5] proved high rate LDPC codes have a good performance.

After that, sub block encoder is designed for high rate LDPC that can keep quality and improve the performance [6]. And the accuracy for sub block encoder of LDPC is improved to enhance the performance for image transmission when code length increasing [7]. Polar code is proposed as channel coding by Erdal Arkan. It can achieve the symmetric capacity with low encoding and decoding complexity [8]. This paper focuses on LDPC code applies with Polar code for image transmission to protect the information from noise and keep efficiency for data transmission. The images sent have error when distance increases. Therefore, this case needs some techniques for improvement of data transmission, and LDPC-Polar code is considered to enhance efficiency in difference distance. The result can find BER and PER to keep the performance at the image destination.

This paper is organized as follows. In Section III, LDPC code algorithm is introduced. In Section IV, we present the methodology. The simulation results are in Section V. Finally, the conclusion is made in Section VI.

### II. POLAR CODE

#### A. Principle of Polar code

It's the unique construction method, named channel polarization; make polar code has the capacity achieving property. By channel polarization, the bit channel will be divided into noiseless bit channels or pure noise bit channel. In the polar code design, only those noiseless bit channels are selected to transmit information. As a family of block codes, polar code can be described as

$$x_1^N = u_1^N G_N \quad (1)$$

Where  $N$  is the block length and is always set to  $2^n, n \geq 0$ .  $G_N$  is the generator of polar code, which comes from the  $n$ -th Kronecker power of. The standard matrix of is

$$G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (2)$$

Polar code can be defined by difference parameter: block length  $N$ , rate  $R = KN$ , and information set  $A$ , denote as  $(N, K, A, u_A^c)$ . Here,  $u_A^c$  is frozen bit. They are kept constant and the decoder can avoid errors in the frozen part. A polar codeword  $x$  may be obtained by mapping a binary  $N$ -tuple  $u$  whose  $i$ th component is set ("frozen") to zero for all  $i \in A$ . Hence, (1) can also written as

$$x_1^N = u_1^N G_N(A) \oplus u_A^N c G_N(A^c) \quad (3)$$

Where  $G_N(A)$  is sub-matrix of formed by the rows with indices of  $A$ . Assuredly, polar codes are channel specific designs. A construction method of polar codes for one channel may be not suitable for another.

The noiseless or pure noise of bit channel is determined by the Bhattacharyya parameter  $Z(W)$  of actual channel  $W$ . For any discrete memory less channel (DMC) channel,  $Z(W)$  defined as

$$Z(W) = W(y_0)W(y_1) \quad (4)$$

And the block error probability of polar code is proved to

$$P_e(N, K, A, u_A^c) \leq Z(W_A^i) \quad (5)$$

### III. LDPC CODE

#### A. LDPC Encoding

The LDPC encoding has a linear complexity. The encoding transforms a message into a codeword. For systematic code, this means adding parity bits. A simple scheme is to exploit the relationship between codeword and the parity-check matrix  $\mathbf{H}$ . We define a codeword with  $c$  dimension of  $1 \times n$  and the corresponding matrix parity check  $\mathbf{H}$  with dimension of  $(n-k) \times n$  and  $\mathbf{0}$  are a vector with dimension of  $(n-k) \times n$ . The encoding process uses modulo-2 addition or exclusive-or (XOR) operation. The relationship between the parity matrix  $\mathbf{H}$  and codeword  $c$  can be written as equation (6).

$$c\mathbf{H}^T = \mathbf{0} \quad (6)$$

Equation (7) is to transpose both sides of equation (6).

$$\mathbf{H}c^T = \mathbf{0}^T \quad (7)$$

We define a systematic codeword  $c$  as the follows,

$$c = [p_1 \ p_2 \ \dots \ p_{n-k} \ | \ m_1 \ m_2 \ \dots \ m_k] = [\mathbf{p} \ \mathbf{m}] \quad (8)$$

Where the parity bits are positioned at the front part and the message bits are at the back part. In this scheme, the encoding can be done efficiently.

The transpose of codeword  $c$  in equation (9) can be written as

$$c^T = \begin{bmatrix} \mathbf{p} \\ \mathbf{m} \end{bmatrix} \quad (9)$$

the parity-check equation can be derived, for example, given that a parity-check matrix  $\mathbf{H}$  of a (3, 7) code is

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (10)$$

Equation (8) is rewritten as

$$\mathbf{H}c^T = \mathbf{0}^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (11)$$

Then the parity-check equations are

$$\begin{aligned} p_1 + m_2 + m_3 + m_4 &= 0 \\ p_2 + m_1 + m_2 + m_4 &= 0, \\ p_3 + m_1 + m_3 + m_4 &= 0 \end{aligned} \quad (12)$$

where '+' is a modulo-2 addition or an XOR operation. The parity bits can be then found from

$$\begin{aligned} p_1 &= m_2 + m_3 + m_4 \\ p_2 &= m_1 + m_2 + m_4, \\ p_3 &= m_1 + m_3 + m_4 \end{aligned} \quad (13)$$



B. LDPC Decoding

The decoding algorithm used for LDPC codes was discovered independently and comes under different names.

The most common are the belief propagation algorithm, the message-passing algorithm and the sum-product algorithm. In the Log-domain Sum-Product algorithm, the message passes between check nodes and variable nodes. In each pass the log likelihood ratio is recorded for its probability of its likely symbol.

Tanner graph is an intuitive way in understanding the LDPC decoder. The graph can be drawn directly from the H matrix as shown in Fig. 1

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

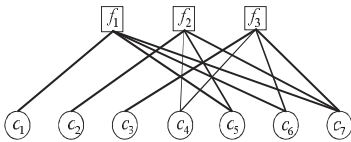


Fig. 1 Parity-check matrix and its Tanner graph

IV. METHODOLOGY

The configuration of image transmission system is shown in Fig 2. We use 512x512-pixels 8 bits gray of Lena, Baboon and Barbara images, and they are tested with various levels of noises. We use code length is 1024. In order to evaluate the performance of these error-correcting codes, first the source image is encoded by using JPEG compression. After JPEG encoder, the compressed image is encoded by LDPC-Polar codes. The data is then modulated using binary phase shift keying (BPSK) scheme before it is transmitted via additive white Gaussian noise (AWGN) channel. The transmission is then improved in efficiency by sub block encoder designed using difference noises such as fading channel. The reverse process is performed by the receiver.

V. SIMULATION RESULTS AND DISCUSSIONS

In this section, the simulation results of the image transmission using the proposed LDPC-polar code as channel coding with BPSK modulation via fading channel at various noise levels are shown.

We examine the system performance with subjective received images visualization, bit-error-rate (BER) and peak-signal-noise-ratio (PSNR), respectively. In the simulation, Lena image is tested with various levels of noises at half rate in order to evaluate the performance of the proposed channel coding.

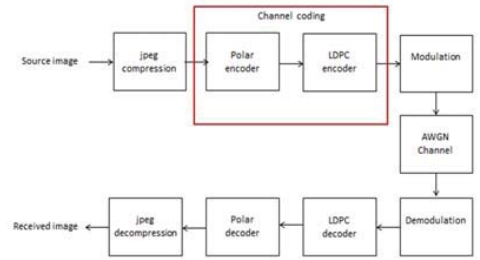


Fig. 2The configuration of image transmission

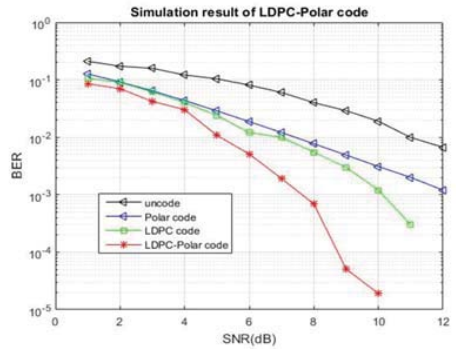


Fig. 3Image transmission BER of three-channel code.

As the result of the bit error rate curves in Fig 3, LDPC-Polar code is proved as the best one, which can achieve good performance for image transmission. The peak signal-to-noise ratios of the received images are shown in Fig.4. Fig. 5 shows the comparison of the received images.

The objective measured quality of the received images can be evaluated by PSNR calculated from the following equation (14).

$$PSNR = \frac{1}{L \times P} \sum_y \sum_x \left[ g(x, y) - \hat{g}(x, y) \right]^2 \quad (14)$$

where  $g(x, y), \hat{g}(x, y)$  represent the gray-scale values of any pixels in an original image and a recovered image, respectively.  $L, P$  represent the width and height of the image, respectively.

From the simulation results, it shows that the LDPC-Polar codes perform well and achieves a good performance in image transmission over AWGN channel. The BER curves of the LDPC-Polar code are shown in Fig. 3, where

Code 1 and Code 2 have the same code length is 1024. We can see that when BER 10<sup>-2</sup>, the performance of coding gain Code 1 and Code 2 is a bit difference. But coding gain of Code 3 is 5 dB. The performance of coding gain Code 3 is 1 dB better than Code 1 and code 2. The PSNR curve of the recovered images at three kinds are shown in Fig.4

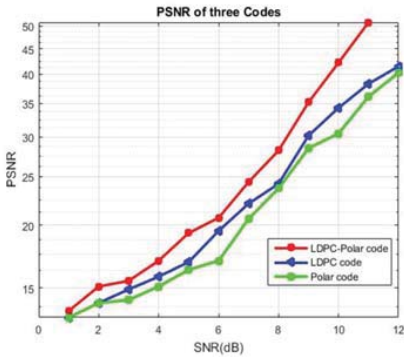


Fig. 4 PSNR of three codes.



Polar code



LDPC code



LDPC-Polar code

Fig.5 Comparison of the transmitted images for three codes at the SNR equal 8 dB.

The Comparison of the transmission images at  $E_b/N_o = 8dB$ , The PSNR of code 1, Code 2 and Code 3 are about 28.5, 24.5 and 24, respectively. It is shown in Fig. 5. The proposed LDPC-Polar code can achieve satisfied performance to keep a good quality of the transmitted image.

VI. CONCLUSIONS

The channel coding applied as LDPC-Polar code to image transmission via AWGN has been presented in this paper. The simulation results show a good performance for correcting data error. The proposed LDPC-Polar code can reduce bit error at various levels of noise and keeps sufficiently quality of the received image. The objective performance is evaluated by BER and PSNR.

REFERENCES

- [1] R. G. Gallager, "Low-Density Parity-Check Codes". MIT Press, Cambridge, MA, 1963.
- [2] D. J. C. Mavkay and R. M. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes", Electronic Letters, Vol. 32, No. 18, pp.1645-1646, August 1996.
- [3] D.J.D.Mackay, M.C.Davey., "Evaluation of Gallager Codes for Short Block Length and Hige Rate Application", In Codes. System and Graphical Models.Ed., vol. 123 of IMA, Springer-Verlag, New York, pp. 113-130, 2000.
- [4] W.E.Ryan, Fan, Wang, e.t., "Optimal Code Rates for the Lorentzian Channle: Shannon Codes and LDPC Codes", IEEE Transactions on Magnetics, vol. 40, pp. 3559-3565, 2004.
- [5] H.Yongqing, P.ENG Qicong and S.Huaizong, "The Application of High-Rate LDPC Codes in Image Transmisson over Wireless Channel", pp.62-65, 2006.
- [6] T.Payommai, W.Chiracharit and K.Chamnonthai, "Sub-Block Encoder Design of High-Rate QC-LDPC code for Image Transmission", APSIPA ASC 2011, China, pp. 1-4, October 18-21, 2011.
- [7] T.Payommai, W.Chiracharit and K.Chamnonthai, "Improved Sub-Block Encoder Design of Irregular LDPC code for Image Transmission", SH 2011, Japan, December 20-22, 2011.
- [8] S. korada, E. Sasoglu and R. Urbanke, "Polar code: Characterization of exponent, bounds, and construction", IEEE International Symposium on Information Theory(ISIT), pp. 1483-1487, 2009.

# ICMP Based IP address Recovery Method for DHCP

Mayoon Yaibuates, Ravi Upa, Rounsang Chaisricharoen

School of Information Technology  
Mae Fah Luang University  
Chiang Rai, Thailand

mayoon@outlook.com, ravi@technocom.co.th, rounsang.cha@mfu.ac.th

**Abstract**— The Dynamic Host Configuration Protocol (DHCP) server is one of the most important network infrastructures. It is used to provide Internet Protocol (IP) address and network parameters dynamically to user device. Since an IP address is limited network resource, a DHCP starvation attack is the method used by an attacker to exhaust all of the available IP Addresses on a DHCP server. This paper proposes a simple and effective method for recovery the IP address that had taken by an attacker. The proposed method applies the Internet Control Message Protocol (ICMP) echo and echo reply message for differentiating between the IP address – has been taken by legitimate network host from an attacker. Moreover, we analyzed our proposed method using probabilistic model.

**Keywords**—DHCP; DHCP Starvation attack ; ICMP echo message; IP address Recovery; Network Security

## I. INTRODUCTION

An IP address is one of the most important network parameters needed by every network devices. Without the IP Address, the network device cannot communicate and obtain the services provided in the computer network. There are two common ways for providing IP Address to the network device: 1) Users configure the IP Address on the network device manually and 2) The DHCP provides the IP Address via network protocol automatically.

The DHCP is a client-server based network protocol that was designed to handle the task for delivering not only IP address but also network configuration parameters such as IP address of default gateway, IP address of Domain Name Service (DNS) Server, and so on. The protocol relies on User Datagram Protocol (UDP). The UDP port number 68 is used by the DHCP Server whereas the UDP port number 67 is used by DHCP client [1].

The DHCP was developed during the early stage of computer network which may lack of security concern, the following protocol has been used by attackers to attack the network [2-4]. DHCP Starvation attack is the simple way for an attacker to prevent the attacked device from accessing to network and services. To perform DHCP Starvation attack, the attack will send several of DHCPDISCOVER packets with spoofing Media Access Control Address (MAC Address) in order to exhaust all of the available IP Addresses on the DHCP

Server. After DHCP Starvation attack had already run successfully, an attacker set up their rouge DHCP server by turning the DHCP Server function on and provided the malicious network parameters such as default gateway, DNS server to the victim. As a consequence, an attacker can capture, modify, and analyze any information which have been sent from the attacked device.

A number of researches have been proposed to mitigation the DHCP Starvation attack. The authentication has been proposed for authenticating the DHCP message between DHCP client and server [5-8]. Even though the authentication can be an efficient way to verify a legitimate client and server, it requires more complex environment to fulfill objectives. In fact, the secret key or certificate used between DHCP client and server must be share before the initial normal DHCP operation. Subsequently, it is hard for network administrators or end users to handle this task when it was moved to another network. The network vendor, Cisco and Juniper Networks, Inc. agree that configuring the security feature in network switch can mitigate DHCP starvation attack [9-10]. However, using port security requires the implementation of an additional expensive network switch [11]. Most of the mitigation techniques were focus on the way to prevent the DHCP Server from being starved while none of any solutions mentions the recovery solution after the DHCP server is being starved.

The purpose of this paper is to propose a simple and effective method for recovery the IP address that taken by an attacker. The proposed method applies the use of ICMP echo and echo reply message for differentiate between the IP address had taken by legitimate network host from an attacker. If the network host be able to respond back to the ICMP echo message, it can be imply that this IP address had taken by legitimate user. Otherwise, the IP address had taken by an attacker.

Apart from the introduction section, this article is divided into 6 main sections as follows. Section II describes the necessary background about DHCP and ICMP. Section III introduces the proposed idea of IP address recovery method. Sections IV analyzes the system in term of probability. Sections V shows the experimental value and discussion. Finally, Section VI presents conclusion and some consideration for future work.

II. BACKGROUND

A. DHCP overview

DHCP, a client-server based network protocol, has been used for automating assigns the network configuration parameters of TCP/IP implementation system. DHCP was developed by the Internet Engineering Task Force (IETF) Dynamic Host Configuration (DHC), is a working group that is recognized as part of the Daft Standard in 1997. The protocol consists of two parts - DHCP server and DHCP-enabled client. DHCP server is responsible for allocating network address and other protocol stack configuration parameters to DHCP-enabled client. The protocol relies on User Datagram Protocol (UDP). The UDP port number 68 is used by the DHCP Server and the UDP port number 67 is used by DHCP client. Normally DHCP protocol provides a hand check by using 4 DHCP messages which are DHCP Discover, DHCP Offer, DHCP Request, and DHCP Ack exchanged between client and server to assign network configuration parameters automatically Figure 1 demonstrates the process of DHCP operation as the following:

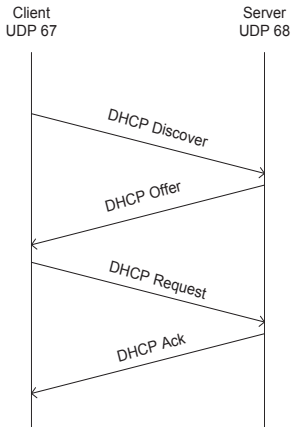


Figure 1. The operation of DHCP.

B. DHCP Starvation Attack

DHCP Starvation attack is considered as Denial of Service (DoS) attack. It is an easy way for an attacker to prevent the attacked device from accessing to network and services. Yersinia is well-known tools for able launch DHCP Starvation Attack. To launch the DHCP Starvation attack, the attacker will send many DHCPDISCOVER packets with spoofing Media Access Control Address (MAC Address) for consuming all of the useable IP addresses on the DHCP Server. After DHCP Starvation attack had done, an attacker set up their own DHCP server, Rouge DHCP, by turning the DHCP Server function on and provided the malicious network parameters such as default gateway, DNS server to the victim.

Thus, an attacker can capture, modify, and analyze any information that sent from attacked device.

C. ICMP overview

Internet Control Message Protocol (ICMP) is a network protocol that is designed for working with Internet protocol (IP) due to the unreliable of IP. The ICMP packet is encapsulated inside the payload of IP datagrams before leaving to lower network layer. The purpose of ICMP is to provide the information about problems which often occur in the transmission environment [12].

Ping is one of the most famous applications that implemented by using ICMP echo and echo reply for checking the network status. The packet size of ICMP Echo service header is 8 Bytes. The Information in Type field is used for differentiating between echo and echo reply message, 8 represents echo message while 0 is echo reply message as shown in figure 2.

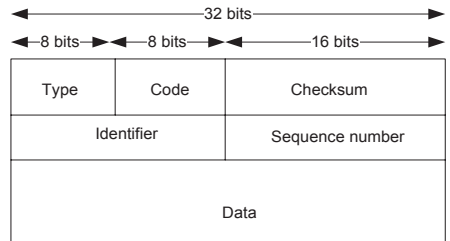


Figure 2. The structure of ICMP echo and echo reply message.

III. PROPOSED METHOD

The proposed IP address recovery method should able to identify the IP address which had taken by an attacker. ICMP echo service is introduced as a method for detecting the status of the network host. Our proposed method will collect all offered IP address then send ICMP echo message to the network host that hold IP address issued from the DHCP server. As ICMP echo service has been implemented in every internet devices [13]. Therefore, the legitimate network host should be able to respond back with ICMP echo reply message while the IP address that had taken by an attacker will never respond. Figure 3 demonstrates the process of IP address recovery from an attacker as shown:

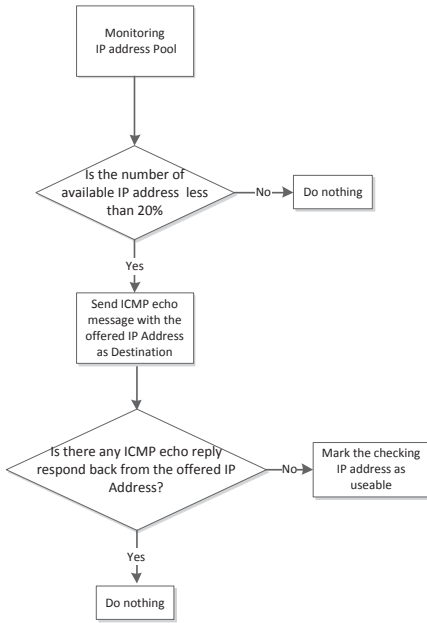


Figure 3. Flowchart of IP address recover method.

The proposed IP address recovery system will monitor the available IP address in the pool to notice if the available is less than the define threshold which is 20% left then the system will send ICMP echo message to all allocated IP address and wait for ICMP echo to reply in a specific period of time. If the measured number of ICMP echo reply is greater than 0, it will mark that IP address as had taken by legitimating user. Otherwise, the IP address will be free and mark as available for serving to future IP address request.

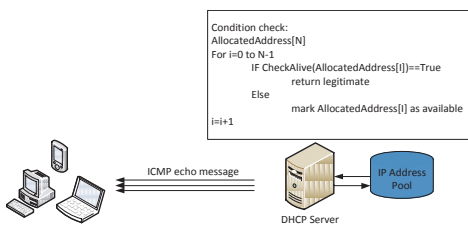


Figure 4. System diagram of IP address recovery.

#### IV. SYSTEM ANALYSIS

According to RFC792, in ICMP echo operation, the host that received the echo message must return the copy information

back to the sender using echo reply message. Normally the application like ping sends 3 ICMP echo messages to the destination host for determining the status. The receiver requires receiving at least one ICMP echo message from the sender to respond the received data back. So, the probability for the receiver to receive any ICMP echo messages out of 3 sent messages can be expressed as the following:

First, the first ICMP echo message was sent. The probability for the receiver receiving any ICMP echo messages from sending one ICMP echo message is expressed as:

$$P_{EM} = p_{em1} \tag{1}$$

Where:

$P_{EM}$  is probability for the receiver to receive any ICMP echo messages.

$p_{em}$  is probability of each ICMP echo messages to be received by the receiver.

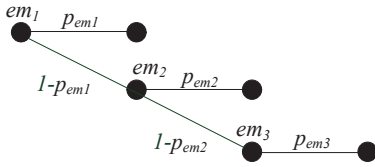
$1-p_{em}$  is probability of each ICMP echo messages unable to reach the destination.

$em$  is ICMP echo message.

However, in a case that the first ICMP echo message was not able to reach the receiver, the probability for the receiver for receiving any ICMP echo messages from sending two ICMP echo messages can be written as:

$$P_{EM} = p_{em1} + (1-p_{em1})p_{em2} \tag{2}$$

Finally, if none of the messages reach the receiver, all of the probability for the receiver to receive any ICMP echo messages can then be calculated by using the equation (3):



$$P_{EM} = p_{em1} + (1 - p_{em1})p_{em2} + (1 - p_{em1})(1 - p_{em2})p_{em3} \quad (3)$$

From equation (3) Let:  $p_{em1} = p_{em2} = p_{em3} = p_{emi}$ ; The probability for the receiver to receive any ICMP echo messages from sending 3 ICMP echo messages can be written as:

$$P_{EM} = p_{emi}^3 - 3p_{emi}^2 + 3p_{emi} \quad (4)$$

The probabilistic model for the receiver to receive any ICMP echo messages from sending 3 ICMP echo messages can be computed as:

$$P_{EM} = p_{emi}(3 - 3p_{emi} + p_{emi}^2) \quad (5)$$

An error in term of false positive error ( $P_{err}$ ), hold IP address by a legitimate client had been seen as an attacker. The probability of the error from our proposed system can be computed as:

$$P_{err} = 1 - P_{EM} \quad (6)$$

V. EXPERIMENT AND DISCUSSION

In this research experiment, IP address request was sent with the varied of  $p_{emi}$  (0.1 to 1) by 100 legitimate DHCP clients for three rounds. The values of  $P_{EM}$  from the experiment ( $P_{EM} Exp$ ) compare to  $P_{EM}$  from calculation ( $P_{EM} Cal$ ) are shown in table I.

TABLE I. THE VALUES OF  $P_{em}$ ,  $P_{err}$  FROM CALCULATION AND EXPERIMENTAL

$p_{emi}$	$P_{EM} Exp$			Avg. of $P_{EM} Exp$	$P_{EM} Cal$	$P_{err} Exp$	$P_{err} Cal$
	Rnd1	Rnd2	Rnd3				
0.1	0.3	0.26	0.33	0.297	0.271	0.703	0.729
0.2	0.57	0.43	0.37	0.457	0.488	0.543	0.512
0.3	0.64	0.58	0.59	0.603	0.657	0.397	0.343
0.4	0.79	0.8	0.81	0.8	0.784	0.2	0.216
0.5	0.88	0.92	0.9	0.9	0.875	0.1	0.125
0.6	0.97	0.91	0.94	0.94	0.936	0.06	0.064
0.7	0.99	0.99	0.96	0.98	0.973	0.02	0.027
0.8	0.98	0.97	1	0.983	0.993	0.017	0.007
0.9	1	1	0.99	0.997	0.999	0.003	0.001
1	1	1	1	1	1	0	0

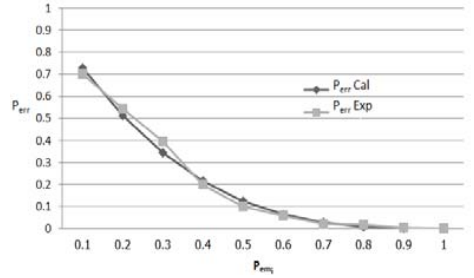


Figure 5. The value of  $P_{err} Cal$  versus the value of  $P_{err} Exp$ .

Figure 5 indicated the similarity of the value of  $P_{err} Cal$  and  $P_{err} Exp$ . The values of the  $P_{err} Exp$  agreed well with the calculated values. Moreover, the value of  $P_{err} Cal$  and  $P_{err} Exp$  depended on the value of  $p_{emi}$  in the same direction. On one hand, if the value of  $p_{emi}$  increases, the value of  $P_{err} Cal$  and  $P_{err} Exp$  will decrease. On another hand, if the value of  $p_{emi}$  decreases, the value of  $P_{err} Cal$  and  $P_{err} Exp$  will increase. The results were then can be conclude that the accuracy of the proposed IP address recovery system depends on the values of  $p_{emi}$  which can recognize from the network bandwidth status. However, the most current local network is implemented by gigabit networks cable (CAT5E and CAT6). Therefore, the value of  $p_{emi}$  would be 1 or almost 1.

VI. CONCLUSIONS

This paper proposed IP address recovery method that could be able to recover the IP address which had taken by an attacker. Since ICMP echo service has already implemented to any network devices, we applied the use of ICMP echo message for recovering the IP address. However, the prevention method still needs to combine with the recovery method in order to make the DHCP systems more resistant to the DHCP Starvation attack. Our proposed method can be applied to any existing DHCP Starvation attack prevention techniques. In future work, we will set up the experiment to examine how much time used by our proposed method for recovery the IP address.

REFERENCES

- [1] R Droms, "Automated configuration of TCP/IP with DHCP", Proceedings of IEEE Internet computing, 1999, pp.45-53.
- [2] M. Yaibuates and R. Chaisrichaoren, "ICMP Based Malicious Attack Identification Method for DHCP", Proceeding of the 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (IJCTEE), Chiangrai, 2014, pp. 1-5.
- [3] H. Altunbasak, S. Krasser, H. Owen, J. Sokol, and J. Grimminger, "Addressing the weak link between layer 2 and layer 3 in the Internet architecture", Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, 2004, pp.417-418.
- [4] L. Senecal, "Understanding and preventing attacks at layer 2 of the OSI reference model", Proceedings of the 4th Annual Communication Networks and Services Research Conference (CNSR 2006), 2006.
- [5] H.Ju, and J. Han, "DHCP Message Authentication with an Effective Key Management", International Journal of Electrical, Computer, Electronics and Communication Engineering Volume 1 Issue 8, 2007, pp.1187-1190.

- [6] S. Duangphasuk, S. Kungpisdan, and S. Hankla, "Design and implementation of improved security protocols for DHCP using digital certificates", Proceeding of the 17th IEEE international conference on Networks(ICON), 2011, pp. 287-92.
- [7] D. D. Dinu and M. Togan, "DHCP server authentication using digital certificates", Proceeding of the 10th International Conference on Communications (COMM), 2014, pp. 1-6.
- [8] O.S.Younes, "A Secure DHCP Protocol to Mitigate LAN Attacks", Journal of Computer and Communications Volume 4, 2016, pp.39-50.
- [9] Juniper Networks, "Example: Configuring MAC Limiting to Protect the Switch from DHCP Starvation Attacks", [Online], Available: [https://www.juniper.net/techpubs/en\\_US/junos12.3/topics/example/port-security-protect-from-dhcp-starvation-attack.html](https://www.juniper.net/techpubs/en_US/junos12.3/topics/example/port-security-protect-from-dhcp-starvation-attack.html)
- [10] Cisco, "Using Port Security to Mitigate a DHCP Starvation Attack", [Online], Available: <http://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/emob30dg/SecInteg.html#wp1005858>
- [11] D.Genkov, "Implementing port security feature using SNMP protocol" Proceeding of the 14th International Conference on Computer Systems and Technologies, 2013, pp. 38-45.
- [12] J. Postel, "INTERNET CONTROL MESSAGE PROTOCOL", RFC 792, 1981, pp. 1-21.
- [13] R. Braden, "Requirements for Internet Hosts -- Communication Layers", RFC 1122, 1989, pp. 42.

# Inter-cell interference avoidance by active coordination of scheduling for downlink on LTE like systems

Ján Kulík

Softea, s. r. o.  
Bratislava, Slovakia  
[jan.kulik@softea.sk](mailto:jan.kulik@softea.sk)

Peter Farkaš

Institute of Telecommunications,  
Slovak University of Technology  
Bratislava, Slovakia  
[p.farkas@icee.org](mailto:p.farkas@icee.org)

**Abstract**—In this manuscript a new simple approach to inter-cell interference mitigation for the OFDMA based cellular networks is proposed, where the prolonged scheduling period and exchange of scheduling information among neighboring eNBs is used to reduce the interference among cells. Exchange of information could be carried over an interface similar to X2 that is standardized for the LTE systems. The process is described with the optimization by weighing the received scheduling information. The advantages of the proposed scheme are demonstrated by simulation results and compared to random resource allocation and soft frequency reuse.

**Keywords**—*inter-cell interference mitigation; interference avoidance; scheduling; X2 interface*

## I. INTRODUCTION

Several approaches [1-4] have been proposed for the solution of interference problem. First group is built around static allocation of resources and is categorized as interference coordination. Neighboring cells are using different sets of resources (usually in patterns), thus the interference does not occur. The same sets are reused among more distant cells and the power of interfering signals fades with distance. To improve the reuse factor of frequency resources, other approach was proposed [5], to share the resources for the users within center of the cell as these are further from sources of interference than cell edge users. Only fractions of resources, those that are dedicated for the cell edge users, are cell specific, thus improving the reuse factor. This is combined with the power allocation where the transmission to central users is done with reduced power to decrease the interference emission to neighboring cells. The cell edge users are given more power to compensate the path loss. Another approach is known as interference averaging and includes techniques such as scrambling, interleaved division multiple access (IDMA) or different methods of frequency hopping. The first two mentioned do not reduce the interference on the physical layer but instead spread its impact on the whole transmitted sequence. This decreases the probability of error during the decoding. Frequency hopping is based on pseudo random assignment of resources. Neighboring cells are using quasi-orthogonal sequences and the interference is mitigated by design. The most complex approach is interference

cancellation. The interference is removed from the received signal in the user equipment (UE) if the inter-cell channel state information (CSI) and interference estimation is known. The main advantage of this method is the ability to remove interference even in systems with high load and the performance increases with the usage of multiple antennas. The performance however drops if there are no dominant interferers or the CSI is inaccurate. ICIC techniques are already included in 4G LTE since Rel. 8 [6]. Recently enhanced ICIC (eICIC) technique was proposed considering also usage of massive MIMO [7]. The research area is identified as one of the main challenges also in connection with 5G development [8].

The new very simple technique is proposed in this paper, which could be categorized as dynamic interference coordination. It minimizes the interference on physical layer while it still maintains high resource reuse. The performance was verified by simulations and compared with two standard interference mitigation schemas the frequency hopping and Siemens proposed schema [6].

## II. INTERFERENCE AVOIDANCE SCHEME BY ACTIVE COORDINATION

While today systems schedule the data transmission in shortest possible intervals usually in the length of transmission time interval (TTI) 1ms. The proposed schema considers much longer scheduling period e.g. 15ms as used in simulations. This scheduling could not be called “fast scheduling” on optimal resources. However, neglecting up to date CQI proves to be beneficial especially in heavy loaded system. The reason for this prolongation is the extra time that the eNBs are given to exchange the information about the resources that are being used and process it. The information that needs to be transmitted consists of  $n \times m$  matrix where  $n$  denotes the number of resource blocks and  $m$  stands for number of TTI's for which scheduler is doing planning (Fig. 1). Each position  $x_{i,j}$  in  $i$ -th row and  $j$ -th column of the matrix indicates whether the eNB will transmit in specified TTI and RB.

This information would be however to no use if the schedulers in neighboring cells were working synchronously.



Thus the neighboring cells that are dominant interferers to each other have time-offset schedulers.

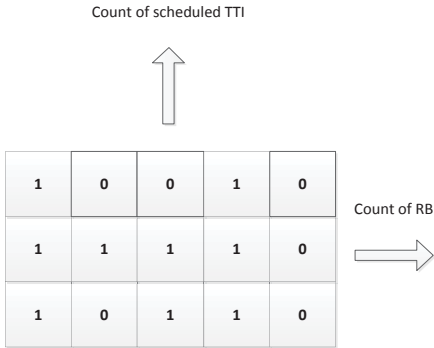


Fig. 1. Transmission indication information matrix (TIIM)

Above described process of active coordination could be adopted to fit the needs of scheduling in cells with multiple sectors. If we denote sectors of hexagonal cells clockwise A, B and C the resulting scheduling algorithm is the same as if the A sector was cell no 1, sector B was cell no 2 and sector C was cell no. 3 (the pattern is seen in Fig. 3).

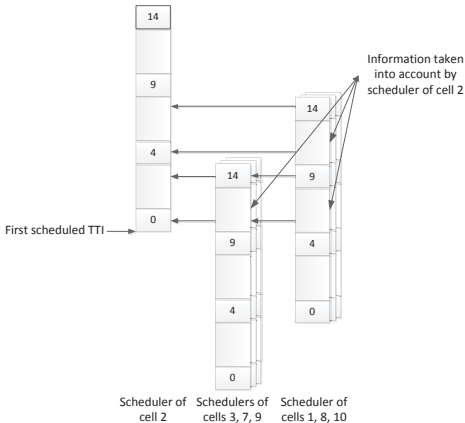


Fig. 2. Scheduler of cell no. 2 takes into account scheduling information from the surrounding cells.

To better describe the scheduling process let us assume that cells 1, 8, 10 etc. start to do the scheduling at the time  $\tau$ . Cells 2, 11, 4 etc. start to do the scheduling at the time  $\tau + 5ms$  and cells 7, 3, 5 etc. start at  $\tau + 10ms$ . So if the cell no. 2 is starting the scheduling for the next 15 transmission intervals, this process could be divided into three smaller stages. For scheduling the first 5 TTIs it already has a complete

transmission information from neighboring cells as cells 3, 7 and 9 are going to transmit the last 5 TTIs from their previous scheduling round and 1, 8 and 10 are going to transmit their 6-th to 10-th TTI. The scheduling for the first third of transmission is potentially the most limited stage as scheduler has to fully adapt to the transmission in surrounding cells. This fact can on the other hand help to maintain systems overall throughput as decreasing the rate or temporary complete stopping of transmission does not lead to heavy interference. Scheduling of the second 5 TTIs provides greater freedom for scheduling as the known interference comes from 10-th to 14-th TTI of cells 1, 8, 10. The scheduling of the last 5 TTIs is independent from the surrounding transmission and thus can be done optimally according to the CQI. As soon as the scheduler finishes the planning for the 15 TTI frame, it sends the information matrix indicating the oncoming transmission to the surrounding cells.

*Weighing of Transmission Indication Information*

One of the questions that arise in this active coordination scheme is what is the interference threshold when the scheduler should decide not to transfer due to interference? If the scheduler would not transmit on the resources that were indicated to be used by nearby cells, the total system throughput would decrease significantly.

To deal with this issue a weighing algorithm can be used to increase the information at scheduler to optimize the coordinated scheduling process in respect to interference in surrounding cells. It is used to distinguish the strong and weak interferers. To start with, the scheduler needs to maintain the interference indication matrix (IIM) that is of the same dimensions as TIIM. The scheduler stores within this matrix the indications of transmission from nearby sectors or cells. These are stored additively and with corresponding weights.

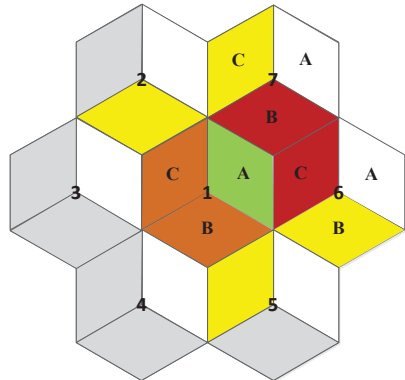


Fig. 3. Weighing of transmission indication as seen by 1st cells sector A

Let us assume we are starting the scheduling for the A sector of first cell (green sector Fig. 3). As we are using frequency reuse 1, sector B of cell 7 and sector C of cell 6 (red sectors) are potentially the strongest interferers due to distance

and overlapping radiation diagrams of directional antennas. Thus when the TIIM information comes from these sectors they are weighed by coefficient 4 and added to the IIM. Sectors B and C of cell no 1 are weighed by coefficient 2 due to common boundary with sector A. Users on the direct edge of sectors could experience interference of the same power as the power of useful signal. However the overlapping of radiation diagrams is small in terms of area coverage. The yellow sectors are the last ones from which the indication is taken into account and is weighed by coefficient 1. White sectors are doing scheduling synchronously and thus the information could not be transmitted.

A simplified case where scheduler has 3 RBs and plans for 3 TTIs ahead is described follows. Sector A of cell 1 is starting the scheduling and from previous scheduling rounds received TIIM information from sector B of cell 2 and sector C of cell 6 (from now denoted as sectors A, B and C). No other transmission is going on in surrounding cells. TIIM from sector B indicates transmission on 1-st and 3-rd RB of the last scheduled TTI which is the only relevant one for sector A. As this information is weighed by 1 the interference indication matrix adds 1 to 1st and 3-rd RB position of first planned TTI in IIM. TIIM from sector C indicates transmission on 1st RB of its second planned TTI and 2-nd and 3-rd RB of its 3-rd planned TTI. As these are weighed by coefficient 4, number 4 is added to the existing interference information for the 1-st RB for the 1st planned TTI of sector A resulting in total interference of 5. 2-nd and 3-rd RB of 2-nd planned TTI for sector A both received interference of 4. This sample can be seen on Fig. 4.

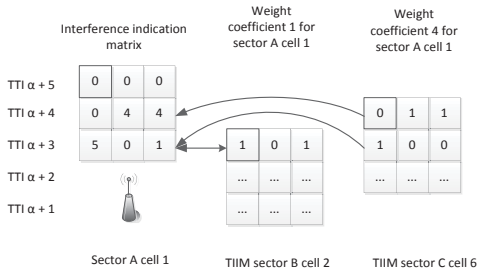


Fig. 4. Construction of interference indication matrix

By having the interference weights scheduler can decide if it should or should not transmit on the available resources and optimize the utilization of resources. In our simulation a threshold of 4 was set to avoid direct interference from opposing sectors. If total interference weight on specific resource is 3 scheduler would use the resource if there was no other option. E.g. eNB needs to transmit data that could be carried by 2 RBs. In such case scheduler of sector A of our sample (Fig. 4) would use 2-nd RB of 1-st TTI and 1-st RB of 2-nd TTI. This would potentially prevent creating interference if the 3-rd planned TTI was used for transmission and some of neighboring cells/sectors needed to use those resources for transmission. If the eNB needed to transmit data over 5 RBs

also the 3-rd TTI would be utilized. If 6RBs were needed the scheduler would use the RB with lowest indicated interference, in our case 3-rd RB in 1st planned TTI.

### III. SYSTEM MODEL AND STANDARD SCHEMES

All tested scenarios are based on conventional cell model consisting of three sectors defined by usage of directional antennas. System consist of 19 cells and to maintain approximately the same conditions of interference throughout the whole system, cell-wrapping technique was used (Fig. 5)

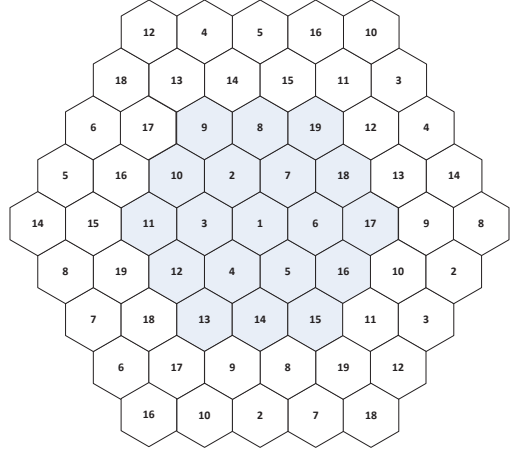


Fig. 5. Cellular model with cell-wrapping

Simulated bandwidth  $B = 10\text{MHz}$  is utilized as 50 resource blocks available for each transmitter to allocate. Maximum available output power of eNB was set to  $46\text{dBm}$ . Mobile radio channel used in simulations included path loss, shadowing and fast fading. Path loss was calculated as recommended in [1] using the formula for urban macro area

$$L = 128.1 + 37.6 \log_{10}(R) \quad (1)$$

, where  $R$  is distance in kilometers. Shadowing parameters were as per recommendation in [1]: standard deviation set to  $8\text{dB}$ , and correlation distances of  $50\text{m}$ . CODIT macro-cell model was used as a propagation model for fast fading. The monitored output of the simulation was average signal to interference plus noise ratio (SINR) generally calculated as

$$\text{SINR} = \frac{P_0 d_0^{-\alpha} 10^{\frac{a_0}{10}}}{N + \sum_i P_i d_i^{-\alpha} 10^{\frac{a_i}{10}}} \quad (2)$$

, where  $P_0$  is power of useful signal,  $d_0$  is distance of UE from the source of useful signal,  $\alpha$  is attenuation coefficient,  $P_i$  and  $d_i$  are power and distance from interfering signal,  $8\text{dB}$

is attenuation characteristics of the shadowing at the point of UE and N is the power of white noise. All scenarios were simulated under different system load conditions where system held 60, 120 or 180 users. Users were moving at the speed of 3km/h. Full-buffer model was considered to simulate services such as video streaming or file transfer where large amounts of data are transmitted. This setting was intentional as it is expected that the proposed algorithm has limited usage if data is transmitted in short bursts. However these types of services are on the rise so these settings could become the real world scenario in short time.

*Random Channel Allocation*

Random channel allocation technique is used as a reference scenario. It belongs to the group of interference averaging techniques as it relies solely on distributing the average interference among users in time. Each user is allocated resource blocks chosen randomly from set of channels with best transmission properties, based on last statistical CSI report. Frequency reuse factor (FRF) of 1 is used. Transmission for each sector is planned separately thus each sector behaves as a separate cell. This method has obvious disadvantage, if the system works under heavy load where it could get into blocking state.

*Soft Frequency Reuse*

Soft Frequency Reuse (SFR) is a scheme that is among the standardized interference mitigation methods [9] used in the 3GPP Release 8. The basic principle of this scheme is the division of the sector or cell into inner and outer zone. Users closer to the eNB are given resources on which lower transmission power is used. Cell-edge users have resources with higher transmission power to overcome the path loss and interference effects.

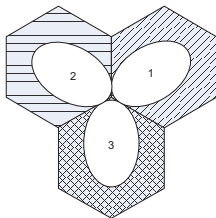


Fig. 6. Soft Frequency Reuse and its power allocation

Resources allocated to be used by cell-edge users within neighboring cells are orthogonal to each other. The FRF is near one.

IV. SIMULATION RESULTS

The constant load from 60, 120 and 180 users corresponds approximately to 20, 40 and 60 percent of total system load. As seen on Fig 7. and Fig. 8 the difference between the random allocation and soft frequency reuse is marginal for low system load. This statistical result could be slightly misleading as it does not show clearly the results for the most vulnerable group of cell edge users.

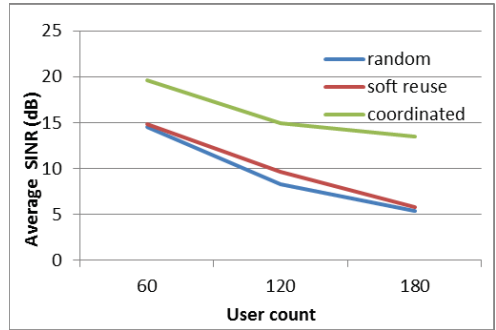


Fig. 7. Average SINR during constant load

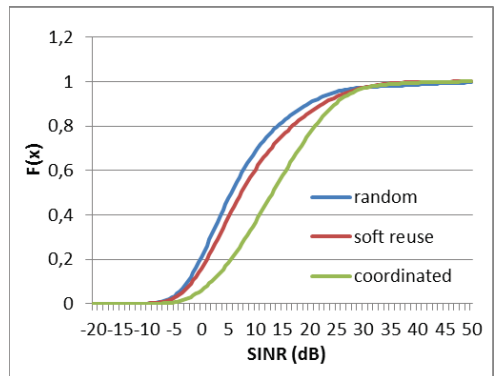


Fig. 8. Cumulative distribution function of a system with 120 users

The proposed scheme shows gain of 5dB compared to the first two scenarios. This is obviously achieved through the transmission on the ideal resources in terms of interference. The gain even increases with the raising system load. This is due to the fact that the active coordination is able to prevent the blocking state when all the eNBs are transmitting on similar resources. While indication of heavy interference is implement in the LTE standard (with the use of X2 interface), its simplistic realization does not allow to continue with the transmission while sustaining relatively high SINR values for all system users.

## V. CONCLUSIONS

Simulation results of different scenarios showed which direction could be the interference mitigation heading. Active coordination of scheduling allows transmission on nearly optimal resources and with combination with interference weighing provides means for balancing total throughput and channel quality. Increased complexity of scheduling is outweighed by gain in average SINR values resulting in higher throughput. As simulations were done for full buffer model with high data transfers, evaluation has to be done for services such as voice data, to validate if the gains are worth the tradeoff for complexity.

## ACKNOWLEDGMENT

This work was supported by EU RTD Framework Program under ICT COST Action IC 1104 and by Scientific Grant Agency of Ministry of Education of Slovak Republic and Slovak Academy of Sciences under contract VEGA 1/0518/13, by Slovak Research and Development Agency under contracts SK-AT-0020-12 and SK-PT-0014-12.

## REFERENCES

- [1] 3GPP TR 25.814 V7.1.0, "Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA) (Release 7)," September 2006.
- [2] 3GPP R1-050738, "Interference Mitigation - Considerations and Results on Frequency Reuse," Siemens, September 2005.
- [3] 3GPP R1-060291, "OFDMA Downlink Inter-Cell Interference Mitigation," Nokia, February 2006.
- [4] Y. Xiang, J. Luo, Ch. Hartmann, "Inter-Cell Interference Mitigation through Flexible Resource Reuse in OFDMA based Communication Networks", European Wireless 2007
- [5] S. Schwarz and M. Rupp, "Exploring Coordinated Multipoint Beamforming Strategies for 5G Cellular," in *IEEE Access*, vol. 2, no. , pp. 930-946, 2014.
- [6] K. Elissa, "Title of paper if known," unpublished.
- [7] Trung Kien Vu et al. Joint In-Band backhoulng and Interference Mitigation in 5G Heterogenous Networks, arXiv: 1604.02750v1 [cs.NI] 10 Apr 2016.
- [8] M. Jaber, M. A. Imran, R. Tafazolli and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," in *IEEE Access*, vol. 4, no. , pp. 1743-1766, 2016
- [9] 3GPP TSG-RAN WG1 R1-050507, "Soft Frequency Reuse Scheme for UTRAN LTE", Huawei, May 2005.

# Challenges in Wireless Technologies Coexistence in Smart Home Environments

A. Marinčić, A. Kerner, D. Šimunić

University of Zagreb, Faculty of Electrical Engineering and Computing,  
Department of Wireless Communications, Green Engineering Laboratory  
Unska 3, 10000 Zagreb, Croatia

**Abstract** – Due to technology development smart home environments become part of our lives. Mostly, smart home environments refer to environments which enable elderly people to live independently. In these environments numerous different devices are being used, whether we talk about medical devices (e.g. heart rate measurement, blood pressure measurement and SPO2 measurement devices) or other sensors (e.g. gas sensors, PIR sensors and illuminance sensors). They all communicate using different technologies. Those technologies are: ZigBee, Bluetooth, Bluetooth Low Energy and WLAN. These technologies should coexist in home environment. Interference becomes the crucial problem in enabling coexistence of these technologies and interoperability of devices using these technologies. In this work, the coexistence of different technologies in home environment using SEAMCAT simulation tool was investigated. To address the coexistence problem, probability of interference in dependency of the number of devices, transmitter power of devices, simulation radius and used technologies is presented. Also, probability of interference in dependence on relative location of interfering transmitters is studied.

**Keywords**—Coexistence, Wi-Fi, Bluetooth, Bluetooth Low Energy, ZigBee, Ambient Assisted Living

## I. INTRODUCTION

Due to technology development and modern way of life, smart home environments become an important part of everyday lives. Main goal of these environments is to enable elderly people to live independently. Smart home environments aim in enhancing the comfort, safety, entertainment and security of the residents. In recent decades, there has been a lot of research in developing smart home environments for healthcare purposes which would enable older people to live more independently.

Smart home environments for ambient assisted living require numerous devices. All of these devices communicate using different technologies. For example, in eWALL system Bluetooth, Bluetooth Low Energy, ZigBee and WLAN technology can be found. In order to deploy and enable these systems to work successfully, all of the above mentioned technologies need to coexist. Interference can be an important issue in these systems. In this work interference issues in smart home environments based on number of devices, transmitter power, simulation radius and relative positioning of interfering transmitter ( $I_t$ ) and victim receiver ( $V_r$ ) was investigated. We use SEAMCAT[1] simulation tool in order to investigate interference issues.

## II. TECHNOLOGIES COMPARISON AND SMART HOME ENVIRONMENT SYSTEM OVERVIEW

A great number of devices for enabling elderly people to live independently is what makes smart home environments a good solution for improving everyday lives. These devices use different technologies for communication. As an example of smart home environment eWALL home installation is described. In Figure 1. eWALL home installation is shown.

From Figure 1. it can be seen that in order to deploy smart home environments a great number of devices is used. Some of them use wire connections, and other wireless. In this work, emphasis is put on wireless communications between devices. As shown in Figure 1. devices used in eWALL system communicate using Bluetooth, Wi-Fi and ZigBee technologies. These devices need to coexist in indoor environment, like our home, so it is important to reduce interference and possibility of communication blocking. Interference occurs as a result of spectrum scarcity, because all of the above mentioned wireless technologies use 2.4GHz frequency band.

### A. Bluetooth (IEEE 802.15.1) Wireless Technology

Bluetooth is a wireless technology for short-range communication. It is a standard of data and speech transmission, which is characterized by low price, low consumption and it is aimed for short range data transmission up to 10 meters. With greater transmission power communication range goes up to 100 meters. IEEE 802.15.1 work group is in charge for Bluetooth standard development [4]. With Bluetooth standard it is possible to achieve transmission rate up to 1Mbps, and this standard uses frequency range from 2.4 to 2.4835 GHz.

### B. Bluetooth Low Energy Wireless Technology

Bluetooth Low Energy is an ultra-low powered feature of Bluetooth 4.0 wireless radio technology. Due to new features Bluetooth Low Energy is aimed at new, principally low power, low cost, low latency and easy to implement applications for wireless devices. This technology has a range of up to 50 meters. [5] Bluetooth Low Energy uses spectrum range from 2402 GHz to 2480 GHz. Difference between Bluetooth Low Energy and classic Bluetooth technology is that Bluetooth has 79 1

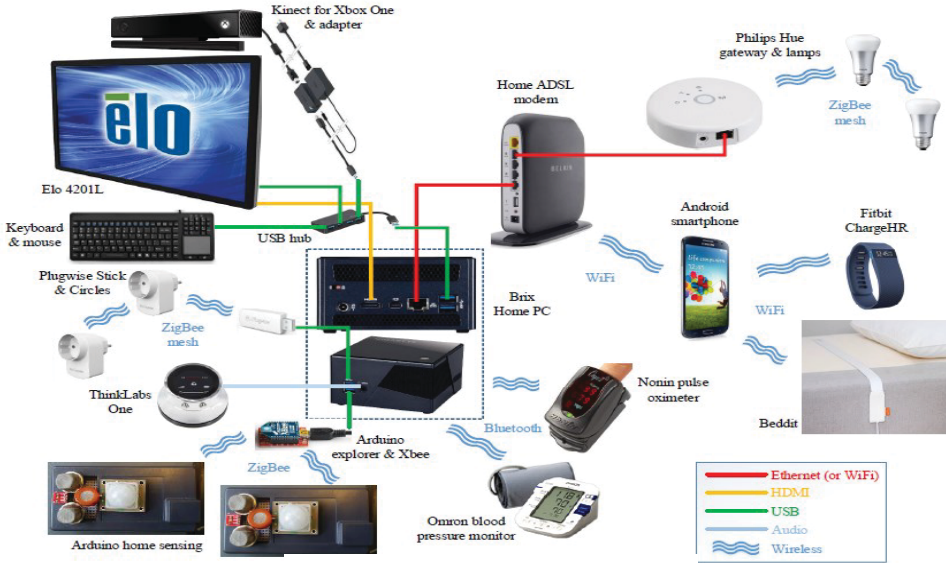


Figure 1. eWALL Home Installation [3]

MHz wide channels and Bluetooth Low Energy has 40 channels that are 2 MHz wide.

### C. Zigbee (IEEE 802.15.4) Wireless Technology

ZigBee was developed as a wireless technology for applications that require a low data rate wireless networking, low battery power consumption and low running costs. ZigBee is based on the IEEE 802.15.4 standard, which is created specifically for control and sensor networks. The data rate is limited to 250 kbps in the global 2.4 GHz Industrial, Scientific, Medical (ISM) band, 20 kbps in the 868 MHz band used in Europe, and 40 kbps in the 915 MHz band used in North America and Australia [6].

### D. Wi-Fi (IEEE 802.11) Wireless Technology

Wi-Fi is a short name for Wireless Fidelity and it refers to any type of IEEE 802.11 Wireless Local Area Network. All Wi-Fi standards use the unlicensed radio spectrum and these standards utilize spread spectrum wireless communication techniques to insure reliable operation in RF environments. Most common IEEE 802. standards are [7]:

- IEEE 802.11a
- IEEE 802.11b
- IEEE 802.11g
- IEEE 802.11n

## III. SEAMCAT SIMULATION TOOL

For purposes of interference investigation between Bluetooth, Bluetooth Low Energy, Wi-Fi and ZigBee technologies SEAMCAT simulation tool is used. In this chapter description of SEAMCAT simulation tool is given.

SEAMCAT (Spectrum Engineering Advanced Monte Carlo Analysis Tool) is a statistical simulation model that uses Monte Carlo analysis method to determine the potential interference between different radio communication systems [11].

SEAMCAT uses following approach to determine potential interference [11]:

- the user defines the distribution of possible values for the system and propagation parameters of the desired and interfering links;
- SEAMCAT uses these distributions to generate random samples based on the Monte Carlo method called snapshots of the subject parameters;
- For each snapshot SEAMCAT calculates the interfering and desired signal levels;
- SEAMCAT then calculates the probability of interference by comparing the relationship of desired and interfering signals at the victim receiver for each snapshot. With SEAMCAT we can consider following studies [11]:
- Sharing and compatibility studies on different equipment operating in the same or adjacent frequency bands;
- Evaluation of different systems transmit and receive masks;

- Evaluation of limits such as unwanted emissions, blocking and intermodulation levels.

IV. WIFI, ZIGBEE, BLUETOOTH AND BUETOOTH LOW ENERGY COEXISTANCE

In this chapter simulation setup for Bluetooth, Bluetooth Low Energy, and Wi-Fi and ZigBee technologies coexistence is given. Also results of interference probability in depending on transmitter density over km<sup>2</sup>, simulation radius, transmitter power and spatial distribution of interferers are given.

When we talk about interference in smart home environments, density of devices plays an important role. In first simulation dependency of interference probability on transmitter density was investigated. Tx density was altered from 1 to 30000. In this work we consider Wi-Fi to be a victim link, while Bluetooth, Bluetooth Low Energy and ZigBee are interfering links. There can be only one victim link, and multiple interfering links. In Table 1. and Table 2. simulation setups are given.

Table 1. Simulation setup for victim link

	Wi-Fi
Power (TX)	20 dBm
Antenna Gain (TX/RX)	0 dBi
Frequency	2.412 GHz – 2.484 GHz
Sensitivity	-74 dBm
Propagation model	Extended Hata (Urban, Indoor)
Room size	4x4
Coverage radius (TX)	0.1km
Density of TX	1

Table 2. Simulation setup for interfering links

	ZigBee	Bluetooth	Bluetooth Low Energy
Power (TX)	0 dBm	0 dBm	10 dBm
Antenna Gain (TX/RX)	0 dBi	0 dBi	0 dBi
Frequency	2.4 GHz-2.483 GHz	2.4 GHz-2.4835 GHz	2.4 GHz
Sensitivity	-99 dBm	-71 dBm	-70 dBm
Propagation model	Extended Hata (Urban, Indoor)	Extended Hata (Urban, Indoor)	Extended Hata (Urban, Indoor)
Room size	4x4	4x4	4x4
Coverage radius (TX)	0.1km	0.1km	0.1km

Simulation results are shown in Figure 2.

From Figure 2, it can be seen, that the probability of interference increases with the higher TX density. The greater the number of sensors in system is, the higher the probability of interference is. Also, due to its features and spectral map Bluetooth Low Energy technology has lower probability of interference, while Bluetooth technology reaches the highest probability of interference. For TX density up to 100 TX over km<sup>2</sup> probability of interference is similar for all technologies, but for higher TX density probability of interference gets rather high, and the difference in density between devices is greater.

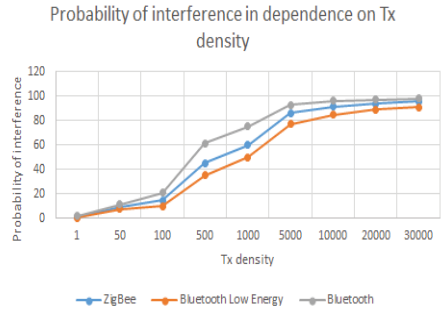


Figure 2. Probability of interference depending on TX density

Also, probability of interference in smart home environments based on spatial distribution of interferes is given in Figure 3.

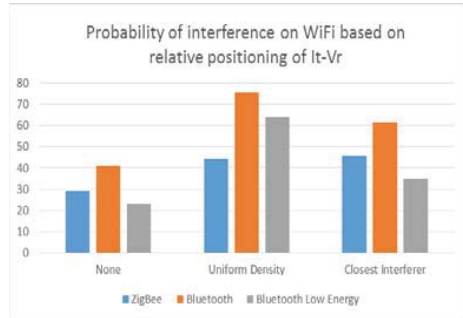


Figure 3. Probability of interference on WiFi based on relative positioning of It (interfering transmitter) to Vr (victim receiver)

In this scenario, three spatial distribution possibilities were considered. When mode “None” is used to represent positioning of It to Vr, active interfering transmitters are located in a circular area. Radius of that area is defined by user, and it is called simulation radius. Interfering transmitters are placed randomly in defined area. When mode “Uniform density” is chosen, active interfering transmitters are

uniformly located in the area with simulation radius. The simulation radius is now calculated based on interferer density and the number of active transmitters. When “Closest interferer” mode is chosen we have just one interfering transmitter. That interfering transmitter is randomly placed in a circular area. Simulation radius for that area is derived from user’s density.

From Figure 3. it can be seen that possibility of interference depends on relative positioning of interfering transmitters and victim receiver. Possibility of interference is the lowest when “None” mode is chosen, and highest when “Uniform density” mode is chosen. Also, it can be seen that no matter what relative positioning mode we chosen probability of interference is the highest for Bluetooth technology and lowest for Bluetooth Low Energy technology. The same result was obtained in first simulation when probability of interference was considered in dependency of transmitter density.

Also, probability of interference depends on simulation radius. Simulation radius defines the area on which probability of interference is examined. On Figure 5. probability of interference on Wi-Fi of different technologies in dependence on simulation radius is shown.

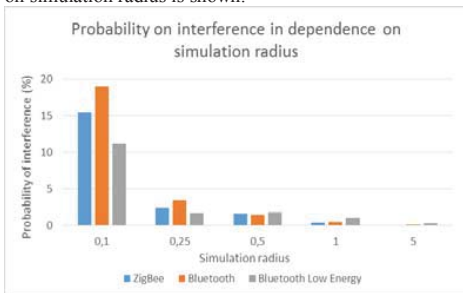


Figure 4. Probability of interference in dependence on simulation radius

It can be seen that with increasing the simulation radius interference decreases due to the fact that interfering links are distributed in a larger area.

Another parameter that could impact the probability of interference on Wi-Fi is transmitter power of different technologies. In order to determine influence of transmitter power, Bluetooth and ZigBee transmitter power were altered. Bluetooth technology can have three different transmitter powers. Based on that Bluetooth devices can be divided in three classes:

- Class 1: 0dBm
- Class 2: 4dBm
- Class 3: 20dBm

The goal of the simulation was to demonstrate how Bluetooth probability of interference on Wi-Fi changes with increasing the transmitter power. The results are shown in Figure 6.

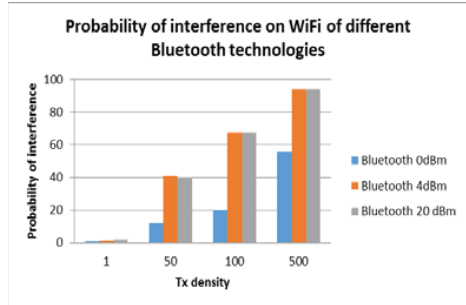


Figure 5. Probability of interference in dependence on Bluetooth transmitter power

From Figure 6. it can be seen that with increasing the power of Bluetooth technology from 0dBm to 4dBm the interference probability is higher, while with increasing the transmitter power from 4dBm to 20dBm probability of interference has similar values. In first case the probability of interference increased by 18% when Tx density equals one. When Tx density equals 50 probability of interference increased by 232%, when Tx density equals to 100, probability of interference increases by 236%, and finally when Tx density equals 500 probability of interference increases by 68 %.

Also, in order to determine ZigBee impact on probability of interference on Wi-Fi, ZigBee transmitter power was also altered. Two transmitter powers were considered: 0dBm and 20dBm. Results are shown in Figure 7.

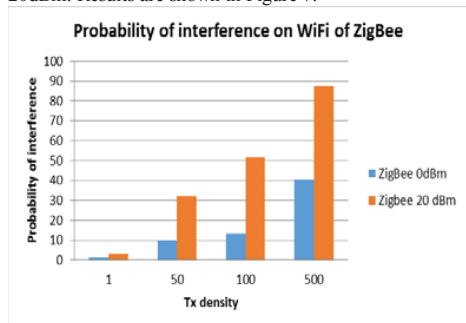


Figure 6. Probability of interference in dependence on ZigBee transmit power

In Figure 7. it is shown that by increasing the power of ZigBee transmitter probability of interference of ZigBee on Wi-Fi is also increasing. In first case the probability of interference increased by 145.7% when Tx density equals one. When Tx density equals 50 probability of interference increased by 222.61%, when Tx density equals to 100, probability of interference increases by 289.6%, and finally when Tx density equals 500 probability of interference increases by 117.15 %. When Bluetooth power is alternating,



probability of interference of Bluetooth Low Energy on Wi-Fi is also changed. Simulation results are shown in Figure 8.

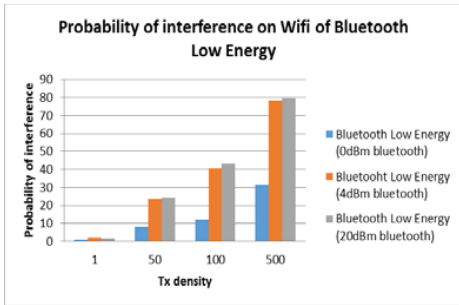


Figure 7. Probability of interference on Wi-Fi of Bluetooth Low Energy

It can be seen that with higher power probability of interference of Bluetooth Low Energy also increases. In first case when Bluetooth power is 4dBm, the Bluetooth Low Energy probability of interference on Wi-Fi increases for 93 % when Tx density equals to one. When TX density equals to 50, probability of interference increases for 192%. For Tx density of 100, probability of interference grows for 234.16%, and for Tx density of 500 probability of interference is bigger for 150.6%.

In Figure 9, probability of interference on Wi-Fi of Bluetooth Low Energy when ZigBee power is changed.

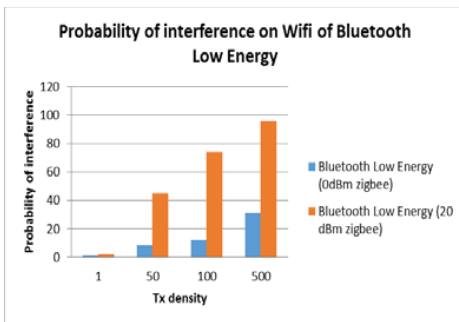


Figure 8. Probability of interference on Wi-Fi of ZigBee

From Figure 9, it can be also seen that with increasing the ZigBee power, probability of interference of Bluetooth Low Energy on Wi-Fi increases. When Tx density equals to one, probability of interference 93%. When TX density equals to 50, probability of interference increases for 469%. For Tx density of 100, probability of interference grows for 508.1%, and for Tx density of 500 probability of interference is bigger for 207%.

V. CONCLUSION

Smart home environments become more important in everyday life and will be part of everyone’s household in near future. These environments are deployed using different devices which communicate using different wireless technologies. In this paper challenges of Bluetooth, Bluetooth Low Energy, Wi-Fi and ZigBee technologies coexistence in 2.4 GHz band were presented. Results showing probability of interference in dependence on transmitter density over km<sup>2</sup>, simulation radius, transmitter power and spatial distribution of interferers are given. Based on simulation results it can be seen that by increasing the number of devices, the probability of interference is higher. Probability of interference goes from 1.232% when Tx density equals to one to 94,183% when Tx density equals to 10000 for ZigBee. Probability of interference goes from 1.112% when Tx density equals to one to 97,442% when Tx density equals to 10000 for Bluetooth. Probability of interference goes from 1.129% when Tx density equals to one to 88,716% when Tx density equals to 10000 for Bluetooth Low Energy. When probability of interference in dependence of simulation radius was investigated it is demonstrated that probability of interference decreases with greater simulation radius. Also, different impacts of transmitter power on probability of interference of every considered technology are given. In most cases higher transmitter power causes higher interference.

VI. REFERENCES

- [1] Seamcat software tool, <http://www.seamcat.org/>
- [2] The eWALL Consortium, “D2.1, Preliminary User and System Requirements,” eWALL for Active Long Living FP7 project, March 2015.
- [3] The eWALL Consortium, “eWALL\_Y2 review WP3”, eWALL for Active Long Living FP7 project, November 2015.
- [4] Bluetooth pages, <http://www.bluetooth.com/Pages/what-is-bluetooth-technology.aspx>
- [5] Specification of the Bluetooth System, Covered Core Package Version: 4.0, Bluetooth SIG, July 2010.
- [6] ZigBee Alliance, “ZigBee Specifications”, January 2007. <http://www.zigbee.org/>
- [7] IEEE 802.11 Wi-Fi Standards, <http://www.radio-electronics.com/info/wireless/wi-fi/ieee-802-11-standards-tutorial.php>
- [8] I. Sakal, D. Simunic, “Simulation of Interference between Bluetooth and 802.11b Systems”, EMC 2003, Vol. 2, pp. 1321 - 1324, May 2003.
- [9] Seung-Joo Yee, Sang-Hyun Chun, Jong-Heon Kim, Chan-Joo Lee “Interference Analysis of 2.4 GHz RF Transceivers Using ZigBee Standard in Multiple Integrated Laser Engagement System (MILES)” IEEE Microwave Asia-Pacific Conference, pp. 195-198, Dec 2007
- [10] On-line SEAMCAT manual <http://tractool.seamcat.org/wikiManual>.
- [11] SEAMCAT Handbook, <http://www.cept.org/eco/ecc-tools-and-services/seamcat>.
- [12] ECC Report 201, Compatibility study on MBANS operating in the 2400-2483.5 MHz and 2483.5-2500 MHz
- [13] ECC Report 175, Co-existence study considering UWB applications inside aircraft and existing radio services in 3.1-4.8 GHz/6.0-8.5 GHz
- [14] ECC Report 107, Regulating Interoperability
- [15] ERC Report 109, Compatibility of Bluetooth with other existing and proposed radiocommunication systems in the 2.45 GHz frequency band
- [16] ERC Report 083, Interference from industrial, scientific and medical (ISM) machines

# Cross-correlation index and multiple-access performance of Gold codes

<sup>1</sup>Olanrewaju B. Wojuola, <sup>2</sup>Stanley H. Mneney and <sup>1</sup>Viranjay M. Srivastava

<sup>1</sup>School of Engineering,  
University of KwaZulu-Natal, Durban, South Africa.

<sup>2</sup>Department of Electronics and Communication Engineering  
St. Joseph University in Tanzania, Dar es Salaam, Tanzania  
wojuolao@ukzn.ac.za, shmneney@gmail.com and viranjay@ieee.org

**Abstract**—Cross-correlation index is a measure of multi-user interference combating a spread-spectrum signal. Its magnitude, relative to auto-correlation index of a reference user, is a critical parameter affecting system performance. If cross-correlation index is much less than auto-correlation index, then there is little interference, and the desired signal survives. On the other hand, if cross-correlation index is much greater than auto-correlation index, then there is significant interference, and the desired signal becomes swamped. In this work, we develop a mathematical model for cross-correlation index for a set of Gold codes, and thereafter deduce the system capacity and bit-error-rate performance. Results show that the codes only have the capacity to support a maximum of about four users, above which bit-error-rate increases rapidly, ultimately resulting in emergence of error floor. The point at which the cross-correlation index equals to auto-correlation index marks the turning point around which the system performance revolves. These outcomes were tested and found to be in close agreement with results of simulations that were obtained for the system performance.

**Keywords**— spread-spectrum system, cross-correlation index, multi-user interference, Gold codes, bit-error-rate, error floor.

## I. INTRODUCTION

Spreading codes (or sequences) play a critical role in determining the performance of spread-spectrum systems. Imperfect code properties yield poor system performance. Code properties are known to be a cause of multiple-access interference, an important limiting factor on the system performance. The significance of spreading codes has made the investigation of code properties, code design [1-4], code performance [5-9], and the search for better spreading codes [10-12] ongoing research activities.

Gold codes are a type of binary sequences used in wireless telephony and satellite navigation. The codes, whose history dates back to the sixties [13-15], have become important elements in spread-spectrum systems and other modern applications. The codes are, for example, used as spreading sequences for multi-user separation in code-division multiple-access (CDMA) systems, as access codes in the global positioning system (GPS), in tomographic imaging [16], in digital terrestrial television [17], in anti-jam underwater communication [18] and in encryption of medical images [19]. Spread-spectrum techniques, whose advent dates back to the 1940s, were developed originally for military applications, but

have over the years found their way into the larger society for commercial and civilian applications [20-22].

Cross-correlation properties of spreading sequences play significant role in determining the system performance. Gold codes are known to have bounded small cross-correlations within a set, making them useful as access codes for multiple users transmitting over the same frequency band.

The cross-correlation between any pair of members a set of Gold codes are known to be triple-valued, but this knowledge is inadequate for determining the multiple-access performance of the codes. The same fact applies to the well-known value of peak cross-correlation between the code pairs. Knowledge of entire cross-correlation spectrum of the codes is necessary to determine their multiple-access performance. Unfortunately, generating the entire cross-correlation spectrum is difficult and intractable. This work provides an effective alternative to this problem.

In this paper, we introduce the concept of cross-correlation index for the estimation of multiple-access performance of the codes. The cross-correlation index provides a good estimation of number of users that a code can support, as well as a measure of critical points around which the system bit-error-rate (BER) performance revolves. Starting from basic principles, we develop a mathematical model for the codes' cross-correlation index. This is then used as a tool for the prediction of the system performance. The outcome of the analysis is found to agree closely with results of software simulations carried out on the system performance.

This paper is part of a larger study. Basic principles and theories behind the work are outcome of a PhD work [23]. This work is a follow-up to a paper presented in a previous edition of the Global Wireless Summit [24]. The rest of this paper is organized as follows. Mathematical model for the system is developed in Section II. Research methods used for the work reported in this paper are outlined in Section III. Results are presented in Section IV. Finally, this paper concludes with a summary of important deductions in Section V.

## II. SYSTEM MODEL

Consider a direct-sequence spread-spectrum (DSSS) system. Spread spectrum signal transmitted by a user  $k$  can be expressed as

$$s_k(t) = A c_k(t) b_k(t) \cos(\omega_c t + \theta_k), \quad (1)$$

Where  $b_k(t)$  is the user binary data,  $c_k(t)$  is the spreading code and  $\omega_c$  is the carrier frequency. The spreading code  $c_k(t)$  for the user can be denoted as

$$c_k(t) = \sum_{i=1}^N c_k^i P_c(t - iT_c), \quad c_k^i \in \{-1, +1\}, \quad (2)$$

where  $N$  is length of the code, and  $P_c$  is a rectangular pulse having a duration  $T_c$ . Let the wireless communication channel be represented by multiple paths having a real positive gain  $\beta_l$ , propagation delay  $\tau_l$  and phase shift  $\gamma_l$ , where  $l$  is the path index. The channel impulse response  $h_k(t)$  for  $L$  independent paths can be modelled as

$$h_k(t) = \sum_{l=1}^L \beta_{kl} e^{j\gamma_{kl}} \delta(t - \tau_{kl}) \quad (3)$$

\*\*\*\*\*

$$r_k(t) = \sum_{l=1}^L A \beta_{kl} e^{j\gamma_{kl}} c_k(t - \tau_{kl}) b_k(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) \quad (5)$$

$$r(t) = \sum_{k=1}^K \sum_{l=1}^L A \beta_{kl} e^{j\gamma_{kl}} c_k(t - \tau_{kl}) b_k(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) \quad (6)$$

$$z_1(m) = \int_{mT_b}^{(m+1)T_b} (r(t) + n(t)) c_1(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) dt \quad (7)$$

$$z_1(m) = \int_{mT_b}^{(m+1)T_b} \left\{ \sum_{k=1}^K \sum_{l=1}^L A \beta_{kl} e^{j\gamma_{kl}} c_k(t - \tau_{kl}) b_k(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) \right\} c_1(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) dt \\ + \int_{mT_b}^{(m+1)T_b} n(t) c_1(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) dt \quad (8)$$

$$z_1(m) = \int_{mT_b}^{(m+1)T_b} \left\{ \sum_{l=1}^L A \beta_{1l} e^{j\gamma_{1l}} c_1(t - \tau_{1l}) b_1(t - \tau_{1l}) \cos(\omega_c t - \theta_{1l}) \right\} c_1(t - \tau_{1l}) \cos(\omega_c t - \theta_{1l}) dt \\ + \int_{mT_b}^{(m+1)T_b} \left\{ \sum_{k=2}^K \sum_{l=1}^L A \beta_{kl} e^{j\gamma_{kl}} c_k(t - \tau_{kl}) b_k(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) \right\} c_1(t - \tau_{1l}) \cos(\omega_c t - \theta_{kl}) dt \\ + \int_{mT_b}^{(m+1)T_b} n(t) c_1(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) dt \\ = z_{11} + z_{12} + z_{13} \quad (9)$$

\*\*\*\*\*

Consider the mathematical expression for  $z_{1l}$  of (9). In this expression, signal recovery at the receiver involves matrix multiplication of code sequences, such that the term  $c_1^2(t - \tau_{1l})$  represents the inner product of the first user-assigned code  $c_1(t)$  with itself. Now, the user's code is a sequence of chip

At the receiving end, the received signal  $r_k(t)$  for the user is obtained by convolving  $s_k(t)$  with  $h_k(t)$ :

$$r_k(t) = \int_{-\infty}^{\infty} s_k(\tau) h_k(t - \tau) d\tau \quad (4)$$

Substituting the expressions for  $s_k(t)$  and  $h_k(t)$  into this integral, and using relevant properties of the Dirac delta function  $\delta(t)$  gives (5). For a multi-user system comprising  $K$  users, the received signal  $r(t)$  is a linear superposition of the signals for the users, and is given by (6).

Let user-1 be the reference user. Assuming coherent demodulation, the receiver output  $z(m)$  for  $m^{\text{th}}$  bit during the bit duration  $T_b$  of the user is given by (7), where  $n(t)$  is receiver noise. Substituting for  $r(t)$  gives (8). Define  $k = 1$  for the reference user. Using this in (8) gives (9), where  $z_{1l}$  represents the desired signal for the reference user,  $z_{12}$  is interference term, and  $z_{13}$  is noise term.

elements (+1,-1). Using this, we can show that  $c_1^2(t - \tau_{1l})$  is equal to the length  $N$  of the code. Also, let  $\omega_c$  be chosen such that  $f_c = \frac{n}{T_b}$ ,  $n \in \mathbf{Z}$ . By substituting these, we can show that the expression for  $z_{1l}$  reduces to:

$$z_{11} = \frac{AN}{2} \sum_{l=1}^L \beta_{1l} e^{j\gamma_{1l}} \left\{ \int_{mT_b}^{(m+1)T_b} b_1(t - \tau_{1l}) dt \right\} \quad (10)$$

If  $b_1(t - \tau_{1l}) = 1$  in the period  $T_b$ , then

$$z_{11} = \frac{AN}{2} \sum_{l=1}^L \beta_{1l} e^{j\gamma_{1l}} T_b \quad (11)$$

For a Gaussian channel, number of paths  $L = 1$ . Consequently, there is no multipath fading. Therefore,  $\beta_{1l} = 1$  and  $\gamma_{1l} = 0$ . Hence, we have

$$z_{11} = \frac{ANT_b}{2}. \quad (12)$$

Similarly, we can show that the interference term  $z_{l2}$  reduces to:

$$z_{12} = \pm \frac{A}{2} \sum_{k=2}^K \int_{mT_b}^{(m+1)T_b} d_{1k} dt, \quad (13)$$

where  $d_{1k} = c_1(t - \tau_{1l})c_k(t - \tau_{kl})$ . Consider the noise term  $z_{13}$ :

$$z_{13} = \int_{mT_b}^{(m+1)T_b} n(t)c_1(t - \tau_{kl}) \cos(\omega_c t - \theta_{kl}) dt \quad (14)$$

In this equation, it can be seen that the spreading code  $c_1(t - \tau_{kl})$  multiplies the receiver noise  $n(t)$ . Effect of this is to spread out the noise. Hence the noise is suppressed greatly. Bit-error-rate (or bit-error-probability)  $P_e$  can now be calculated using:

$$\begin{aligned} P_e &= Q\left(\sqrt{2 \frac{E_b}{N_o + I_o}}\right) = Q\left(\sqrt{2 \frac{|z_{11}|^2}{N_o + |z_{12}|^2}}\right) \\ &= Q\left(\sqrt{2 \frac{\left|\frac{ANT_b}{2}\right|^2}{N_o + \left|\frac{A}{2} \sum_{k=2}^K \int_{mT_b}^{(m+1)T_b} d_{1k} dt\right|^2}}\right) \\ &= Q\left(\sqrt{2 \frac{\left|\frac{ANT_b}{2}\right|^2}{N_o + \left|\frac{A}{2} \sum_{k=2}^K \int_{mT_b}^{(m+1)T_b} d_{1k} dt\right|^2}}\right) \quad (15) \end{aligned}$$

where  $I_o$  is interference term. This expression shows that BER depends on code length  $N$ . Since the  $Q$ -function is a monotonically decreasing function, it implies that as  $N$  increases, BER decreases. That is to say, longer codes give better BER. The equation also shows that BER depends on the cross-correlation  $d_{1k}$ . This implies that BER for a multi-user system is affected by cross-correlation index of the codes.

Gold codes are a type of pseudo-noise (PN) sequences derived from combination of certain pairs of  $m$ -sequences called *preferred sequences*, implemented using linear feedback shift registers. A set of Gold code sequences consists of  $2^n - 1$

sequences, each one with a period of  $2^n - 1$ . A Gold code has a period  $N = 2^n - 1$ , where  $n$  is the length of the shift register. Gold codes exhibit triple-valued cross-correlation function [13, 25, 26] with values  $\{-1, -(n), (n)-2\}$ , where

$$t(n) = \begin{cases} 2^{(n+1)/2} + 1, & n \text{ odd} \\ 2^{(n+2)/2} + 1, & n \text{ even} \end{cases} \quad (16)$$

### III. METHODOLOGY

Graphs of cross-correlation index for the codes were generated direct software implementation of the mathematical model that was developed in the previous section. The model was tested using a set of 63-chip Gold codes. The outcome of this was compared with results of simulations.

For the software simulations, various sets of Gold codes were generated from appropriate combinations of preferred pairs of  $m$ -sequences, using linear feedback shift registers (Table 1). The software simulations were carried out for the transmission of typically about one million random QPSK symbols for uncoded as well as coded data transmission, through a Gaussian channel of zero mean and unit variance Gaussian noise. Following signal recovery, original transmitted data was compared with recovered data for the determination of system (BER) for the various sets of codes. The simulations for the various code lengths were carried out simultaneously during the same simulation session, thereby ensuring consistency of simulation conditions.

TABLE I. GENERATOR POLYNOMIALS FOR THE GOLD CODES

$n$	$P_1^n(x)$	*Generator polynomial	$N$
6	$P_2^6(x)$	$x^6 + x^5 + 1$	63
	$P_2^6(x)$	$x^6 + x^5 + x^4 + x + 1$	
8	$P_2^8(x)$	$x^8 + x^7 + x^6 + x + 1$	255
	$P_2^8(x)$	$x^8 + x^7 + x^5 + x^3 + 1$	
10	$P_2^{10}(x)$	$x^{10} + x^7 + 1$	1023
	$P_2^{10}(x)$	$x^{10} + x^9 + x^8 + x^5 + 1$	
12	$P_2^{12}(x)$	$x^{12} + x^{11} + x^{10} + x^4 + 1$	4095
	$P_2^{12}(x)$	$x^{12} + x^{11} + x^{10} + x^2 + 1$	

\* $P_1^n(x)$  and  $P_2^n(x)$  are the generator polynomials of the preferred pair used for obtaining corresponding set of Gold codes of degree  $n$ .

### IV. RESULTS

Cross-correlation index for a set of spreading code is an important parameter affecting the system performance.

Figure 1 shows the graph of cross-correlation index  $D_{lk}$  for a set of 63-chip Gold codes ( $N = 63$ ). This graph was generated by direct implementation of (15) at zero lag. For the purpose of comparison, the value of auto-correlation index  $D_{ll}$  was also obtained and plotted. Both quantities were normalised by code length. This figure shows that  $D_{lk}$  increases linearly with number of users, reaching a peak value of 15 at maximum loading (63 users) for the code. Figure 1 also shows that  $D_{lk} = D_{ll}$  for four users. Therefore, the system is expected to saturate when number of users approaches four. These predictions agree with results obtained for the system performance. This is explained as follows.

REFERENCES

- [1] G. Song, Y. Tsujii, J. Cheng, and Y. Watanabe, "Finite Field Spreading for Multiple-Access Channel," *IEEE Transactions on Communications*, vol. 62, pp. 1001-1010, 2014.
- [2] S. Jos, J. P. Nair, D. Sen, and A. Naniyat, "Method of Generating Multiple Sets of Orthogonal Codes with Wide Choice of Spreading Factors," *IEEE Wireless Communications Letters*, vol. 1, pp. 492-495, 2012.
- [3] H. Chenggao, N. Suehiro, and T. Hashimoto, "A Systematic Framework for the Construction of Optimal Complete Complementary Codes," *IEEE Transactions on Information Theory*, vol. 57, pp. 6033-6042, 2011.
- [4] T. Yifeng, F. Pingzhi, H. Li, and T. Xiaohu, "A Simple Method for Generating Optimal Periodic Complementary Sequence Set Based on Phase Shift," *IEEE Signal Processing Letters*, vol. 17, pp. 891-893, 2010.
- [5] N. Rahnama and S. Talebi, "Performance comparison of chaotic spreading sequences generated by two different classes of chaotic systems in a chaos-based direct sequence code division multiple access system," *IET Communications*, vol. 7, pp. 1024-1031, 2013.
- [6] O. B. Wojuola and S. H. Mneney, "Performance of even- and odd-degree Gold codes in a multi-user spread-spectrum system," presented at the Global Wireless Summit 2014, Aalborg, Denmark, 2014.
- [7] O. B. Wojuola, "Performance of DS-CDMA and MC-CDMA systems in Gaussian and Rayleigh fading channels," (to be submitted).
- [8] O. B. Wojuola, S. H. Mneney, and V. Srivastava, "Performance of a space-time coded MC-CDMA system in a Rayleigh fading channel," in *5th IEEE International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, ed. Hyderabad, India: IEEE, 2015.
- [9] O. B. Wojuola and S. H. Mneney, "Multiple-access interference of Gold codes in a DS-CDMA system," *SAIEE African Research Journal*, vol. 106, pp. 4-10, 2015.
- [10] H. Su, L. Zilong, G. Yong Liang, X. Wenhui, B. Guoan, and L. Shaoqian, "Sequence Design for Cognitive CDMA Communications under Arbitrary Spectrum Hole Constraint," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1974-1986, 2014.
- [11] C. Kwonhue and L. Huaping, "Polyphase Scrambled Walsh Codes for Zero-Correlation Zone Extension in QS-CDMA," *IEEE Communications Letters*, vol. 16, pp. 429-431, 2012.
- [12] Z. Zhang, F. Tian, F. Zeng, S. He, and G. Xuan, "Multiple Orthogonal Subsets With Three-Valued In-Phase Cross-Correlation for HF Communications," *IEEE Communications Letters*, vol. 20, pp. 1377-1380, 2016.
- [13] R. Gold, "Characteristic Linear Sequences and Their Coset Functions," *SIAM Journal on Applied Mathematics*, vol. 14, pp. 980-985, 1966.
- [14] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 13, pp. 619-621, 1967.
- [15] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 14, pp. 154-156, 1968.
- [16] T. M. S. x, oeu, and M. R. Inggs, "Fully Parallel Electrical Impedance Tomography Using Code Division Multiplexing," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, pp. 556-566, 2016.
- [17] X. Feng, H. C. Wu, Y. Wu, and X. Wang, "Kasami sequence studies for DTV transmitter identification," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 1138-1146, 2012.
- [18] S. Kalita and P. P. Sahu, "An anti-jamming underwater communication transceiver model using uncoordinated direct sequence spread spectrum technique," in *Electronics and Communication Systems (ICECS), 2015 2nd International Conference on*, 2015, pp. 972-976.
- [19] A. B. Mahmood and R. D. Dony, "Adaptive encryption using pseudo-noise sequences for medical images," in *Third International Conference on Communications and Information Technology (ICCIT)*, 2013, pp. 39-43.
- [20] R. Prasad, *CDMA for Wireless Personal Communications* Artech House, 1996.
- [21] A. J. Viterbi, *CDMA: principles of spread spectrum communication*: Addison-Wesley Pub. Co., 1995.
- [22] A. J. Viterbi, "Spread spectrum communications: myths and realities," *IEEE Communications Magazine*, vol. 40, pp. 34-41, 2002.
- [23] O. B. Wojuola, "Performance of a space-time coded multicarrier CDMA system in frequency-selective Rayleigh channel" PhD, Electrical,

Consider a family of BER curves for a direct-sequence CDMA system for a family of Gold codes. If there is no system saturation, horizontal spacing between adjacent (neighbouring) performance curves for the various code lengths are expected to remain the same. Figure 2(a) and (b) illustrate this for all the code lengths under consideration. Now, if any of the codes undergoes system saturation, it will deviate from this normal behaviour, and will drift away from the other characteristic curves. On Figure 2(c), this anomalous behaviour can be observed for code  $N = 63$ . The performance curve for this code can be seen to drift away gradually from the others. This is an evidence of system saturation.

It is interesting to note that this figure (i.e. Figure 2(c)) represents the performance for four users. Thus we see that the anomalous behaviour agrees exactly with what the mathematical model predicts for this code. The model predicts that system saturation should set in after four users. This situation becomes more obvious when number of users increases to five and above. This can be seen on Figures 2(d) to (f). These figures show that when number of users become increasingly higher than 4, system performance degrades rapidly, ultimately resulting in emergence of error floor. Going back to Figure 1, we see that peak value of  $D_{ik}$  is 15. This implies that at that point, cross-correlation index is 15 times larger than autocorrelation index. This means that multi-user interference is 15 times larger than the desired signal. Therefore the signal becomes swamped, and level of error floor worsens.

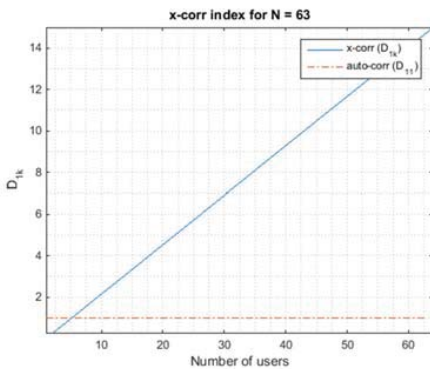


Figure 1. Cross-correlation index for  $N = 63$

V. CONCLUSION

This paper presented a mathematical model for cross-correlation index of Gold codes. The model was tested for a set of 63-chip Gold codes and was found to agree with results of simulations for the system performance. The results confirmed that the cross-correlation index is an accurate tool for predicting level of multiple-access interference experienced by a spread-spectrum signal. It is also useful for the estimation of loading capacity of a spreading code, as well as the onset of system saturation and emergence of error floor. This work is part of a larger study. Other results not presented here confirmed that the model is valid for other sets of Gold codes. Further results shall be presented in future communications.

Electronic and Computer Engineering, University of KwaZulu-Natal, Durban, South Africa, 2015.

- [24] O. B. Wojuola, S. H. Mneye, and V. Srivastava, "A review of bit-error-rate performance of interleavers in IDMA systems," in *2015 IEEE International Conference on Computing and Network Communications (CoCoNet'15)*, Trivandrum, India, Under Review (259), 25 - 30 Sept 2015.

- [25] R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Transactions on Information Theory*, vol. 13, pp. 619-621, 1967.
- [26] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions (Corresp.)," *IEEE Transactions on Information Theory*, vol. 14, pp. 154-156, 1968.

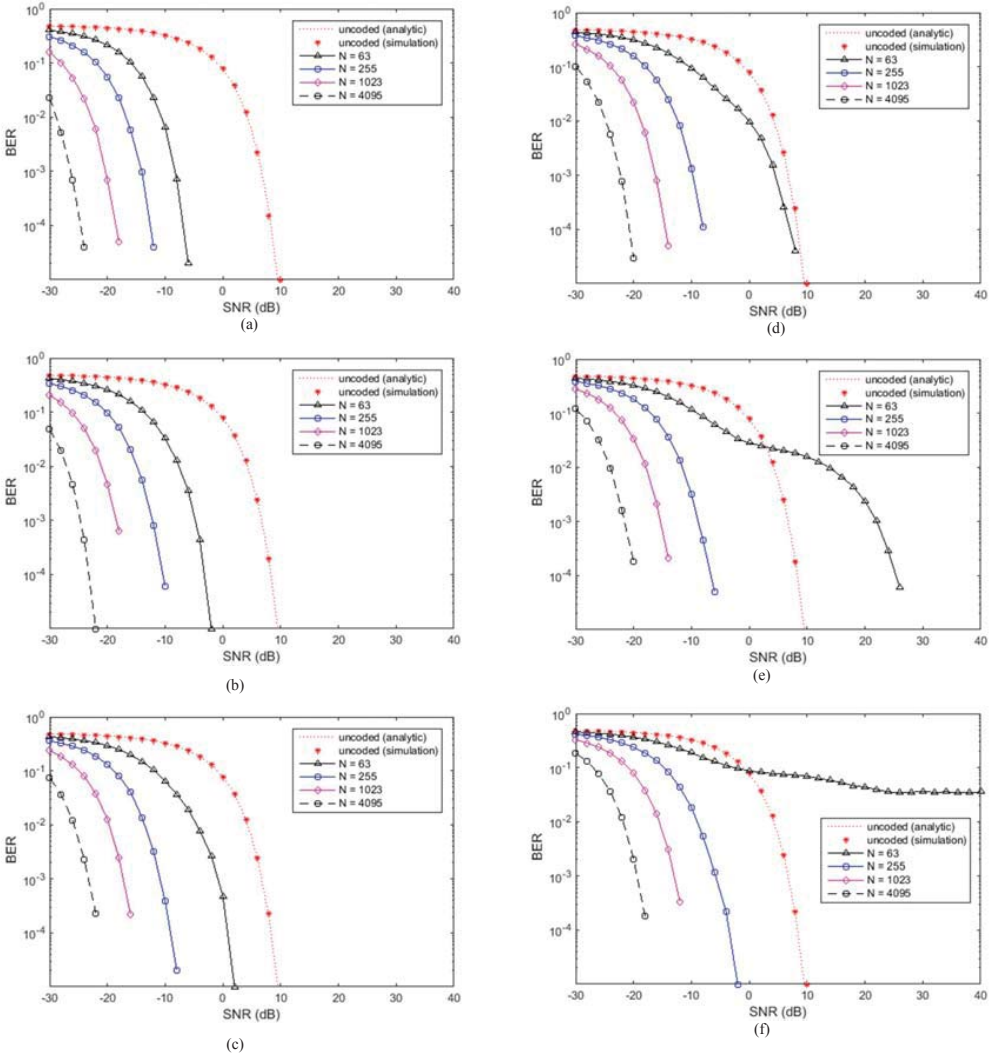


Figure 2. Performance for (a) two, (b) three, (c) four, (d) five, (e) six and (f) ten users.

# Leveraging TV White Spaces as a Tool for Improved Rural Broadband Connectivity in Developing Countries: An Operational Perspective

Alexander Osei-Owusu  
Department of Electronic Systems  
Aalborg University  
Copenhagen, Denmark  
[ao@es.aau.dk](mailto:ao@es.aau.dk)

Kwesi Fosu Biney-Assan  
Faculty of Engineering - Graduate School  
Ghana Technology University College  
Accra, Ghana  
[bineyassankwesi@gmail.com](mailto:bineyassankwesi@gmail.com)

**Abstract**—Around the globe there is a coordinated move for digital switch over (DSO). The DSO is to discontinue analogue television transmission. Ghana which is chosen as a case study is among the pioneers of the sub-Saharan African countries that signed the Geneva agreement on DSO. The DSO is expected to free some portions of the transmission spectrum and such portions of the spectrum left unused is called “tv white space”. This paper sought to determine whether “tv white space” can be used as a tool for improved rural broadband connectivity. In achieving the above objective, the paper focused on the operational perspective of the “tv white space” using the first ever established commercial TV white space broadband service at Koforidua, one of the regional capitals in Ghana. The research asked the question, does the current mode of tv white space provide adequate throughput? In so doing an observational experiment was conducted within a week/7days 24hr period over the network to record the signal strength and the throughput of the system. The research concluded that TV white space broadband provides better signal strength, very stable link with very minimal packet drops and consumes less power compared to a Wi-Fi over a 3G system.

**Keywords** □ □ □ hite □ ace; digital switch over; □ ural □ onnectivity; through ut; signal strength; wifi; □ systems

## I. INTRODUCTION

Internet connectivity provides access to knowledge and information, which is crucial for social and economic development. Internet broadband services in today’s era of information explosion, has been conceived as the most important resource that contributes greatly towards a nations development, (Saeed, 2011). Information provides the foundation for innovations, development of knowledge, and the resource for an informed citizenry which is an essential commodity for the progress of any society, (Pejovic, 2012).

Rural areas provide challenging environment for commercial broadband or network service providers’. As service providers face the challenge of huge infrastructure cost, most rural areas typically in sub-Saharan Africa are characterized by low income, highly scattered and low population density, (Simba, 2011). This situation drives

network operators to establish network infrastructures in urban centers leaving rural areas as underserved communities. Only government subsidies will make private investment viable, unless a more innovative approach is considered and the “tv white space” promises to be the option. Ghana is among few countries in the sub-Saharan region who is at the fore front of bridging the rural connectivity gap. The government through the telecom regulator, the National Communication Authority (NCA) has been involved in several projects aimed at spreading internet across most rural areas. Some of these projects include the Community Information Centre Initiative which was aimed at facilitating internet to 230 constituencies; and the wiring Ghana Initiative, aimed at constructing underground fiber-optic backbone across the country.

The switch over from analogue to digital television transmission taking place all over the world has led to a significant amount of unused spectrum. The unused band or spectrum from the digital switch over (DSO) is called “tv white space”. These TV white spaces present developing countries with some opportunities to improve rural internet coverage. Ghana is the first country in sub-Saharan Africa to establish a commercial broadband service using “tv white space”, as such presents as with a solid foundation and case for the assessment of this research.

Whilst tv white space is plausible from both regulatory and technical point of view, this research is focused on the operational and economic viability. The operational feasibility sought to measure how well the proposed TVWS can solve the problems as well as satisfy the requirement identified in the requirements analysis phase of the system development. In particular, the operational feasibility analysis will test the signal strength for transmitting/receiving, and the throughput over a period, as well as the transmission speed of the system.

This research has chosen a Wi-Fi broadband connection over a 3G system as the candidate system for caparison because most developing countries especially it remote communities are characterize with low human density, low income and as such traffic for data or voice calls are always

low. Operators have tended to maintain or set up a Wi-Fi connections over the old 3G systems in such areas so to avoid huge cost associated with latest systems while still serving their purpose in such communities.

## II. LITERATURE REVIEW

### A. Digital Switchover

According to the Digital Migration Report (2010), Ghana in 2006 signed the Geneva (GE06) Agreement, establishing the Digital Terrestrial Broadcasting Plan in the bands 174 – 230 MHz and 470 – 862 MHz at the Regional Radio Communications Conference (RRC-06). The Geneva Agreement set 17 June 2015 as the deadline for the cessation of international protection for analogue broadcasting transmissions in the said bands.

The migration from analogue to digital was necessary and urgent for the following reasons: to comply with and adopt the tenets of the GE-06 Agreement; to rapidly adopt spectrum efficient methods in the management of the scarce RF spectrum to broaden its utility as a resource in the interest and benefit of stakeholders; to prevent dumping of obsolete analogue transmission equipment into the country to protect the environment, investors and consumers; To enhance the quality and experience of TV viewers in Ghana by improving terrestrial TV transmission and reception; To promote environmental sanity through co-location of broadcast transmission infrastructure.

### B. TV White Space

[1] defined TV white space as the portion of the TV bands that is unused by licensed services. There are many unused TV channels in the VHF and UHF TV bands. Regulatory agencies, have been creating regulations to permit wireless networks to get access to these unused channels while guaranteeing that these wireless networks do not cause harmful interference to the licensed services in the TV bands. The TV White Space spectrum will provide a much better Radio Frequency propagation than systems that have been deployed in the Industrial, Scientific and Medical (ISM) bands allowing for a reliable, cost-effective, and better coverage in rural areas. This is mainly because the TV white space range, exists underneath 1GHz in both the VHF and UHF band. Cognitive radio technology has been proposed to compliment TV white space, this will ensure its devices do not cause interference to the services operating in the occupied TV channels. Some of the cognitive technologies include a combination of geo-location positioning and a database of incumbent systems and spectrum sensing. Geo Location deals with the ability of a device (TV White Space) to know its location in terms of latitude and longitude. The database stores information about all licensed services in the TV bands. Spectrum Sensing, is a technology embedded in the TV White Space device that makes measurements of the radio frequency (RF) TV channels to determine which channels are occupied by incumbent systems and which are unoccupied and are hence TV white Space,(Zennaro, 2013).

### C. Potentials of TV White Space

Zennaro (2013) conducted a study to assess the performance of the TV white spaces network under harsh African environment for example intermittent power and constrained Internet bandwidth resource. It was observed that unlike other fixed broadband services, TVWS services demonstrated 2.6 times better data rates given the same operating conditions. A similar study conducted by (Roberts et al, 2010) identified the benefits of using TV broadcast spectrum. They stated that TVWS signals can easily cover large regions as it transmits in lower frequencies. Comparatively, a typical Wi-Fi (i.e. over a 3G systems) signal can cover approximately 100 meters while a TVWS signal at the same power level can easily travel 400 meters and with higher power can cover several kilometers. Robert also stated from the study that TVWS signals can easily penetrate common obstructions, their research showed that a single TVWS base station operating at 20 dBm with a 2 dBi gain antenna was able to cover all parts of a building spanning a usable area of 222,000 sq. ft over 4 floors which had 150 access points to provide high speed access research. They concluded that TVWS signals travel farther than 3G broadband services, a TVWS deployment would cost less, and would consume less power than 3G and this makes it the best option for covering rural or less developed communities.

### D. Rural Connectivity

Simba (2011) conducted a research on the available connectivity technologies with potentials to offer broadband access network to rural areas. They stated that, rural areas especially those of the developing countries provide challenging environment to implement communication infrastructure for data and Internet based services and such challenges include high cost of network implementation and lack of customer base (as rural areas are characterized by low income, highly scattered and low population density).(Kenney,2000) examined the importance of reducing cost associated with internet connectivity and increasing the number of rural access points to increase rural internet access. He emphasized on how the internet has the potential to play a significant role in reducing poverty and promoting sustainable development, especially in rural areas. Kenny stated that active competition between providers will become important in reducing cost and also the introduction of low cost rural access options. (Pejovic, 2012), conducted a thorough examination of Internet usage at three locations in rural Africa consisting of network traffic analysis and Internet use. Interviews revealed several major obstacles to efficient broadband usage. These obstacles included restrictions on the locality of access, a lack of locally relevant content, unfamiliarity with new concepts, shortage of trained personnel, high cost of Internet access, and limited connection capacity with respect to the Internet structure and content.



E. Case of Ghana

In May 2014, Microsoft announced a commercial pilot in Ghana at the Dynamic Spectrum Alliance Global Summit held in Accra. Microsoft intends to work with public and private sector partners to deploy networks using low-cost wireless technologies, including TV White Space radios, to better enable the next generation of cloud-connected experiences. The networks used TV White Space-enabled radios and other wireless technologies to connect campus buildings as well as off-campus hostels where students live to ensure they have access to fast broadband. The pilot in Ghana was a commercial partnership with Spectra Wireless and a research partnership with Facebook. The pilot made use of TVWS under a test license from the National Communication Authority (NCA) to enable university and faculty members at universities in Koforidua to enjoy always-on and fast internet access. After the successful end to the trail in 2014, Africa’s first commercial service network utilizing TV White Spaces was launched on the 26th of January 2015 in Ghana. This service allow students to buy affordable, high speed internet bundles and devices.

III. TVWS NETWORK SETUP AT KOFORIDUA POLYTECHNIC

Figure 1 is a block diagram illustrating the connection from Accra to the Koforidua campus. Koforidua polytechnic is connected to the internet through a fiber connection from Accra to Obuotabri (Koforidua Mountains). From the mountain top, a point to point link was setup from the base station to the Central Classroom Block (CCB) block on campus. This link served as the main uplink to the Koforidua campus.

On the campus various point to point and point to multipoint installations have been setup to various blocks and hostels nearby over TV white spaces.

Figure 2 is a pictorial view of a point to multi point connection from the TV White space base to the Lincoln hostel and vineyard hostels.

Figure 1: block diagram of TWS network at Koforidua polytechnic

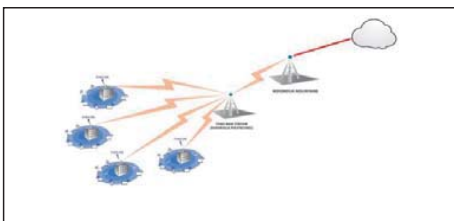


Figure 2: A pictorial view of a TVWS sample set-up at Koforidua Polytechnic

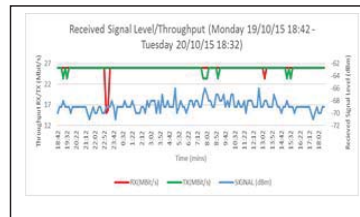


IV. OPERATIONAL FEASIBILITY ANALYSIS

A. Signal Strength Throughput Test fo TV White Spaces

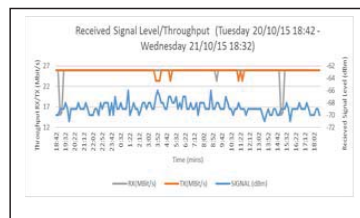
A test was carried out on the link from CCB block of Koforidua Polytechnic (TV White Space base station) and a nearby hostel called Vineyard hostel. Vineyard hostel is 496 meters from the Koforidua polytechnic. The test was conducted from 19/10/2015 at 18:42GMT to 26/10/15 at 18:32GMT.

Figure 3: RSL/Throughput test for Day1



Source: researcher’s field data (2015)

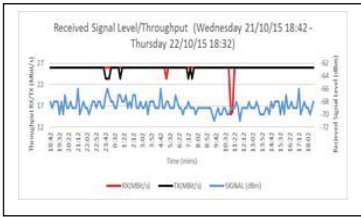
Figure 4: RSL/Throughput test for Day2



Source: researcher’s field data (2015)

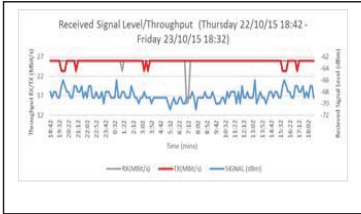
Figure 5: RSL/Throughput test for Day3

Source: researcher's field data (2015)



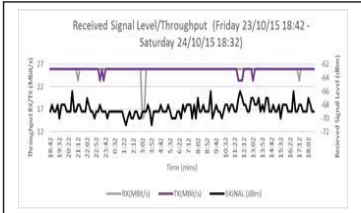
Source: researcher's field data (2015)

Figure 6: RSL/Throughput test for Day4



Source: researcher's field data (2015)

Figure 7: RSL/Throughput test for Day 5



Source: researcher's field data (2015)

Figure 8: RSL/Throughput test for Day6

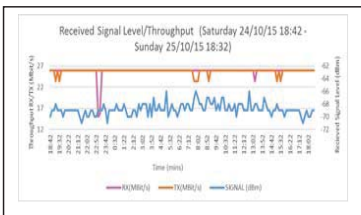
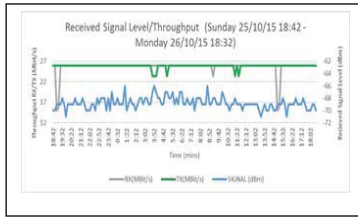


Figure 9: RSL/Throughput test for Day7



Source: researcher's field data (2015)

Table 1: Summary chart for (Rx MBit/s) and (Tx MBit/s) over 7days period

Period(dns)	Rx1	Tx1	Rx2	Tx2	Rx3	Tx3	Rx4	Tx4	Rx5	Tx5	Rx6	Tx6	Rx7	Tx7
18:42-19:32	26	23	15	26	26	26	26	26	26	26	26	23	15	26
19:32-20:22	26	26	26	26	26	26	26	23	26	26	26	26	26	26
20:22-21:12	26	26	26	26	26	26	26	23	23	26	26	26	26	26
21:12-22:02	26	26	26	26	26	26	26	26	26	26	26	26	26	26
22:02-22:52	26	26	26	26	26	26	26	26	26	26	26	26	26	26
22:52-23:42	15	26	26	26	26	23	26	26	26	23	15	26	26	26
23:42-0:32	26	26	26	26	26	23	26	26	26	26	26	26	26	26
0:32-1:22	26	26	26	26	26	23	23	26	26	26	26	26	26	26
1:22-2:12	26	26	26	26	26	26	26	26	26	26	26	26	26	26
2:12-3:02	26	26	26	26	26	26	26	23	15	26	26	26	26	26
3:02-3:52	26	26	26	23	26	26	26	23	26	26	26	26	26	23
3:52-4:42	26	26	26	26	26	26	26	26	26	26	26	26	26	26
4:42-5:32	26	26	26	23	23	23	26	26	26	26	26	26	26	23
5:32-6:22	26	26	26	26	26	26	26	26	26	26	26	26	26	26
6:22-7:12	26	26	26	26	26	26	15	26	26	26	26	26	26	26
7:12-8:02	26	23	26	26	26	26	26	26	26	26	26	26	23	26
8:02-8:52	26	26	26	26	26	26	26	26	26	26	26	26	26	23
8:52-9:42	26	23	23	26	26	26	26	26	26	26	26	26	23	26
9:42-10:32	26	26	26	26	26	26	26	26	26	26	26	26	26	26
10:32-11:22	26	26	26	23	26	15	26	26	26	26	26	26	26	23
11:22-12:12	26	26	26	26	26	26	26	26	26	26	26	23	26	26
12:12-13:02	26	26	26	26	26	26	26	26	26	26	26	26	26	26
13:02-13:52	23	26	26	26	26	26	26	26	26	26	26	26	26	26
13:52-14:42	26	26	26	26	26	26	26	26	26	26	26	26	26	26
14:42-15:32	26	26	15	26	26	26	26	26	26	26	26	23	15	26
15:32-16:22	26	26	26	26	26	26	26	23	26	26	26	26	26	26
16:22-17:12	26	26	26	26	26	26	26	23	26	23	26	26	26	26
17:12-18:02	26	26	26	26	26	26	26	26	26	26	26	26	26	26

Rx = throughput over 24 hourly period of a TVWS downlink channel. Tx = throughput recorded over 24 hourly period of a TVWS uplink channel. Occasional low throughput in red.

Table 2: Signal strength (S/dBm) over 7days period

Period(hrs)	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	S <sub>4</sub>	S <sub>5</sub>	S <sub>6</sub>	S <sub>7</sub>
18:42-19:32	-70	-66	-66	-66	-66	-70	-66
19:32-20:22	-70	-70	-66	-66	-66	-66	-66
20:22-21:12	-66	-66	-66	-66	-66	-66	-66
21:12-22:02	-71	-66	-66	-66	-66	-66	-66
22:02-22:52	-70	-68	-66	-66	-66	-66	-66
22:52-23:42	-70	-67	-66	-66	-66	-66	-66
23:42-0:32	-66	-66	-66	-66	-66	-66	-66
0:32-1:22	-70	-66	-66	-66	-66	-66	-66
1:22-2:12	-70	-66	-66	-66	-66	-66	-67
2:12-3:02	-66	-66	-66	-70	-66	-66	-66
3:02-3:52	-66	-66	-66	-66	-66	-66	-66
3:52-4:42	-66	-66	66	-66	-66	-66	-66
4:42-5:32	-66	-66	-67	-66	-70	-66	-66
5:32-6:22	-66	-66	-66	-66	-66	-66	-66
6:22-7:12	-67	-66	-66	-70	-66	-66	-66
7:12-8:02	-66	-70	-66	-71	-66	-66	-66
8:02-8:52	-66	-67	-66	-66	-66	-66	-66
8:52-9:42	-66	-66	-66	-66	-66	-67	-68
9:42-10:32	-67	-68	-66	-66	-70	-68	-66
10:32-11:22	-68	-68	-68	-66	-66	-66	-66
11:22-12:12	-68	-68	-71	-66	-68	-66	-66
12:12-13:02	-67	-67	-70	-66	-68	-66	-68
13:02-13:52	-66	-66	-65	-65	-66	-68	-67
13:52-14:42	-66	-66	-66	-66	-67	-68	-68
14:42-15:32	-66	-66	-66	-67	-66	-66	-66
15:32-16:22	-66	-66	-66	-66	-70	-66	-70
16:22-17:12	-68	-66	-66	-66	-66	-68	-68
17:12-18:02	-68	-66	-66	-66	-66	-67	-66

Results from Table 1 and Table 2 clearly indicates good received signal levels (RSL) under the period of observation. The highest RSL was -66dBm with -71dBm being the lowest.

An average of 26MBit/s uplink/downlink packet throughput was observed, which is the maximum expected throughput of the radio that can be achieved with few occasional low throughput of 15 MBit/s uplink/downlink packet.

*B. Throughput and a Signal Strength of a 3G Broadband*

For an ideal case scenario for a 3G broadband, throughput ranges from 14Mbit/s and 7Mbit/s for a downlink and an uplink respectively, even in an upgraded 3G such as 3.5G, it is between 42Mbit/s and 22Mbit/s for both downlink and uplink respectively.(Goldsmith, 2005)

*C. Ping Test*

A ping test was conducted to further test the network stability or robustness, jitter and latency of the connection and also the time it took for packets to make a round trip to the public DNS (Domain Name System) server. A ping test is a tool used to send small packets to a server or a computer to measure the amount of time it takes to get there. For the purpose of the study a ping test was conducted to reach Google’s public DNS server [8.8.8.8].From the results it took a minimum of 146ms, a maximum of 361 and an average of 185ms to make a roundtrip to the server. The round trip can be further explained as the time between when the packet was sent and when the packet was received.

A total amount of 36 packets were sent, 36 packets were received. This indicates no packets were lost and the link was stable. A further case for system quality and service quality. Below is a snapshot of the test conducted using the Command prompt; a total amount of 36 packets were sent, 36 packets were received. This indicates no packets were lost and the link was stable. A further case for system quality and service quality.

```

C:\Users\User>ping 8.8.8.8 -t

Pinging 8.8.8.8 with 32 bytes of data:
Reply from 8.8.8.8: bytes=32 time=188ms TTL=64
Reply from 8.8.8.8: bytes=32 time=188ms TTL=64
Reply from 8.8.8.8: bytes=32 time=247ms TTL=64
Reply from 8.8.8.8: bytes=32 time=154ms TTL=64
Reply from 8.8.8.8: bytes=32 time=179ms TTL=64
Reply from 8.8.8.8: bytes=32 time=184ms TTL=64
Reply from 8.8.8.8: bytes=32 time=170ms TTL=64
Reply from 8.8.8.8: bytes=32 time=157ms TTL=64
Reply from 8.8.8.8: bytes=32 time=175ms TTL=64
Reply from 8.8.8.8: bytes=32 time=184ms TTL=64
Reply from 8.8.8.8: bytes=32 time=179ms TTL=64
Reply from 8.8.8.8: bytes=32 time=184ms TTL=64
Reply from 8.8.8.8: bytes=32 time=169ms TTL=64
Reply from 8.8.8.8: bytes=32 time=153ms TTL=64
Reply from 8.8.8.8: bytes=32 time=146ms TTL=64
Reply from 8.8.8.8: bytes=32 time=175ms TTL=64
Reply from 8.8.8.8: bytes=32 time=184ms TTL=64
Reply from 8.8.8.8: bytes=32 time=167ms TTL=64
Reply from 8.8.8.8: bytes=32 time=188ms TTL=64
Reply from 8.8.8.8: bytes=32 time=173ms TTL=64
Reply from 8.8.8.8: bytes=32 time=184ms TTL=64
Reply from 8.8.8.8: bytes=32 time=181ms TTL=64
Reply from 8.8.8.8: bytes=32 time=168ms TTL=64
Reply from 8.8.8.8: bytes=32 time=206ms TTL=64
Reply from 8.8.8.8: bytes=32 time=178ms TTL=64
Reply from 8.8.8.8: bytes=32 time=155ms TTL=64
Reply from 8.8.8.8: bytes=32 time=188ms TTL=64
Reply from 8.8.8.8: bytes=32 time=261ms TTL=64

Ping statistics for 8.8.8.8:
    Packets: Sent = 36, Received = 36, Lost = 0 (0% loss),
    approximate round trip times in milliseconds:
        Minimum = 146ms, Maximum = 361ms, Average = 185ms
C:\Users\User>
    
```

Source: researcher’s field test data (2015)

Remarks:

Throughput for both downlink and uplink of a TVWS by far is better than that of a WIFI 3G broadband network. A good signal strength for any radio transmission ranges between -55dbm and -67dbm (Goldsmith, 2005)

TVWS exhibited very good signal strength with just a fraction of the period, falling little below the standard range of a good signal strength.

CONCLUSIONS

The research concluded that TV white space broadband provides a better signal strength, and throughput and would cost less for deployment which makes it the best option for an improved rural broadband connectivity.

REFERENCES

Rashid A. Saeed, S.J.S., 2011. TV White Space Spectrum Technologies: Regulations, Standards, and Applications, Pejovic, V. et al., 2012. The Bandwidth Divide : Obstacles to Efficient Broadband Adoption in Rural Sub-Saharan Africa. International Journal of Communication, 6(2012), pp.2467–2491

Simba, F. et al., 2011. Broadband Access Technologies for Rural Connectivity in Developing Countries. International Journal of Research and Reviews in Computer Science (IJRRCS), 2(2), pp.312–319.

Zennaro, Marco, Pietrosemoli, E, E.A., 2013. An Assessment Study on White Spaces in Malawi Using Affordable Tools. IEEE 2013 Global Humanitarian Technology Conference

Roberts, S., Garnett, P. & Chandra, R., Connecting Africa Using the TV White Spaces : From Research to Real World Deployments.

Kenny, C.J., 2000. Expanding Internet access to the rural poor in Africa\*. Information Technology for Development, 9(1), pp.25–31

Andrea Goldsmith (2005). Wireless Communications. Cambridge University Press cited.

# Performance of Non-binary Tail-biting LDPC Convolutional Code Encoder Design for Image Transmission

Tanaporn Payommai

Department of Electronic and Telecommunication  
Engineering, Faculty of Engineering  
King Mongkut's University of Technology Thonburi  
126 Pracha-uthit Rd., Bangmod, Thungkhru, Bangkok  
10140, Thailand  
E-mail: Tanaporn@vru.ac.th

Kosin Chamnongthai

Department of Electronic and Telecommunication  
Engineering, Faculty of Engineering  
King Mongkut's University of Technology Thonburi  
126 Pracha-uthit Rd., Bangmod, Thungkhru, Bangkok  
10140, Thailand  
E-mail: Kosin.cha@kmutt.ac.th

**Abstract**—In image transmission, an encoder needs to code image data in terms of a low bit error rate (BER), low-rate loss and robustness against noise. From an information theory perspective, image compression is highly sensitive to bit errors arising from interference. We focus on the problem of high BER when the large scale of data is transmitted over additive white Gaussian noise (AWGN) channel. The proposed scheme employs a non-binary tail-biting low-density parity-check convolutional (NB TB-LDPC) code as channel coding to deal effectively with bit error. NB TB-LDPC is a form of error control coding (ECC), an important tool for improving data transmission. The advantage of tail-biting is that it avoids rate loss. In addition, the design of a base matrix with column weight two shows that the encoding performance depends on the column weight and the length of the code over a finite file. The performance of the proposed method is evaluated by comparing the BER as the code length increases. The simulation results for the method's performance for correcting data error at a low signal-to-noise ratio (SNR) show that the design of NB TB-LDPC code with a base matrix column weight two can improve the performance by at least 0.5 dB compared with TB-LDPC and can adequately maintain the quality of the received image and high peak signal-to-noise ratio (PSNR) at low SNR.

**Keywords**—non-binary tail-biting low-density parity-check convolutional (NB TB-LDPC) code; image transmission.

## I.

## INTRODUCTION

The demands of image transmission over wireless channels are recently increasing, and efficient wireless image transmission is crucial for high quality communication. Nevertheless, a wireless channel is usually sensitive to time-varying and multi-path fading, which causes the transmitted images to exhibit a high bit error rate (BER). Loss of either the completeness or quality of the images can occur at a receiver if the images were compressed. The BER is influenced by receiver imperfection, and a high bit error makes the reconstructed image unacceptable. In future communication systems, a lower BER will be required to provide high-quality multimedia service; hence, more powerful channel coding is needed. In the past few years, research on channel coding has

attracted considerable interest for the purpose of improving performance at a low cost. The introduction of redundant bits into the input data can help remove errors during transmission over the channel.

A low-density parity-check (LDPC) code [1] is a type of channel coding, and is an important tool for improving data transmission. The main advantage of the code is to improve the performance close to a limited capacity for many different channels by employing linear time complex algorithms. LDPC is used in many applications such as mobile communication, DVB2 (Digital Video Broadcast2), magnetic recorder systems and optical storage system.

LDPC code is one of many types of linear block codes with a sparse matrix and is a subclass of prominently iterative decoder. Gallager [1] first introduced LDPC code in 1963 and for communication systems, LDPC code exhibits excellent performance near the channel limit via an additive white Gaussian noise (AWGN) channel. Mackay and Neal [2] proposed and explored [3] LDPC code as valuable for high rates ( $R > 2/3$ ) and short block lengths ( $N < 5000$ ) and they demonstrated better performance than the Reed-Solomon code. Ryan [4] and Yongqing [5] proved that high-rate LDPC codes exhibit excellent performance. However, the structure of LDPC code has high complexity employing quadratic encoding with the block lengths. The structure of LDPC code can be used for necessary reductions in encoding problems. In previous works, many researchers have considered the complexity of encoding problems [6-9]. [10-12] the paper proposes a design of sub-block encoder for high rate QC-LDPC for image transmission. It can keep a transmission quality, and improve its performance by measuring BER. The LDPC code has the disadvantage of being a high complexity encoder. In practical communication systems, low BER is a necessary but not a sufficient requirement. Additionally, the latency is considered to be caused by introducing channel coding, which is necessary. Convolution code has been considered as a good choice for applications with strong latency constraints [13]. LDPC convolutional codes (LDPCCCs) were introduced in [14]. LDPCCCs have some advantages in comparison with LDPC

block codes, particularly for transmitting data [15]. An important feature of LDPC codes is that the same encoder can be used to obtain code sequences of varying lengths with quite good performance by choosing different termination lengths. However, the introduction of a zero-tail for the termination results in the so-called rate loss. The tail-biting LDPC code with column weight 3 and row weight 6 ((3, 6) TB-LDPC codes) for hardware design [16] is proposed to avoid rate loss and outperforms LDPC codes in terms of performance and complexity. However, the column weight is not minimized, and the proposed (3, 6) TB-LDPC code is binary so that the performance is not as high. Meanwhile, a finite field with column weight two is proved [16] to offer the best LDPC performance. The main problems of the binary LDPC method are the complex air gap between the grain and the distribution of the moisture content that is completely different. These problems will lead to the error for the measurement. Recently, the multi-media resolution is dramatically increased as the current trend. The more resolution increased, the more serious the errors are. To quantize time-discrete and amplitude-continuous data into digital one which is initially convenient for hardware design may not be always suitable, especially in very high resolution so that the authors select to implement a method on non-binary basis in order to preserve high-resolution information as possible. Therefore, the authors focus on the development of non-binary TB-LDPC code for image transmission with column weight two to attain low BER. In developing the image transmission method, tail biting matrix and base matrix are created to provide a codeword for image transmission. In creating a tail biting matrix consisting of base matrices with code rate 0.5, we try to avoid low girth which causes low performance so that the identity matrices are assigned as the base matrix in the first row of the tail biting matrix, and the base matrices in the second row are one-bit-shifted identity matrices under the condition of non-low girth. In the case of low girth, it is necessary to shift by one bit at a time until the girth is greater than or equal to 8.

The rest of this paper is organized as follows: The image transmission system is introduced as an overview in section I. Section II is problem analysis. Proposed Non-binary LDPC Convolution Code Method is presented in section III. The experimental results are presented in section IV. Section V contains a discussion. Finally, the research works and results are concluded in section VI.

II. PROBLEM ANALYSIS

The conventional method of the tail-biting LDPC code with column weight 3 and row weight 6((3, 6) TB-LDPC codes) is an excellent theory, especially in the views of communication, hardware design and so on. However, in high-resolution image and video transmission, quantization errors occurring during binarization and bit errors should be seriously considered, since those errors may lead to misunderstanding of information in term of picture and video. The authors therefore try to consider modification for targeting better bit error rate and peak-signal-to-noise ratio.

Recently, the multi-media resolution is dramatically increased as the current trend. The more resolution increased, the more serious the errors are. To quantize time-discrete and

amplitude-continuous data into digital one which is initially convenient for hardware design may not be always suitable, especially in very high resolution so that the authors select to implement a method on non-binary basis in order to preserve high-resolution information as possible.

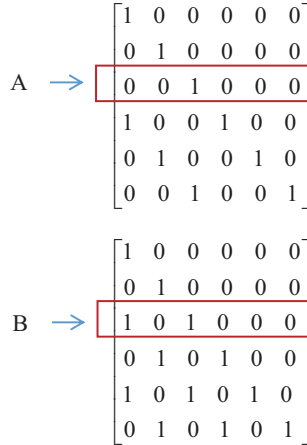


Fig. 1. Comparison between A (3, 6) and B (2, 6) column

In tail biting, the viewpoint of column weight of (3, 6) TB-LDPC code [13], 3x3 matrix is theoretically appropriate for hardware design with low computational cost. Basically, the more number of “1”, the less the BER in communication is when column weight of (3, 6) TB-LDPC code is analyzed comparing with (2, 6) TB-LDPC as shown by samples in A and B respectively in Fig. 2, number of “1” representing accuracy indicates by A and B in Fig. 1 shows that 2x2 matrix takes more advantage than 3x3 matrix in term of accuracy. However, computational cost of 2x2 matrix case would be increased which is considered as trade-off. This paper therefore modifies the TB-LDPC code to (2, 6) column weight.

III. PROPOSED NON-BINARY LDPC CONVOLUTION CODE METHOD

The concept of non-binary tail-biting LDPC convolutional code and the proposed method for designing a base matrix with column weight two are introduced in this section. We alleviate the aforementioned problems by considering non-binary tail-biting LDPC convolutional code defined over GF (28) with a parity matrix of column weight two. This paper seeks to reduce the bit error rate for image transmission to improve performance at low SNRs. The image transmission system recommended in this paper is shown in Fig. 2, and the proposed code encoder method is allocated in the 2nd process of the system, surrounded by a solid line. First of all the source color image is encoded by JPEG compression because it is currently a popularly used compression algorithm for images. The compressed image is then encoded by designed non-binary

tail-biting LDPC convolutional codes for the proposed method. The data are subsequently modulated using a binary phase shift keying (BPSK) scheme and transmitted over an additive white Gaussian noise (AWGN) channel. The image data received via the AWGN channel on the receiver side would be performed in reverse to reconstruct the original color image.

processes of base matrix construction, and tail biting LDPC convolution code construction.

A. Constructing Finite fields

A finite field with q elements is called a Galois field (GF(q)). A non-binary of code C over GF(2m) is defined by the null space of a sparse MxN parity-check matrix H = (hij) defined over GF(2m) as in equation (1).

$$C = \{x \in GF(2^m)^N \mid Hx = 0 \in GF(2^m)^M\} \quad (1)$$

Where  $x = (x_1, \dots, x_N)$  is a codeword. The  $i^{th}$  parity-check equation for  $c = 1, \dots, M$  is written as  $h_{i1}x_1 + h_{i2}x_2 + \dots + h_{iN}x_N = 0$  where  $h_{i1}, \dots, h_{iN} \in GF(2^m)$  are the entries of the  $i^{th}$  row of **H**. The parameter  $N$  is the code word length in symbols. A non-binary of  $(d_v, d_c)$  is regular, if it has a constant column weight  $d_v$  and a row weight  $d_c$ . In this paper, we focus on  $(2d_v = 2, d_c)$ , the regular non-binary code defined over  $GF(2^8)$  which is empirically known as the best performance code, particularly for short code lengths [18]. The non-binary of **H** can be represented by a Tanner graph with variable and check nodes [19]. Each variable and check node represents a coded symbol and parity-check equation, respectively. The code rate of the non-binary can be computed by code rate  $(R=K/N)$  where  $K$  is message bit and  $N$  is block length. best performance obtaining by Girth 8 [14], 4x4 identity matrix is recommended as the first sub-matrix in this paper so that  $p$  should be set up as the parameter by 4. The next sub-matrix is then created at  $p+1$ :  $p \times 2$  in row direction until  $k$  times. A sample in repeating the sub-matrix in row direction is shown by matrix B in the Fig. 4. In the column direction of the parity check matrix, the sub-matrices are repeated  $j$  times from the most left to the most right in all rows by employing the previous 4x4 identity matrix which all elements in all rows are shifted one bit in the right direction. A couple of samples in creating sub-matrices in column direction of the parity matrix are shown by matrices C and D in the Fig. 4. In determining  $k$  and  $j$  representing size of the parity check matrix, first of all the users have to select expected code rate, and then calculate number of columns and rows. The code rate is equal to the ratio between row and column; code rate = row / column. For instance, 0.5 and 2 are selected as the code rate and number of rows respectively, the number of columns becomes 4 as shown as a sample of parity check matrix in the Fig. 4. In order to evaluate designed parity check matrix, girth that is represented by number of "1"s in clockwise direction in the parity check matrix is used as an indicator for the coding accuracy. In the sample of parity check matrix shown in the Fig. 5, its girth has been counted as eight "1"s which means the designed parity check matrix is a good quality one.

B. Base matrix

The base matrix is a parity check matrix (H). The structure of matrix H basically has a great effect on the encoder, decoder and code performance. A design of the parity check matrix

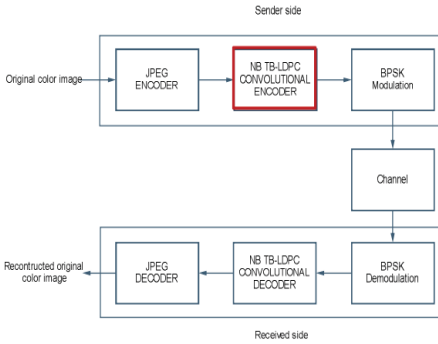


Fig. 2. Block diagram of image transmission.

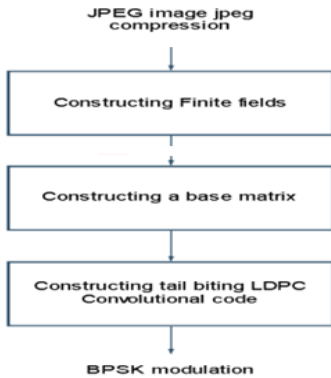


Fig. 3 Block diagram of the proposed method as channel coding.

The proposed method is non-binary tail biting LDPC convolutional code designed with a base matrix with column weight two. We present a coding strategy for half rate for  $x_v$  and The processes of proposed method is described in the following sub-sections according to the flowchart in Fig. 3. The proposed method of non-binary tail-biting LDPC convolutional coding is explained especially in the first process surrounded by a dash line, and followed by a couple of

outlined in this paper is based on column weight two because it applies to the non-binary. In previous work [14], the best performance was exhibited for column weight two. Thus, we design a base matrix with column weight two for non-binary tail biting LDPC convolutional code with a code rate of 0.5 as shown in Fig. 4. In order to determine a parity check matrix, the parity check matrix basically comprises many sub-matrices starting from a sub-matrix in the top left of the parity check matrix as shown by matrix A in the Fig. 4. The first sub-matrix is initially determined by defining the number of sub-matrix rows (p) as a parameter, and the first sub-matrix with the determined rows is then used as the initial sub-matrix to repeat in the row direction k times, where k x p is the total number of rows of the parity matrix. Due to the best performance obtaining by Girth 8 [14], 4x4 identity matrix is recommended as the first sub-matrix in this paper so that p should be set up as the parameter by 4. The next sub-matrix is then created at p+1: p x 2 in row direction until k times. A sample in repeating the sub-matrix in row direction is shown by matrix B in the Fig. 4. In the column direction of the parity check matrix, the sub-matrices are repeated j times from the most left to the most right in all rows by employing the previous 4x4 identity matrix which all elements in all rows are shifted one bit in the right direction. A couple of samples in creating sub-matrices in column direction of the parity matrix are shown by matrices C and D in the Fig. 4. In determining k and j representing size of the parity check matrix, first of all the users have to select expected code rate, and then calculate number of columns and rows. The code rate is equal to the ratio between row and column; code rate = row / column. For instance, 0.5 and 2 are selected as the code rate and number of rows respectively, the number of columns becomes 4 as shown as a sample of parity check matrix in the Fig. 4. In order to evaluate designed parity check matrix, girth that is represented by number of "1"s in clockwise direction in the parity check matrix is used as an indicator for the coding accuracy. In the sample of parity check matrix shown in the Fig. 5, its girth has been counted as eight "1"s which means the designed parity check matrix is a good quality one.

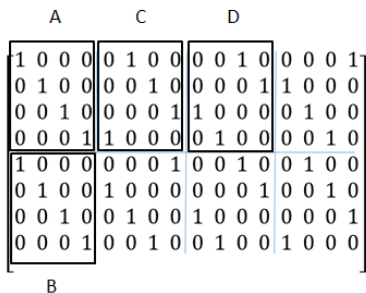


Fig. 4 to determine a parity check matrix

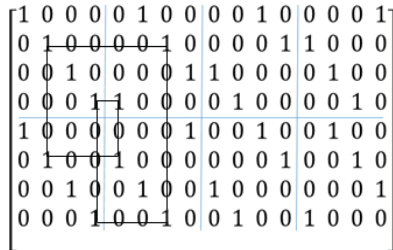


Fig. 5 Girth of the parity check matrix

**Algorithm (Based Matrix Code)**

```

Input: p is an NxN Identity matrix
Output: H is the based matrix
Create the first group of H matrix
Let p1=p
Circular shift left of matrix p and attach to p1 (n-1) times
on the second row
for i = i : n-1
    p1= [ p1 circshift(p ; i) ];
end
create the second group of H matrix utilize circular shift left
of p1 k time and attach to p1 as a second row
H = [ p1 ; circshift(p1 , n) ];
    
```

*C. Constructing Tail-biting LDPC convolutional codes*

In this section, The LDPC convolutional code is defined as LDPC code with rate where c is length of tail biting LDPC code and b is message bits of tail biting LDPC code as in equation (2):

$$u_{[0,i]} = (u_0, u_1, \dots, u_i) \tag{2}$$

Where  $u_{[0,i]}$  is the information sequence that enters the encoder. The encoded sequence is shown in (3):

$$v_{[0,i]} = (v_0, v_1, \dots, v_i) \tag{3}$$

where  $v_i \in GF(2)$ .

The time-varying LDPC code is defined as the set of all sequences  $v_{[0,\infty]}$  satisfying the equation  $v_{[0,\infty]} H_{[0,\infty]}^T = 0$ , where



$$H_{[0,\infty]}^T = \begin{bmatrix} H_0^T(0) \cdots H_{m_s}^T(0) & 0 \\ \vdots & \vdots \\ 0 & H_0^T(t) \cdots H_{m_s}^T(t+m_s) \end{bmatrix} \quad (4)$$

is a semi-infinite transposed parity check matrix, called the syndrome former. For a rate code  $R = b/c$ , the elements of  $H_{[0,\infty]}^T$  are sub matrices of dimension  $c \times (c-b)$  given by

$$H_j^T(t) = \begin{bmatrix} h_j^{(l,1)}(t) \cdots h_j^{(l,c-b)}(t) \\ \vdots \\ h_j^{(c,1)}(t) \cdots h_j^{(c,c-b)}(t) \end{bmatrix}, j=0, \dots, m_s \quad (5)$$

where, at least for one time-instant  $t$ , we have  $H_{m_s}^T(t) = 0$ . The value  $m_s$  is the syndrome former memory. Moreover, the syndrome former defining a regular  $(J, K)$  LDPC code has exactly  $J$  ones in each row and  $K$  ones in each column. For practical applications, periodic syndrome former matrices with period  $T$  are used. In [16], the methods that can be used to construct a periodical LDPC code based on the unwrapping procedure are described.

The code sequences  $\hat{V}_{[0,N-1]}$  of the tail-biting LDPC code satisfy the quality

$$\hat{V}_{[0,N-1]} \tilde{H}_{[0,N-1]} = 0 \quad (6)$$

where the transposed parity check matrix  $\tilde{H}_{[0,N-1]}$  of rate  $R=1/2$  TB LDPC code with block length  $2N$  is obtained by wrapping the last  $m_s$  columns of the sub matrices of the syndrome former in (3) after  $N$  time instants. Finally, it can get  $H$  of tail biting matrix.

#### IV. EXPERIMENTAL RESULTS

In this section, the simulation results of the NB TB-LDPC code method as channel coding for image transmission with modulation using BPSK over an AWGN channel at various noise levels are shown.

We consider the performance of the system using by subjective received image visualization, bit error rate (BER) and peak signal noise ratio (PSNR), respectively.

A dataset of containing standard 512x512-pixel jpeg images as shown in Fig. 6 is tested as source images for

simulation [17]. The NB TB-LDPC convolutional code as channel coding is used with the same code rate of 0.5 and the same code length of 1024 to evaluate the performance in comparison with previous works as follows.



Fig. 6 Source images.

We use code lengths of 4096 which is generally used to evaluate the performance for large block sizes because the performance for a large block in the original method is not high. The code rate is 0.5 for simplicity of computing with 30 iterative decoding, which can be improved. The results of the bit error rate curves are given in Fig. 7, and the received images are compared in Fig. 8. The bit error rate curves of the proposed method compared to previous works are shown in Fig. 7, which show that the proposed TB-LDPC convolutional code with non-binary provides the better performance compared with other codes. The proposed TB-LDPC convolutional code with non-binary shows a small improvement in performance from the TB-LDPC convolutional code in the simulation due to basic finite field of  $GF(2^2)$ . The improvement is considered to increase in case of

bigger finite fields.

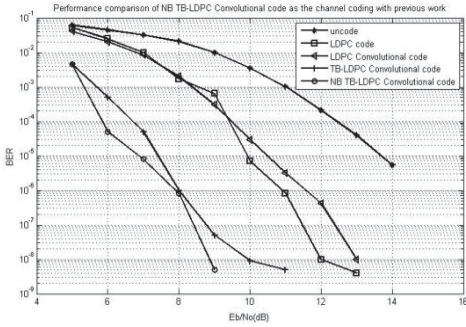


Fig. 7 Performance comparison of the proposed method with previous works on Lena.

A. PSNR

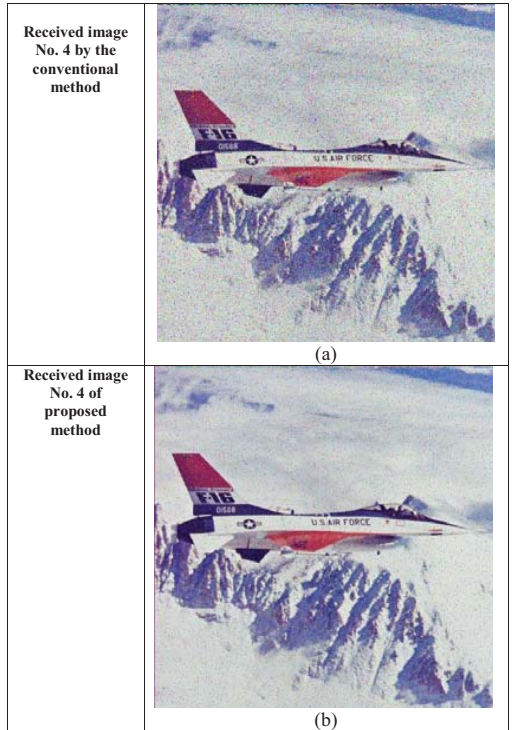
The PSNR between proposed and conventional methods is compared at as low SNR as 2 dB that includes many noise, as shown in Table I.

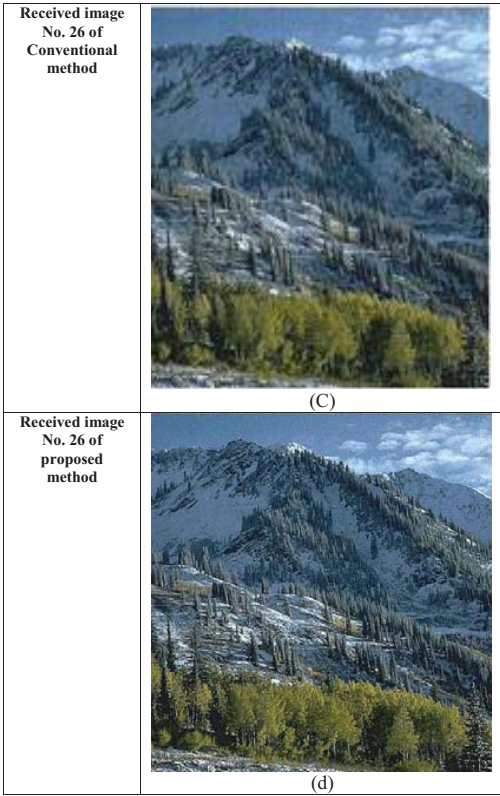
TABLE I. PSNR values of each source image.

Image no.	PSNR Of Conventional method	PSNR Of Proposed method	Difference Between Conventional And Proposed method
1	24.78	27.75	2.97
2	25.82	29.30	3.48
3	24.59	28.51	3.93
4	22.92	24.93	2.01
5	24.34	27.92	3.58
6	24.93	27.87	2.94
7	25.65	29.02	3.37
8	23.97	28.26	4.29
9	23.25	28.26	5.01
10	24.91	29.60	4.69
11	24.14	27.55	3.41
12	24.41	26.76	2.35
13	22.96	26.01	3.05
14	22.40	24.52	2.21
15	20.67	23.15	2.48
16	22.92	26.52	3.60
17	21.66	24.36	2.70
18	23.23	26.51	3.28
19	23.76	27.24	3.48
20	23.53	28.08	4.55
21	23.69	27.63	3.94
22	27.53	29.56	2.03
23	21.22	24.14	2.92
24	22.32	25.21	2.89
25	23.80	27.02	3.22
26	22.27	27.76	5.49
27	23.23	25.83	2.60
28	24.33	26.92	2.59
29	22.16	25.62	3.46
30	21.36	23.98	2.62

As seen in the results shown in Table 1, all PSNR of conventional method are obviously less than the ones obtained from the proposed method. It is therefore proved that the proposed method improves image quality during transmission. Comparison of received image conventional method and proposed method at 2 dB is shown in Table II.

TABLE II. COMPARISON OF RECEIVED IMAGES PERFORMED BY CONVENTIONAL METHOD AND PROPOSED METHOD AT 2dB

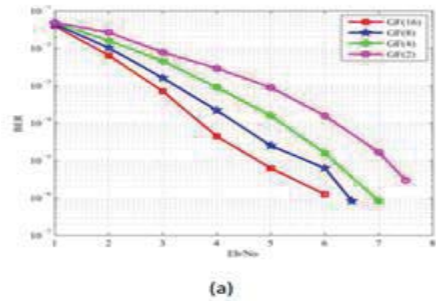




LDPCC code using code gain when BER is 10<sup>-4</sup>. At N=2048, with  $m_s=31$ , the performance of NB TB-LDPCC code is better than TB-LDPCC code by approximately 0.5 dB. At N=4096 with  $m_s=63$ , the performance of NB TB-LDPCC code is better than TB-LDPCC code by approximately 0.6 dB. At N=8192, the performance of NB TB-LDPCC code is better than TB-LDPCC code by approximately 0.45 dB. The best performance at half rate is N=8192, with  $m_s=127$ . In addition, to improve the efficiency of the robust transmission of image at half rate, the proposed method designed with N=8192 and  $m_s=127$  is applied with the field size increased up to GF(2<sup>8</sup>) to find the best performance for robust image transmission at half rate. The objective measured quality of the received images can be evaluated by PSNR, as calculated using equation (8), as follows:

$$PSNR = 10 \log \frac{K^2}{MSE} \quad (8)$$

where K is the maximum grayscale value of the image.



(a)

Field size	Used Memory
GF (2 <sup>4</sup> )	4 GB
GF (2 <sup>4</sup> )	4 GB
GF (2 <sup>4</sup> )	4 GB
GF (2 <sup>4</sup> )	4 GB
GF (2 <sup>4</sup> )	8 GB
GF (2 <sup>4</sup> )	8 GB
GF (2 <sup>4</sup> )	8 GB
GF (2 <sup>4</sup> )	8 GB

(b)

Fig. 8 (a) Image transmission BER of TB LDPCC code with various field sizes and (b) Used memory

*B. Comparison of BER & Memory in each Finite field with various block length*

Next, we use the proposed method to design a base matrix with column weight two, because Davey and Mackay [14] proved that column weight two can improve performance over a finite field. The block length is 1008, and the code rate is 0.5 for evaluating the performance of field size. After examining the performance within the limits of the finite field, the performance in small fields is presented. The results in Fig. 8 show that GF (2<sup>4</sup>) exhibits the best performance, and we can observe that BER is improved when the field size is increased. The performance of the base matrix (64, 2, 4) at half rate with various  $m_s$  designs, as shown in table 1, is considered with the block length given by  $N = 64(m_s + 1)$ . Based on the simulation results, Fig. 9 shows the bit error rate curves of TB LDPCC code with GF (2<sup>4</sup>). Based on the BER shown in Fig. 9, we can see that good performance of the TB LDPCC code over the finite field is better than the performance of the TB

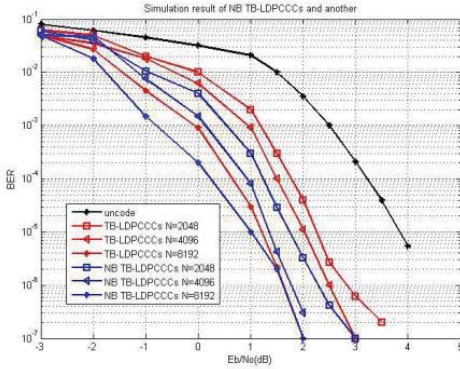


Fig. 9 Image transmission BER of TB LDPCC code with GF (2<sup>4</sup>).

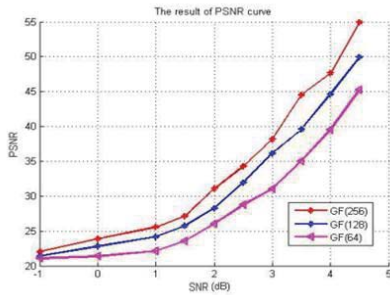


Fig. 10 The PSNR curves of the recovered images.

Based on the simulation results, the PSNR curves of the recovered images over three finite fields are shown in Fig. 10. Compared with the transmitted images at 2 dB, the average of PSNR values of GF (256), GF (128) and GF (64) are approximately 32, 27 and 21, respectively

### V. DISCUSSION

Comparing with previous works, the TB-LDPCC code has a zero-tail for the termination result that can maintain rate loss. However, the performance of the binary TB-LDPCC code is not high compared with the proposed NB TB-LDPCC code, as shown in Fig. 7. At the same BER, the original work has a code gain of approximately 5 dB on Lena image. However, our proposed method has a code gain of approximately 6 dB, meaning that the proposed method outperforms the original

method by approximately 14% due to quantization error of binary method. The performance advantage of non-binary code is up to approximately 0.5 dB when the block length is increasing. The NB TB-LDPCC code with column weight two over GF (2<sup>8</sup>) with N=8192 with =127 can maintain better quality of the received image at low SNR. In this paper, image transmission is implemented using NB TB-LDPCC code. A dataset of standard image 512x512 pixel JPEG images are tested as the original images, as shown in Fig. 6. PSNR of proposed method can improve the performance of receive image around 25 % comparing with the received image resulting from NB TB-LDPCC code over GF(4) and other conventional codes. The results show that NB TB-LDPCC code has a high PSNR and better performance in terms of the received image. As seen in Table 2, obviously image No. 4 with low resolution in brightness axis is improved by our proposed method less than the one in No. 26 with high resolution. This means our proposed method is especially effective in high-resolution images. The performance of NB TB-LDPCC code with column weight two can improve when the field size increases. Inevitably, the good performance of NB TB-LDPCC code involves a trade-off in the system between memory and field size. That is when field size is increased; bigger memory is required, as seen in Fig. 8 (b). However, GF (2<sup>8</sup>) obtains the best PSNR while memory is required as same as 8 Giga bytes as other GFs. Thus, we recommend using at least 8 Giga bytes memories in order to increase efficiency in image transmission.

### VI. CONCLUSIONS

Image transmission implemented using non-binary TB-LDPCC code with column weight two for the goal of low BER has been presented in this paper. The proposed design method can reduce bit error when code length increases at various levels of noise and maintain adequate quality of the received image. Good performance in correcting data error at low SNR is shown in the simulation results. In addition, NB TB-LDPCC code with base matrix column weight two that presents a performance improvement of at least 0.5 dB compared with TB-LDPCC code and can adequately maintain the quality of the received image at low SNR. The best performance at half rate for image transmission is N=8192 with =127 over GF (2<sup>8</sup>).

### REFERENCES

- [1] R. G. Gallager, "Low-Density Parity-Check Codes". MIT Press, Cambridge, MA, 1963.
- [2] D. J. C. Mavkay and R. M. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes", Electronic Letters, Vol. 32, No. 18, pp.1645-1646, August 1996.
- [3] D.J.D.Mackay, M.C.Davey., "Evaluation of Gallager Codes for Short Block Length and Hige Rate Application", In Codes. System and Graphical Models. Ed., vol. 123 of IMA, Springer-Verlag, New York, pp. 113-130, 2000.

- [4] W.E.Ryan, Fan, Wang, et., "Optimal Code Rates for the Lorentzian Channel: Shannon Codes and LDPC Codes", IEEE Transactions on Magnetics, vol. 40, pp. 3559-3565, 2004.
- [5] H.Yongqing, P.ENG Qicong and S.Huaizong, "The Application of High-Rate LDPC Codes in Image Transmission over Wireless Channel", pp.62-65, 2006.
- [6] S. J. Johnson and S. R. Weller, "Quasi-cyclic LDPC Codes from Difference Families", 3rd AusCTW, Canberra, Australia, February 4-5, 2002.
- [7] M. P. C. Fossorier, "Quasi-Cyclic Low-Density Parity-Check From Circulant Permutation Matrices", IEEE Trans. Information Theory, Vol. 50, No. 8, August 2004.
- [8] Z. Wang and Z. Cui, "A Memory Partially Parallel Decoder Architecture for QC-LDPC Codes", IEEE 2005.
- [9] N. Kodani, "Traveling Crane Using H2 Controller Based on Minimum Order Observer", SICE Annual Conference, Fukui University, Japan, August 4-6, 2003.
- [10] P.Tanaporn, C.Werapon and C.Kosin, "Sub-block Encoder Design of High-rate QC-LDPC Code for Image Transmission", Proceedings of the 3rd Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Association (APSIPA ASC'11) - Short Paper, Grand New World Hotel Xian, Xi'an, China, pp. 1-4, October 18-21, 2011.
- [11] P.Tanaporn, C.Werapon and C.Kosin, "Improved Sub-block Encoder of Irregular LDPC Code for Image Transmission", The 4th IEEE/SICE International Symposium on System Integration (SII'11), Kyoto University, Yoshida Campus, Kyoto, Japan, pp. 41-44, December 20-22, 2011.
- [12] P.Tanaporn, C.Werapon and C.Kosin, "Sub-block Encoder of High-Rate LDPC Code over Fading Channel for Image Transmission", The 9th Annual International Conference of the Electrical Engineering/Electronics, Computer, Telecommunications, and Information Technology Association (ECTI-CON'12), Novotel Hua Hin - Cha Am Beach Resort & Spa, Cha-Am, Phetchaburi, Thailand, pp. 1-4, May 16-18, 2012.
- [13] T.Hehn and J.Huber, "LDPC codes and convolutional codes with equal structural delay: A comparison", IEEE transactions on communication, 57, 1683-1692, 2009.
- [14] D. J. Costello, S. Bates Pusane and K. SH. Zigangirov, "A comparison between LDPC block and convolutional codes", Proc. Information Theory and Applications Workshop, 2006.
- [15] Najeeb ul Hassan, L.Michael, and Gerhard P. Fettweis, "Comparison of LDPC Block and LDPC Convolutional Codes Based on their Decoding Latency", International symposium on turbo code and iterative information processing (ISTC), 225-229, 2012.
- [16] tavares, marcos b. s., zigangirov, kamil sh. and fettwiss, gerhard p. "tail-biting ldpc convolutional codes", isit, 2341-2345, 2007.
- [17] <http://sipi.usc.edu/database/database.php>

# Teaching Chinese as a Foreign Language Teachers' Epistemic, Pedagogical and Technological Beliefs: The Case Study of Northern Thailand

Fang Yuan <sup>[1]</sup>  
fang.yua@mfu.ac.th  
School of Sinology  
Mae Fah Luang University

Nikom Rongbutrsri <sup>[2]</sup>  
nikorn@mfu.ac.th  
School of IT  
Mae Fah Luang University

Thanon Phunsapphaisan <sup>[3]</sup>  
thanon.phu@mfu.ac.th  
School of Sinology  
Mae Fah Luang University

*Abstract—This paper demonstrates the study of three belief dimensions: epistemology, pedagogy, and ICT of Thai teachers in Chiang Rai, who teach Chinese as a foreign language. The researchers implemented survey and use basic statistical analysis for data processing. The results indicate that Thai teaching Chinese as a foreign language teachers are more traditional in their epistemological and pedagogical beliefs. They implement ICT in their traditional teaching; however, to utilize ICT in their education implementing constructivist pedagogy should be considered.*

*Keywords—CFL, Epistemology, Pedagogy, ICT, Teaching*

## I. INTRODUCTION

Research motivation and objectives are discussed here including the current stage of education system and potential of ICT integration to education in a developing country-Thailand.

### A. Thai education, ICT integrated education

Evidence shows that Thailand has started their education system since Sukhuthai Era (1285) [1]. Education models are different in different eras, depending on their social. In 1999 the first construction of Thailand education was announced formally; in its 9th topic declared about ICT application in the education system [2]. However, this was not the first ICT application in education policy in Thailand, in 1996, the ministry of ICT, announced its first ICT national policy call IT 2000; there are three main missions, one of them is the government investment in good education and information personnel [3]. ICT (Information and Communication Technology) has been integrated into education widely in different countries including Thailand. Education is not limited only classroom activities, but the ability to access knowledge, information and news at any time and any place through technology. "21st century skills" is another framework to enable new ways of ICT integration which can provide freedom to learners. Schools construct their own curricula which response to the needs of their local communities. ICT enables them to sustain their learning material (1).

Chiang Rai as a province locates in Northern Thailand which have applied ICT into their school level education; specially schools in city area, they equipped with good facilities and

infrastructure; therefore, they are much more ready than schools in rural area. However, at secondary and high school levels, most of them integrate ICT only into their traditional education, e.g. to support their presentation (presentation tools and streaming video), to support their provided learning material. It is less application in primary school level, teachers do not allow students to utilize ICT in their formal learning activities; however, students who can access to the technology, use it for entertainment purposes e.g. chatting, games, movies, songs.

Teachers' beliefs have been perceived as an important field that needs to be investigated in order to reform education, teachers' beliefs influence their behavior of teaching and learning [4]. Many studies suggested teachers' beliefs could affect their practice and learning. Ching Sing Chai (2006) [5] shows that education system should move forward from traditional to constructivist. Even though ICT can be applied in both pedagogy: traditional and constructivist, the potential of ICT is more suitable for democratic, participation and personalized, which are characteristics of constructivist. Ching Sing Chai(2010) [6] suggests that there is relationship between the three belief domains. Therefore, when the three domains of belief can be understood, ICT integration can be more effective.

There are very few research about Thai teachers' beliefs, a current examination of these beliefs is very rare. In this study, we investigated Thai Chinese language teachers' epistemic beliefs (beliefs about knowledge and learning), their pedagogical beliefs (beliefs about teaching) and their beliefs of use ICT.

The aim of this research is to how Thai teachers who teach Chinese as a foreign language, how do they believe about knowledge, pedagogy and application of ICT.

## II. LITERATURE REVIEW

In this section we demonstrate the existing concepts of teachers' beliefs towards technological and how relevant beliefs including epistemic beliefs.

### A. Teachers' epistemic beliefs

The *epistemic beliefs concept* is typically thought of as concerning the nature of knowledge and learning. Including beliefs about how to know, how construct knowledge, and how to evaluate it [7].

Perry (1970) [8] provided the epistemological developmental model which consists of four stages: (a) dualistic, (b) multiplicative, (c) relativistic and, (d) commitment with relativism. Schommer (1990) [10] contended that the developmental model may not be able to simulate the complication of epistemic beliefs because of its unidimensional nature. She proposed a multidimensional model that conceives epistemic beliefs as an independent beliefs including: (a) innate ability, (b) quick learning, (c) simple knowledge, (d) certain knowledge.

Epistemological thinking is a major factor of lifelong learning; it influences an individual's learning and eventually teaching. The understanding of individual epistemological belief leads us to understand the person's behaviors of learning and teaching others [11], [12]. There are two models: developmental model [8] and the multidimensional model [10] to understand pre-service, in-service teachers' epistemic beliefs.

Several researchers had implemented multidimensional model to investigate pre-service teachers, in-service teachers in Asian context mainly in Hong Kong, Singapore and South China [13] [15]; some found the relation factor that influences individual epistemic belief including culture [13], educational context [14]. The literature suggests that teachers' epistemic beliefs may influence their ways of teaching [22] and their pedagogical beliefs are the link between their epistemic beliefs and their teaching practice. [14]; pedagogical beliefs are in the next section.

### B. Teachers' pedagogical beliefs

Teachers' pedagogical beliefs refer to their ways of teaching and learning that they find it is best suit them; including the meaning of teaching and learning, and how teachers and students are supposed to interact during teaching and learning process.

Kember (1997) [23] categorizes pedagogical beliefs as either being teacher-centred or student-centred. The teacher-centred perspective focuses on transmission or direct instruction [24] where learners are passive recipients of content knowledge [25].

The learner-centred perspective advocates active learning, guided discovery, focusing on understanding instead of memorization, application of knowledge, alternative forms of assessment, self-assessment, reflection, etc.. The above indicates that the teacher-centred conception refers to the behaviouristic perspective or traditional perspective, whereas the learner-centred conception refers to constructivism [13], [25].

In Asia context, the studies of teachers' pedagogical beliefs indicate that the pre-service teachers from Singapore, China, Hong Kong and Taiwan generally are inclined to agree with the constructivist beliefs.

### C. Teachers' ICT beliefs

Teachers' ICT beliefs refers to the belief of individual teachers towards employing ICT into their daily life; ICT is composed of devices, applications, infrastructure and users. ICT capacity can be classified to different levels: basic operation e.g. office application skill, internet skills, social media, customization, and creation. The application of ICT in their teaching and their learning are also curious whether an individual teacher perceives ICT as a tool to support his/her activities or it can enhance or even make a new way of classroom interaction.

ICT which can be used as a pedagogical tool can be applied for different pedagogies or teaching approaches. Ertmer (2005) [26] mentioned that teachers' use of ICT is influenced by their beliefs. The researches indicate that teachers' beliefs have significant influences over their espoused use of ICT. Teachers who hold the traditional pedagogical beliefs would use of ICT, mainly to support students' acquisition of knowledge and traditional teaching. The teachers who are inclined towards constructivist teaching tend to use the ICT as cognitive tools to support students' personal construction of knowledge [27]. While in Asian context, Chai (2010) [21] found that the Singaporean teachers' constructivist beliefs are correlated with both constructivist use and traditional use of ICT. Chai explained that the teachers may see that the traditional use of ICT could be a way to enhance students' understanding before engaging the students in collaborative co-construction of knowledge.

Deng et al. [28] found that the Chinese teachers' constructivist-oriented pedagogical beliefs are a significantly positive in using technology from different cultures. Deng's study did not discuss about traditional pedagogical beliefs in relation to ICT integration. Epistemic beliefs are related to the ICT in constructivist but not in traditional.

In summary, In Asian countries including Singapore, Taiwan, Hong Kong and China have tested their teachers in relation to Epistemic, Pedagogical and technological beliefs. They found relationship between the three beliefs dimensions with also the contextual variables e.g. culture, political perspectives.

## III. METHOD

This section we discuss the research setting including selecting participants, research procedures and research instruments.

### A. Participants

27 in-service teachers from Chiang Rai Northern Thailand who teach Chinese as a foreign language participated in this study. There are 2 teachers have more than 15 years teaching experience, 2 teachers have 2-3 months teaching experience. Others have 5-6 years teaching experience. After obtaining the permission from the school of Sinology, Mae Fah Luang University, the authors distributed the questionnaire during an ICT training in the university. All the training participants participate the questionnaire.

**B. Procedure**

This study used a survey method. The participants were asked by getting questionnaires to rate their responses for items of the epistemological, pedagogical and technological beliefs on a five-point scale which can identify from 1 is Strongly Disagree and 5 is Strongly Agree. The survey questions are in Thai- the local language. The survey instrument consisted of four sections: Section A-Background Informaion, Section B-Epistemological beliefs, Section C-Pedagogical Beliefs, Section D-Technological Beliefs. Section A asking about their background information such as age and the years, subjects of teaching before starting the questionnaire. The first questionnaire (Section B)was measure the epistemological beliefs ,the second questionnaire( Section C) were adopted from Chan and Elliott(2004) [13]. These two questionnaire were to examine the concepts of the pedagogy held by the Thai in-service teachers, and the third questionnaire (Section D) the use of ICT questionnaire.

**C. Research instruments**

**1) Epistemological beliefs questionnaire (EBQ)**

The EBQ was adapted from Chan and Elliott(2004) [13] and Ching Sing Chai [6] and translated into Thai. The questions under this section is composed of four dimensions:

- Innate/ Fixed Ability,*
- Learning Effort/Process,*
- Authority/ Expert Knowledge and*
- Certainty Knowledge.*

*Innate/Fixed Ability* refers to ability which is got at birth and unchangeable on one pole, while at the other ability is dynamic and developable.

*Learning Effort/Process* refers to effort and hard work to get into understanding.

*Authority/Expert Knowledge* refers to the structure of handling knowledge whether it is passed through experts or learners obtain by their own decision.

*Certainty Knowledge* refers to the understanding that knowledge is unchanged for one pole or dynamic, contextualized at the other.

**2) Pedagogical beliefs questionnaire (PBQ)**

The PBQ is to test teachers’ pedagogical beliefs which has two poles: traditional and constructivist pedagogies; 30 items were initiated by Chan & Elliott [13].

**3) The use of ICT questionnaire (UIQ)**

The UIQ comprised two subscale. General ICT competencies (GIC) and pedagogical-oriented ICT competencies (PIC) are based on the TPACK framework. The pedagogical-oriented ICT competencies subscale including 2 parts: traditional use of ICT (5 items) and Constructivist use of ICT (6 items).this paper adapted the questions under this belief domain from Ching Sing Chai [6] and add some more questions regarding the pedagogical-oriented ICT competencies.

**IV. ANALYSIS**

**A. Basic information**

There are 27 participants; 48% have a 3-7 years teaching experience, 30% have a long time teaching experience, and 19% are new teachers. 56 % are secondary or high school (Mathayomsuksa) teachers; the others are primary school (Prathomsuksa) teachers. Every teacher teaches Chinese Speaking; most of them teach Chinese Characters, Chinese Writing, Chinese Calligraphy, Chinese Conversation, Chinese Listening and Chinese Reading.

**B. Epistemology’s belief**

Questions	AVG	STD
1. You cannot do much to make yourself starter, because the ability is fixed at birth.	1.64	0.91
2. Learning abilities are fixed at birth	1.92	0.81
3. We are limited by our innate ability.	2.20	0.91
4. Some are good since their birth, some are limited by their abilities.	2.48	1.19
5. Some cannot learn well in certain subjects since their born.	3.08	1.12
6. Learning ability is innate.	2.60	1.19
7. Students with average ability will remain the same level of their study.	1.32	0.56
8. Smart students do not need to work hard to learn well.	2.44	1.23
9. Keep trying if you don’t understand a thing.	4.20	0.91
10. Learning ability is important than the acquired facts or knowledge.	3.72	0.84
11. We learn little if not try hard.	4.12	0.78
12. Thinking process and understand learning material are more important than acquired knowledge.	3.52	0.92
13. Everyone need to improve their learning skills.	3.88	0.93
15. We learn better if concentrate on the process of understanding, not the acquired knowledge.	3.68	0.80
16. To really understand a thing, you need much time and effort.	4.60	0.50
17. The more effort you put, the more learning you get.	4.64	0.57
18. Progression need a lot of work.	3.16	1.07
19. If try hard enough, you will always understand the material.	3.84	0.62
20. Wisdom is knowing how to get the answer, not the recalling of the answer.	4.32	0.75
21. A textbook written by an authority sometimes may not be reliable.	3.60	0.87



22. Advice from experts should be critically consider.	3.96	0.89
23. I often wonder how much experts really know	3.60	1.00
24. I rely on teachers' words more than my own.	3.40	1.00
25. Suggestion from experts is important than my knowledge.	2.60	0.71
26. I always believe in whatever the experts say	2.56	0.58
27. Scientists will get to the truth if they keep trying.	4.08	0.81
28. If scientists try hard enough, they can find any truth.	4.00	0.82
29. One can understand difficult concepts if one tries hard enough	4.24	0.72
30. I think there is a pedagogy that applicable to all learning situations	3.84	1.11
31. Scientific knowledge is unchangeable	3.00	0.96

From the results can see that the teachers have positive about potential of human can overcome any unknown and improve themselves; anyone can be improved by keep trying and not giving up. Additionally, they reject ability is fixed at birth which confirms the first point. However, they are not totally convinced about the importance of learning processes over the content; this can show signs of traditional education perspective. This is more significant when we look at the perspectives of knowledge; they are not totally convinced that knowledge is dynamic and still rely on expert.

### C. Pedagogy belief

Questions	AVG	STD
1. To teach well, a teacher must understand the feelings of the students.	4.68	0.48
2. Good teachers always let students to think for answers by themselves.	4.64	0.49
3. Learning is opportunities for students to explore, discuss and express their ideas	4.60	0.65
4. A democratic and free atmosphere is needed in a classroom to stimulates students to think and interact	4.44	0.58
5. A child needs their personalized learning to serve their particular needs.	4.28	0.84
6. Teaching can be effective if there is more discussion and practical activities for students.	4.24	0.83
7. Teaching should focus on knowledge construction rather than knowledge transmission.	4.16	0.80
8. Teachers should define instruction which is flexible enough to accommodate individual differences of students	4.32	0.69
9. Teachers should define different objectives and expectations in different students.	4.04	0.93

10. Students should often express their idea.	4.36	0.70
11. Students' ideas are important and should be carefully considered.	4.20	0.91
12. Good teachers always make their students feel important.	4.24	0.93
13. A teacher's major task is transferring knowledge to students and test their memories.	3.56	1.00
14. In classroom, students must confine to the text book and sit properly on their desks.	2.44	1.12
15. Learning is remembering teachers' teaching.	2.52	0.65
16. A Good student must follow instruction and work quietly.	2.32	0.85
17. Lecturing is the best pedagogy because it can cover all topics in lesson plan.	2.20	0.71
18. Classroom should be utilized by practicing exercises.	3.60	0.91
19. Teachers should talk the most in a classroom.	2.36	0.70
20. Learning is receiving knowledge as much as you can.	3.20	0.91
21. A good teacher should provide students with complete knowledge rather than let them find out themselves	2.56	0.77
22. A good teacher must correct misunderstanding of the students right the way, rather than let them correct it themselves.	3.12	1.13
23. Students cannot learn without controlling	3.04	0.84
24. Students must be control during their learning all the time.	3.00	0.87
25. Learning to teach means following the teacher's tasks without questioning them	2.44	0.87
26. When I can recall something, means I learnt it	2.72	0.89
27. Telling, presenting or explaining the subject are teaching methods.	3.32	0.90
28. A teacher's role is to transfer knowledge to his/her students.	3.56	1.23
29. Learning occurs primarily from drilling and practice.	4.12	0.73

The results reveal that teachers have very strong believe in constructivist as shown in their answers in question 1 to 12. However, the other questions which reflects on their perspectives of traditional education model, show that they still rely on it even though they aware that there is a better way of teaching and learning. From the these results reveal that these teacher aware of the potential of constructivist which can give better result than traditional model; however, they still are not totally convinced to give up their traditional model; even though they are not impressed on traditional model but it is acceptable for them.

*D. Technology belief*

Questions	AVG	STD
1 I can use computers effectively.	3.68	0.80
2 I learn to use technology easily.	3.56	0.77
3 I solve technical problems by myself.	3.00	0.96
4 I always look for information about important new technology	3.32	0.90
5 I can develop a web page.	2.00	0.96
6 I use social media (e.g. Facebook, Youtube, Blog, Wiki).	4.36	0.76
7 I usually use Microsoft office for my work.	4.28	0.89
8 I always look for new application for my entertainment, social and work.	3.40	0.96
9 My school has good ICT support, facilities and infrastructure.	3.72	1.10
10. I can master skills just the teacher taught me.	3.08	0.76
11. I prefer to use ICT to clarify what I cannot understand well.	4.00	0.76
12. ICT is good for presenting information to students.	3.36	1.04
13. I am able to use technology to searches for factual information.	4.36	0.70
14. I am able to handle tutorials drill and practice by ICT.	3.40	0.82
15. Using computer distract me from learning.	3.28	0.84
16. ICT is good for fun but not for deep learning.	2.88	1.01
17. ICT can provide collaborative learning environment.	3.88	0.73
18. I like to use ICT to analyze data.	3.32	1.03
19. I can use ICT for basic learning activities in a classroom.	4.04	0.89
20. I can use computer supported collaborative learning systems	3.84	0.85
21. A learner-centred environment could be created by using ICT.	4.24	0.60
22. ICT improves the quality of teaching	4.28	0.61
23. ICT enables new ways of teaching.	4.28	0.61
24. All learning material can be digitized.	4.12	0.60
25. ICT is an effective way of teaching	4.56	0.58
26. ICT motivates students in learning	4.44	0.65
27. ICT encourages students to be active in their learning	4.16	0.75

The results show that most schools provide good infrastructure and equipment; but teachers have only basic ICT skills; they apply ICT for their social, entertainment and some basic operations in their profession. Their way of IT integration to their teaching is still in traditional model e.g. search for information, presentation. However, they see the potential of ICT can improve class participation, motivation and

eventually the quality of learning. ICT for constructivist is still not clear for these teachers.

V. CONCLUSION

Chiang Rai is a rural area and far from Bangkok; people live in their traditional life and mostly rely on agriculture. Schools in Chiang Rai are active in teaching Chinese even though it is not compulsory according to the curriculum from the education ministry. Chinese teaching in Chiang Rai becomes good marketing for schools. Several collaborations with China educational organizations have been made locally. Teaching and learning activities rely on face-to-face interaction. Little of ICT is integrated in their teaching and learning mostly to support their traditional teaching. This study aims to understand the current stage of teachers' perspectives on epistemology, pedagogy and technology in order to further facilitate the teachers to utilize ICT in their constructivist model which is more sustain than traditional one. The three dimensions relate to the potential to implement an ICT – supported pedagogy which is suitable for the context. The study reveals school teachers concern and believe in the potential of constructivist which roots in participatory and democracy. However, as the life style is in traditional way, they perceive the disadvantages of traditional education but still rely on it as their comfort zone, which is difficult to change or get over it. Teacher utilize technology for their entertainment and social but only use ICT for basic professional operations e.g. communication, searching, documentation, spreadsheet, and presentation. They also implement ICT into their classroom activities but mostly to support their traditional way of teaching which relies on knowledge transferring. ICT mostly is used for finding information and presenting information to students but very rare on discussing and constructing knowledge. As we can see the school teachers in Chiang Rai aware of constructivist and its potential but not enough to put it into practice or leave their comfort zone. They probably need to gain more experience and some knowhow on constructivist and ICT integration; a hand-on workshop or training of implementing constructivist is required. Then participation and democracy will dominate Chinese teaching in Chiang Rai schools; deep learning according to bloom' taxonomy can be expected. The only then can utilized the available teaching and learning tools on the markets which can support different activities rather than just searching, presenting, documentation, and storing. Online discussing, co-drawing, co-author for children story, story-telling, commenting could be seen on their virtual learning environments.

In conclusion, this study is a case study from a developing country (lacking of resources, low GDP-Gross Domestic Product, relying on low-income sectors-agriculture) perspectives in three dimensions: epistemology, pedagogy and technology towards integration ICT into secondary language learning. Tradition dominates most sections in the region including education and how technology is used. This shows political perspective effecting learning and education. Implementing constructivist and democracy can be perceived as a treat to their tradition. Therefore, changing to constructivist can be sensitive. Implementing ICT to support their traditional educational model is done, but cannot get the

full potential of the technology. On the other hand, implementing ICT requires a big amount of budget for devices, infrastructure, applications, and training. Therefore, optimizing ICT for their full potential is required to ensure the efficient of investment.

## REFERENCES

- [1] อดิสรณ์ สุรพล, วัฒนาการและแนวโน้มการบริหารการศึกษาไทย : อดีตสู่อนาคต. Bangkok: มหาวิทยาลัยบูรพา, 2000. (Evolution and Trends of Thai Education Administration: From the Past to the Future by Surapol)
- [2] กระทรวงศึกษาธิการ, พระราชบัญญัติการศึกษามัธยมศึกษา พ.ศ. 2542. Bangkok: โรงพิมพ์จุฬาลงกรณ์, 1999. (National Education Act 1999 by Thai Ministry of Education)
- [3] สำนักงานเลขาธิการและคณะกรรมการเทคโนโลยีสารสนเทศแห่งชาติ, กรอบนโยบายเทคโนโลยีสารสนเทศและการสื่อสาร ระยะ พ.ศ.2554-พ.ศ.2563 ของประเทศไทย. Bangkok: สำนักงานเลขาธิการและคณะกรรมการเทคโนโลยีสารสนเทศแห่งชาติ, 2011. (Thailand Policy Framework for Information Technology and Communication Act of 2011-2020 by National Information Technology Committee Secretariat (NITC))
- [4] A. Woolfolk Hoy, H. Davis, and S. J. Pape, "Teacher knowledge and beliefs," *Handb. Educ. Psychol.*, vol. 2, pp. 715-737, 2006.
- [5] C. S. Chai, M. S. Khine, and T. Teo, "Epistemological beliefs on teaching and learning: A survey among pre-service teachers in Singapore," *Educ. Media Int.*, vol. 43, no. 4, pp. 285-298, 2006.
- [6] C. S. Chai, "The relationships among Singaporean preservice teachers' ICT competencies, pedagogical beliefs and their beliefs on the espoused use of ICT," 2010.
- [7] B. K. Hofer and P. R. Pintrich, "The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning," *Rev. Educ. Res.*, vol. 67, no. 1, pp. 88-140, 1997.
- [8] W. G. Perry Jr, *Forms of Intellectual and Ethical Development in the College Years: A Scheme. Jossey-Bass Higher and Adult Education Series*. ERIC, 1999.
- [9] M. F. Belenky, B. M. Clinchy, N. R. Goldberger, and J. M. Tarule, "Connected teaching," *Womens Ways Knowing N. Y. Basic*, pp. 214-229, 1986.
- [10] M. Schommer, "Effects of beliefs about the nature of knowledge on comprehension," *J. Educ. Psychol.*, vol. 82, no. 3, p. 498, 1990.
- [11] P. M. King and K. S. Kitchener, *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults. Jossey-Bass Higher and Adult Education Series and Jossey-Bass Social and Behavioral Science Series*. ERIC, 1994.
- [12] D. Kuhn, *The skills of argument*. Cambridge University Press, 1991.
- [13] K.-W. Chan and R. G. Elliott, "Relational analysis of personal epistemology and conceptions about teaching and learning," *Teach. Teach. Educ.*, vol. 20, no. 8, pp. 817-831, 2004.
- [14] B. K. Hofer, "Personal epistemology and culture," in *Knowing, knowledge and beliefs*, Springer, 2008, pp. 3-22.
- [15] M. M. Cheng, K.-W. Chan, S. Y. Tang, and A. Y. Cheng, "Pre-service teacher education students' epistemological beliefs and their conceptions of teaching," *Teach. Teach. Educ.*, vol. 25, no. 2, pp. 319-327, 2009.
- [16] C. C. Sing and M. S. Khine, "Assessing the epistemological and pedagogical beliefs among pre-service teachers in Singapore," in *Knowing, Knowledge and Beliefs*, Springer, 2008, pp. 287-299.
- [17] C. S. Chai, F. Deng, B. K. S. Wong, and Y. Qian, "South China education majors' epistemological beliefs and their conceptions of the nature of science," 2010.
- [18] L. D. Bendixen and F. C. Feucht, *Personal epistemology in the classroom: Theory, research, and implications for practice*. Cambridge University Press, 2010.
- [19] N.-H. Kang and C. S. Wallace, "Secondary science teachers' use of laboratory activities: Linking epistemological beliefs, goals, and practices," *Sci. Educ.*, vol. 89, no. 1, pp. 140-165, 2005.
- [20] L. J. Olafson and G. Schraw, "Some final thoughts on the epistemological melting pot," *Issues Educ.*, vol. 8, no. 2, pp. 233-247, 2002.
- [21] C. S. Chai, "Teachers' epistemic beliefs and their pedagogical beliefs: A qualitative case study among Singaporean teachers in the context of ICT-supported reforms," *TOJET Turk. Online J. Educ. Technol.*, vol. 9, no. 4, 2010.
- [22] M. F. Pajares, "Teachers' beliefs and educational research: Cleaning up a messy construct," *Rev. Educ. Res.*, vol. 62, no. 3, pp. 307-332, 1992.
- [23] D. Kember, "A reconceptualisation of the research into university academics' conceptions of teaching," *Learn. Instr.*, vol. 7, no. 3, pp. 255-275, 1997.
- [24] P. A. Ertmer, "Ottenbreit - Leftwich, AT, Sadik, O., Sendurur, E., and Sendurur, P.(2012). Teacher beliefs and technology integration practices: A critical relationship," *Comput. Educ.*, vol. 59.
- [25] J. G. Brooks and M. G. Brooks, *In search of understanding: The case for constructivist classrooms*. ASCD, 1999.
- [26] P. A. Ertmer, "Teacher pedagogical beliefs: The final frontier in our quest for technology integration?," *Educ. Technol. Res. Dev.*, vol. 53, no. 4, pp. 25-39, 2005.
- [27] R. Hermans, J. Tondeur, J. van Braak, and M. Valcke, "The impact of primary school teachers' educational beliefs on the classroom use of computers," *Comput. Educ.*, vol. 51, no. 4, pp. 1499-1509, 2008.
- [28] F. Deng, C. S. Chai, C.-C. Tsai, and M.-H. Lee, "The Relationships among Chinese Practicing Teachers' Epistemic Beliefs, Pedagogical Beliefs and Their Beliefs about the Use of ICT," *Educ. Technol. Soc.*, vol. 17, no. 2, pp. 245-256, 2014.
- [29] K.-W. Chan and R. G. Elliott, "Exploratory study of epistemological beliefs of Hong Kong teacher education students: Resolving conceptual and empirical issues," *Asia-Pac. J. Teach. Educ.*, vol. 28, no. 3, pp. 225-234, 2000.
- [30] K. Chan and R. G. Elliott, "Exploratory study of Hong Kong teacher education students' epistemological beliefs: Cultural perspectives and implications on beliefs research," *Contemp. Educ. Psychol.*, vol. 27, no. 3, pp. 392-414, 2002.

# Extend SDN Network by Multi-controller Based on Fibonacci Heap

Chan Wang<sup>1</sup>, Zhuqing Jiang<sup>1</sup>, Yi Zhao<sup>2</sup>, Yang Liu<sup>1</sup>, Rui Qu<sup>1</sup>, Aidong Men<sup>1</sup>, Chonghua Liu<sup>2</sup>, Zhendong Li<sup>2</sup>

<sup>1</sup>Beijing University of Posts and Telecommunications, Beijing, China

<sup>2</sup>China Academy of Space Technology, Beijing, China

**Abstract**—In Software Defined Networking (SDN) architecture, the control and data planes are separated, which makes controller manage the data plane in a logically centralized way. However, it is difficult to scale SDN network because of the centralized control plane, and when network is large-scale, sheer propagation delay will be the most related factor arguing a single monolithic control point for the entire global network. In this paper, we proposed a way to extend SDN network by multi-controller which is based on Fibonacci heap originating from Fibonacci series (a recursive series). And we also simulate the optimized Dijkstra's algorithm based on Fibonacci heap considering the node weight for a graph derived from SDN topology. We also compare it with original Dijkstra's algorithm(OSPF used) and results indicate the optimized Dijkstra's algorithm outperforms the original one when the network is dense.

**Keywords:** SDN; Network Virtualization; Extend; Dijkstra's algorithm; Fibonacci Heap

## I. INTRODUCTION

Since traditional networking technologies have many limitations, such as complexity that leads to stasis, inconsistent policies, inability to extend and vendor dependence [1], Software Defined Networking is an emerging network which transforms networking architecture. Over the past few years, SDN and Network Functions Virtualization (NFV) have become the hottest topics in networking.

According to [2], SDN is based on three architectural principles and the definition of open interfaces, as follows:

- a. Decoupling of traffic forwarding and processing from control
- b. Logically centralized control
- c. Programmability of network services
- d. Open interfaces

Based on these 4 basic principles, SDN can greatly reduce the fragility of the network and the likelihood of configuration mistakes that can compromise security[3]. It also can be applied to different networking environments but with customized solutions and enable the research community and third-party software developers to contribute to the creation of algorithms [4].

Currently, many researchers and enterprises pay their attention to study how to scale or extend SDN network. [5] proposed a framework for recursive abstraction of SDN controlplane for large-scale production networks which is named FRACTAL. In FRACTAL, a large network is divided into multiple small networks, each of which is abstracted as a single virtual switch. [6] proposed a comprehensive security

architecture for SDN to provide security services such as enforcing mandatory network policy correctly and receiving network policy securely for SDN in order to solve these common security issues and new security challenges, and it can also help the developers to implement security functions to provide security services when developing the SDN controller. [7] proposed an abstract model of SDN architectures enabling comprehensive performance comparisons and derived variations in terms of composition unit (single or multiple), processing principle (sequential or parallel), or location (intraor inter-node). [8] proposed a SDN virtualization architecture with flexible hypervisor function allocation which is called HyperFlex. HyperFlex provides control-plane virtualization by adding a control-plane isolation function, either in software or on network elements. The isolation function ensures that the resources of the control-plane are shared correctly between each virtual SDN network while it also protects the hypervisor resources from resource exhaustion. [9] proposed a recursive approach for scaling SDN using Fibonacci heap. [10] proposed a recursive abstraction of SDN control plane for large-scale production networks which was called RAON. However, they all focused on SDN architecture without the corresponding routing algorithm.

In this paper, we proposed a way to extend SDN network by multi-controller which is based on Fibonacci heap originating from Fibonacci series (a recursive series). And we also simulate the optimized Dijkstra's algorithm based on Fibonacci heap. The algorithm considers the node weight for a graph derived from SDN topology. We use Eclipse to implement the optimized Dijkstra's algorithm, so that we can embed it to OpenDayLight in the future work. And then we also compare the time complexity with the original Dijkstra's algorithm. As shown by the comparisons, the optimized Dijkstra's algorithm outperforms the original algorithm.

The remainder of this paper is organized as follows: section 2 discusses SDN's architecture, OpenFlow, Fibonacci heap and how to extend SDN network; section 3 describes the optimized Dijkstra's shortest path algorithm and evaluation result and its analysis; section 4 indicates the conclusion and our future work.

## II. EXTEND SDN NETWORK

### A. Software Defined Network

1) *Basic Model:* Figure 1 shows the basic model of SDN. The service consumer (client, user, customer) include

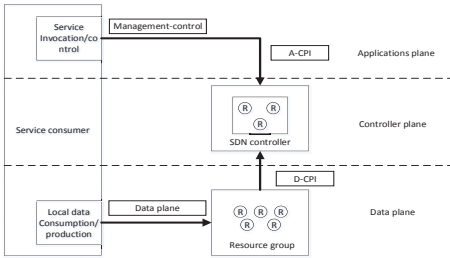


Fig. 1. Basic model

two parts, one is service invocation/control which exchanges management-control operations with SDN service provider through application-controller plane interface A-CPI, and another is local data consumption/production which exchanges data with some set of resources in Resource group. The resources that are owned by some SDN server or provider forward or process user data ultimately, but service consumer controls user data's service via a session contained in a management-control association by invoking actions on a set of virtual resources (resources in SDN controller) through Data-controller plane interface D-CPI. In controller plane, the SDN controller virtualizes and orchestrates the virtual resource.

2) *OpenFlow*: OpenFlow provides an open protocol to program the flow table in different switches and routers [11]. In the newest OpenFlow1.5 [12], a flow table consists of flow entries including match fields, priority, counters, instructions, timeouts, cookie, flags. An OpenFlow switch consists of Flow Tables, Group Table, Meter Table and so on, as shown in Fig.2

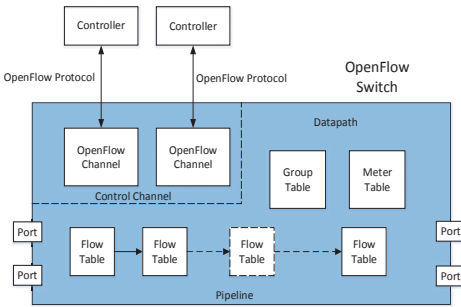


Fig. 2. Main components of an OpenFlow switch

Once receiving a package, an OpenFlow Switch performs some functions. The switch starts by performing a table lookup in the first flow table, and it also may perform table lookups in other flow tables through pipeline processing. Then

match fields are analyzed to see if the packet matches to the flow table entry. If yes, the action set in the table will be executed, otherwise it will be uploaded to the controller through OpenFlow channel.

*B. Fibonacci Heap*

1) *Fibonacci Series*: Fibonacci heap [13] which is proposed by Michael L. Fredman and Robert E. Tarjan comes from Fibonacci series. Fibonacci series is a naturally series which can be expressed in the following integer sequence:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233 \dots \dots \quad (1)$$

In mathematical terms, the sequence  $F_n$  of Fibonacci series is defined by the recursive relation, as expressed in this equation:

$$\begin{cases} F_1 = 1 \\ F_n = F_{n-1} + F_{n-2} \quad (n \geq 3) \end{cases} \quad (2)$$

2) *Definitions*: Before introducing Fibonacci heap, there are some definitions we need to acquire, as follows:

- $x$ : a node in a graph;
- $p(x)$ : node  $x$ 's parent;
- $H.min$ : a pointer points to the root of the tree containing a minimum key;
- $x.p$ : a pointer points to  $x$ 's parent node;
- $x.child$ : a pointer points to  $x$ 's one child;
- $x.left$ : a pointer points to  $x$ 's left sibling;
- $x.right$ : a pointer points to  $x$ 's right sibling;
- $x.rank$ : the number of  $x$ 's children;
- $x.mark$ : indicating whether  $x$  loses any child when  $x$  becomes other node's child. If yes,  $x.mark$  is marked to be *TRUE*, otherwise *FALSE*;

$H.n$ : the number of nodes currently in the heap;  
 $D(H.n)$ : the upper bound of any node's rank.

3) *Fibonacci Heap*: A Fibonacci heap is a collection of rooted tree which has min-heap ordered [14], and it means each tree has min-heap property: if  $x$  is any node in the tree, then the key of the item in  $x$  is no less than its parent  $p(x)$ .

Figure 3 shows an example of Fibonacci heap, (a) is a Fibonacci heap includes 5 min-heap ordered trees and 14 nodes. The dashed line indicates root list. The minimum node in the heap is the one include key 3, as the  $H.min$  points to. As is shown in (b), each node  $x$  includes a pointer  $x.p$  and a pointer  $x.child$ . All children of  $x$  are doubly linked together into a circular which is called child list of  $x$ . Each child  $y$  in child list has a pointer  $y.left$  and a pointer  $y.right$  respectively. If  $x$  has just one child  $y$ , then  $y.left = y.right = y$ . A given Fibonacci heap  $H$  is accessed by  $H.min$  pointing to the minimum node of the Fibonacci heap. If there are more than one root containing a minimum key, anyone can be the minimum key. If a Fibonacci heap  $H$  is empty, then  $H.min = NIL$ . The roots of all the trees in a Fibonacci heap are linked together using their left and right pointers into a circular, doubly linked list called the root list of the Fibonacci heap. The order of the trees within a root list is arbitrary.

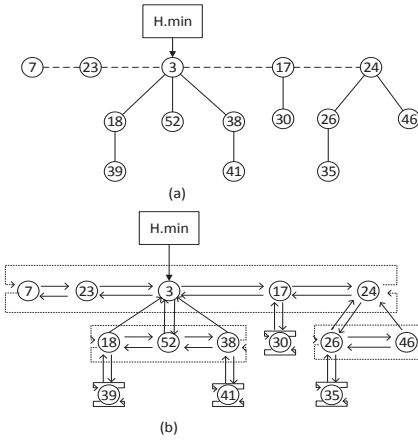


Fig. 3. An example of Fibonacci heap

### C. How to extend SDN network

Fibonacci Heap has a method *Fibonacci - Heap - Union*( $H1, H2$ ) to combine two heaps into one. The method is shown as follows:

---

#### Algorithm 1 Fibonacci-Heap-Union

---

**Require:**  $H1, H2$

**Ensure:**  $H$

```

1: function FIBONACCI-HEAP-UNION( $H1, H2$ )
2:   create  $H$ ;
3:    $H.min = H1.min$ ;
4:   concatenate the root list of  $H2$  with the root list of  $H$ ;
5:   if ( $H1.min = NIL$ ) or ( $H2.min \neq NIL$  and
6:      $H2.min.key < H1.min.key$ ) then
7:      $H.min = H2.min$ ;
8:   end if
9:    $H.n = H1.n + H2.n$ ;
10:  return  $H$ 
11: end function
    
```

---

Fig.4 shows how to combine two heaps into one. In the real network, routers correspond to nodes in the heap and controllers correspond to  $H.min$ . That means every slice of the network has a controller to manage them, then the network can be extend.

## III. EVALUATION

### A. Algorithm

1) *An Optimized Dijkstra's Algorithm*: It is known that Dijkstra's algorithm solving the single-source shortest path problem for a graph with non-negative length edges [15].

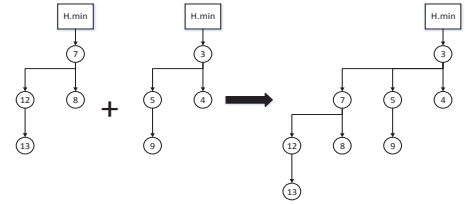


Fig. 4. Combine two heaps into one

Given a directed, edge-weighted graph  $G = (V, E, W)$ , where  $V$  is the set of nodes,  $E$  is the set of edges and  $W$  is one edge's weight which is non-negative, the original Dijkstra's algorithm will return a shortest path from the source node  $S$  to each other node in the graph.

---

#### Algorithm 2 The optimized Dijkstra's algorithm

---

**Require:**  $G = (V, E, W), S$

**Ensure:** *ShortestPath*

```

1: function OPTIMIZEDDIJKSTRA ( $G = (V, E, W), S$ )
2:   create  $H$  and  $H.min = 0$ ;
3:    $d[s] \leftarrow 0$ ;
4:   for each  $u \neq s, u \in V$  do
5:      $D[u] \leftarrow \text{infinity}$ ;
6:     Insert( $u$ );
7:   end for
8:   while  $H$  is not an empty set do
9:      $u \leftarrow \text{ExtractMin}(H)$ ;
10:    for each  $v$  adjacent to  $u$  do
11:      if  $d[v] > d[u] + w(u, v)$  then
12:         $d[v] \leftarrow d[u] + w(u, v)$ ;
13:        previous[ $v$ ]  $\leftarrow u$ ;
14:      end if
15:    end for
16:  end while
17:  return ShortestPath
18: end function
    
```

---

The algorithm 1 shows the optimized Dijkstra's algorithm based on Fibonacci heap, where *previous*[ $v$ ] means the previous node of one node in the shortest path tree. If we find each node's previous node, we can get the shortest path. Line 2 create a empty Fibonacci heap to insert each  $u \neq s, u \in V$ , the function *Insert*( $u$ ) is shown as algorithm 2, and the function *ExtractMin*( $H$ ) is shown as algorithm 3.

Algorithm 2 shows how to inset a node  $x$  into a Fibonacci heap  $H$  when presuming that  $x$  has been assigned and the value of  $x.key$  has been given.

Algorithm 3 shows how to extract  $H.min$ , which is the most complex part of the whole algorithm. We also presume that the pointer remains in the root list must be updated, but the pointer of the node extracted from a root list will not be changed. The function *Consolidate*( $H$ ) is shown in algorithm

---

**Algorithm 3** Fibonacci-Heap-Insert

---

**Require:**  $H, x$

```

1: function INSERT( $H, x$ )
2:    $x.rank \leftarrow 0$ 
3:    $p(x) \leftarrow NIL$ 
4:    $x.child \leftarrow NIL$ 
5:    $x.left \leftarrow x$ 
6:    $x.right \leftarrow x$ 
7:    $x.mark \leftarrow FALSE$ 
8:   concatenate the root list containing  $x$  with root list  $H$ 
9:   if  $H.min = NIL$  or  $H.min.key > x.key$  then
10:     $H.min \leftarrow x$ 
11:   end if
12:    $H.n \leftarrow H.n + 1$ 
13: end function

```

---

**Algorithm 4** Fibonacci-Heap-ExtractMin

---

**Require:**  $H$

```

1: function EXTRACTMIN( $H$ )
2:    $z \leftarrow H.min$ 
3:   if  $z \neq NIL$  then
4:     for each child  $x$  of  $z$  do
5:       add  $x$  to the root list of  $H$ 
6:        $p(x) \leftarrow NIL$ 
7:     end for
8:     remove  $z$  from the root list of  $H$ 
9:     if  $z = z.right$ 
10:       $H.min \leftarrow NIL$ 
11:     else
12:       $H.min \leftarrow z.right$ 
13:       $Consolidate(H)$ 
14:     end if
15:      $H.n \leftarrow H.n - 1$ 
16:   end if
17: end function

```

---



---

**Algorithm 5** Consolidate

---

**Require:**  $H$

```

1: function CONSOLIDATE( $H$ )
2:   for  $i = 0$  to  $D(H.n)$  do
3:      $A[i] \leftarrow NIL$ 
4:   end for
5:   for each node  $z$  in the root list of  $H$  do
6:      $x \leftarrow z$ 
7:      $r \leftarrow x.rank$ 
8:     while  $A[r] \neq NIL$  do
9:        $y \leftarrow A[r]$ 
10:      if  $y.key < x.key$  then
11:        exchange  $y$  and  $x$ 
12:      end if
13:      remove  $y$  from the root list of  $H$ 
14:      make  $y$  a child of  $x$ , incrementing  $x.rank$ 
15:       $y.mark \leftarrow FALSE$ 
16:       $A[r] \leftarrow NIL$ 
17:       $r \leftarrow r + 1$ 
18:     end while
19:      $A[r] \leftarrow x$ 
20:   end for
21:    $H.min \leftarrow NIL$ 
22:   for  $i = 0$  to  $D(H.n)$  do
23:     if  $A[i] \neq NIL$  then
24:       add  $A[i]$  to the root list of  $H$ 
25:     end if
26:     if  $H.min = NIL$  or  $A[i].key < H.min.key$  then
27:        $H.min \leftarrow A[i]$ 
28:     end if
29:   end for
30: end function

```

---

4.

Algorithm 4 shows how to consolidate the nodes whose rank is the same in the  $H$ 's root list. After consolidated, each rank has at most one root in the root list.

*B. Simulation*

In this simulation, because operation time algorithm costs is a key point, we choose time as a criterion to compare the original Dijkstra's algorithm with the optimized Dijkstra's algorithm based on Fibonacci heap. It indicates how long it takes to find the shortest path from the source node to the other nodes in the graph. Both algorithms were run in Eclipse(Mars) on server, so that we can present a natural extension of our method into OpenDayLight [16] in the future work. The datasheet of the server is as follow: 3.2G of quad-core processor; 8GB memory; Windows 7 of OS.

We generate the random topology whose edge weights are created randomly from a distribution [1, 1000], and the random graph of  $n$  vertices has also randomly decided topology. The

random graph is generated in the following steps. Firstly, two parameters are input: one is the number of vertices  $n$ , the other is density  $d$  whose value is between 0 and 1. Then, two nodes are chosen randomly, and linked as an edge if their index are different. And then we give a weight from 1 to 1000 randomly. At last, we keep generating random edges until all vertices are connected. As a result, we get the random graph used to test operation time of the original Dijkstra's algorithm to the optimized Dijkstra's algorithm. It should note that a certain node (eg.node 5) is chosen to be the single source node, and we compute the time of finding the shortest path from the single source node to every other node in the topology, like multi-broadcast in the real network.

The simulation results are shown from Figure 4 to 6. In each figure, X-axis represents the number of nodes in each graph, while Y-axis indicates the operation time(millisecons). Blue lines are results using original Dijkstra's algorithm, and red lines are results implementing optimized Dijkstra's algorithm.

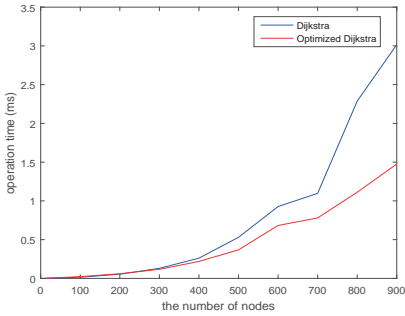


Fig. 5. The comparison of operation time when density is 0.3

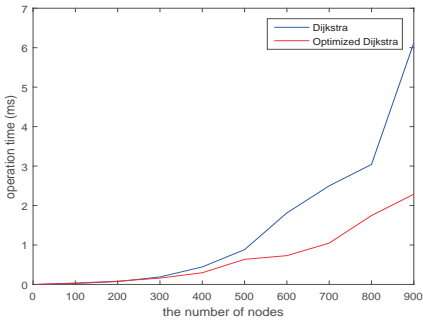


Fig. 6. The comparison of operation time when density is 0.6

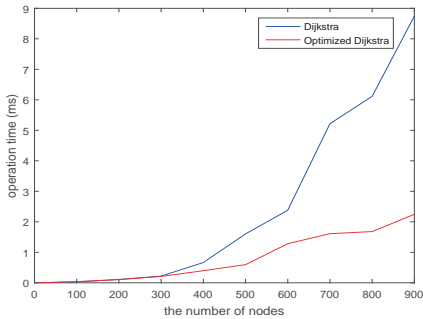


Fig. 7. The comparison of operation time when density is 0.9

### C. Analysis

In figure 4, 5 and 6, we can come a conclusion that optimized Dijkstra’s algorithm perform better in the cost of time. Corresponding to the real network, the nodes represent routers which need to communicate with other ones. Thus, as is indicated in the figures, with the increasing in density of nodes, the optimized Dijkstra’s algorithm will perform better. In fact, the complexity of algorithms determines their different performance. Let  $V$  be number of vertices, and  $E$  represents the amount of edges. The complexity of original Dijkstra’s algorithm can be expressed as  $O(V^2 + E)$ , while the complexity of optimized Dijkstra’s algorithm can be expressed as  $O(V \lg V + E)$

## IV. CONCLUSIONS

In this paper, we proposed a way to extend SDN network by multi-controller which is based on Fibonacci heap. Next we use Eclipse to implement the optimized Dijkstra’s algorithm which can be applied in the extended network, and compare the running time with the original Dijkstra’s algorithm. As the simulation shown, the optimized Dijkstra’s algorithm costs less time than original one. When the graph is dense, the optimized Dijkstra’s algorithm outperforms the original one. In the future work, we will present a natural extension of our method into OpenDayLight or Beacon[17] and simulate it in a real SDN network.

## V. ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (61671077 and 61671264).

## REFERENCES

- [1] O. N. Fundation, “Software-defined networking: The new norm for networks,” *ONF White Paper*, 2012.
- [2] E. Haleplidis, K. Pentikousis, S. Denazis, J. H. Salim, D. Meyer, and O. Koufopavlou, “Software-defined networking (sdn): Layers and architecture terminology,” tech. rep., 2015.
- [3] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, et al., “B4: Experience with a globally-deployed software defined wan,” in *ACM SIGCOMM Computer Communication Review*, vol. 43, pp. 3–14, ACM, 2013.
- [4] A. Greenberg, G. Hjalmtysson, D. A. Maltz, A. Myers, J. Rexford, G. Xie, H. Yan, J. Zhan, and H. Zhang, “A clean slate 4d approach to network control and management,” *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 5, pp. 41–54, 2005.
- [5] M. Kwak, J. Suh, and T. Kwon, “Fractal: A framework for recursive abstraction of sdn control-plane for large-scale production networks,” in *Network Softwareization (NetSoft), 2015 1st IEEE Conference on*, pp. 1–5, IEEE, 2015.
- [6] Z. Hu, M. Wang, X. Yan, Y. Yin, and Z. Luo, “A comprehensive security architecture for sdn,” in *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on*, pp. 30–37, IEEE, 2015.
- [7] T. Sato, S. Ata, I. Oka, and Y. Sato, “Abstract model of sdn architectures enabling comprehensive performance comparisons,” in *Network and Service Management (CNSM), 2015 11th International Conference on*, pp. 99–107, IEEE, 2015.



- [8] A. Blenk, A. Basta, and W. Kellerer, "Hyperflex: An sdn virtualization architecture with flexible hypervisor function allocation," in *Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on*, pp. 397–405, IEEE, 2015.
- [9] W. Dai, G. Shou, Y. Hu, and Z. Guo, "R-sdn: A recursive approach for scaling sdn," in *Information and Communications Technologies (ICT 2014), 2014 International Conference on*, pp. 1–6, IET, 2014.
- [10] M. Kwak, H. Lee, J. Shin, J. Suh, and T. Kwon, "Raon: A recursive abstraction of sdn control-plane for large-scale production networks," in *Information Networking (ICOIN), 2015 International Conference on*, pp. 426–427, IEEE, 2015.
- [11] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.
- [12] O. S. Specification, "Version 1.5. 1," *Open Networking Foundation*, 2015.
- [13] M. L. Fredman and R. E. Tarjan, "Fibonacci heaps and their uses in improved network optimization algorithms," *Journal of the ACM (JACM)*, vol. 34, no. 3, pp. 596–615, 1987.
- [14] T. H. Cormen, *Introduction to algorithms*. MIT press, 2009.
- [15] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische matematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [16] J. Medved, R. Varga, A. Tkacik, and K. Gray, "Opendaylight: Towards a model-driven sdn controller architecture," in *2014 IEEE 15th International Symposium on*, pp. 1–6, IEEE, 2014.
- [17] D. Erickson, "The beacon openflow controller," in *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*, pp. 13–18, ACM, 2013.

# In-Service Assessment of Mobile Services QoE from Network Parameters

Sibila Isak-Zatega  
BH Telecom  
Sarajevo, Bosnia and Herzegovina  
sibila.isakzatega@bhtelecom.ba

Vlatko Lipovac  
University of Dubrovnik  
Dubrovnik, Croatia  
vlatko.lipovac@unidu.hr

**Abstract**— In this paper, we analyze the relationship between the achieved Quality of Experience (QoE) of the web browsing network service and in-service network parameters obtained by passive monitoring. With this regard, we conducted experiments where the user QoE and in-service parameters are measured simultaneously, where the latter are derived from traffic data captured by passive probe on the Gn interface of the mobile network, and correlated to detailed records of HTTP and TCP protocols. Based on the subjective evaluations by the participants in the experiment about the web browsing service quality, it was evident that there was strong relationship between the QoE and specific network parameters such as e.g. *Average Time to get first data* and *Average Time to connect*. However, no such relationship was found between the *Cancellation rate* parameter and QoE, as the hypothesis that there was no relationship between these two parameters, could not be rejected.

**Keywords**— QoE; network parameters, in-service monitoring

## I. INTRODUCTION

The last decade was marked by the explosive growth of mobile data traffic. It is estimated that in 2015 global mobile data traffic increased by 74 percent, and at the end of the year has reached about 3.7 EB (exabytes) on a monthly basis. According to the latest forecasts for the period 2015-2020 the overall mobile data traffic is expected to grow to 30.6 EB per month by the end of 2020 [1]. Video streaming and social networking services will dominate mobile data traffic by 2019, followed by web browsing [2]. With increased use of smart mobile devices that can support high-speed data transfer and larger demand for mobile broadband, there is an increase in user expectations regarding the quality of services. The Quality of Experience (QoE) has become one of the most important factors for the customer when choosing an operator. Consequently, user satisfaction is becoming one of the most important topics considered by telecom operators, who therefore are very determined to find an appropriate way of assessing the level of customer satisfaction in using the services. The QoE is subjective in nature, but it is very important to find a link to technical quality parameters that are measurable and can be monitored in the operator environment, allowing to estimate the QoE by objective methods.

Monitoring web browsing QoE is complex and was considered in many conducted researches such as e.g. [3] - [8]. Particularly, QoE assessment from measured network parameters can reduce the time needed to eliminate the causes of poor user experience and may impact the overall customer satisfaction.

The main objective of this research is to explore the possibility of estimating the web browsing QoE of mobile network by measuring in-service parameters derived from the traffic data collected on interfaces in the mobile core network. The parameters are defined in the performance management system with passive probes. The main research questions are:

- To what extent is there a relationship between parameter *Web page download time* and customer satisfaction?
- To what extent is there a relationship between the in-service parameters defined from the traffic data and user satisfaction?

The method we used here is empirical research, i.e. based on the experiments that we conducted under live network conditions, collecting and analyzing the appropriate real-life data. In Section II, we describe how the experiment was performed and the tools we used for conducting it. In Section III, we describe the chosen in-service parameters while in Section IV, we present the test results and the analysis. Conclusions are drawn in section V.

## II. TEST SETUP AND ANALYSIS TOOLS

### A. Participants

The experiment was conducted by ten users, five female and five male. The age of participants ranged between twelve and forty-five years. All participants use the Internet at least one hour a day and usually via WiFi. They used the mobile internet access only if WiFi was not available.

### B. Test Setup

The test setup is presented in Fig. 1.

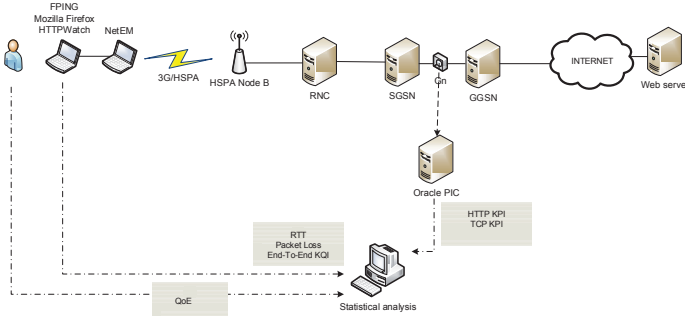


Fig. 1. Test setup for web browsing

The participants in the experiments used a client which was operating on Windows 8, with Intel (R) Core (TM) i5-3210M CPU @ 2.5GHz x 2 processor, 4 GB RAM. For the purpose of measuring *Web page download time* parameter, a software tool called HttpWatch was installed on the client device. The client device was connected to the gateway via Ethernet full duplex link, the bandwidth of 100 Mbps. The gateway ran on Ubuntu 14.04 LTS OS (Linux OS) with an Intel Core 2 Duo T7500 CPU @ 2.2GHz x 2, 2 GB RAM-a and was connected via USB modem to the 2G/3G network.

We used the NetEm network emulator on Ubuntu on the gateway to change network conditions by adding delay and packet loss. The Huawei E3272 LTE USB modem was used for testing, managing it by the embedded software Connection Manager, which allowed setting the preferred access network.

We enabled the 3G/HSPA to be the preferable access network in the experiment. The client system was connected to the internet through the mobile network via the gateway. In both laptops, automatic software updates were disabled. The participants in the experiment used Mozilla Firefox 35.0.1 web browser. HTTP and TCP extended Detailed Records (xDR) from data captured on Gn interface were available using ProTrace application on the Oracle PIC platform. Using ProTraq application on this platform, we defined and activated new statistical sessions which generated the in-service parameters' values aggregated over 5-minute intervals. The parameters were defined from HTTP and TCP xDR's for MSISDN of test SIM card.

To investigate the relationship between the QoS and in-service parameters, we activated two test scenarios. Each participant tested web browsing six times under different network conditions, which were changed using Netem to add delay or packet loss during the experiment.

In one test, a participant tested web browsing according to two scenarios, which we later considered to be two paired observations. The duration of both scenarios was limited to 5 minutes. In the first test scenario the participant was asked to open a web browser and download three predefined web pages, specifically *sportspost.ba*, *klix.ba* and *vijesti.ba*.

After opening the web browser HttpWatch automatically started to record the log file. After testing, we disconnected the USB modem from the mobile network and also emptied the cache using the appropriate option available in HttpWatch. The participant was asked to give his opinion about the quality of web browsing for each of the three web pages, and the overall opinion as well. After that, the values of defined in-service parameters were collected by running statistics sessions in Oracle PIC platform. In the second test scenario, all steps were carried out as in the first test scenario, with the only difference that participants accessed web pages chosen by themselves, and provided only the overall opinion. The parameter *Web page download time* was not considered, and the quality rating was according to the ordinary 5-point Likert-scale (Excellent; Good; Fair; Poor; Bad).

### C. Test tools

We used the Oracle Performance Intelligence Center (PIC) as a monitoring and data gathering system that helps service providers to manage their assets, encompassing network performance, quality of service (QoS), and customer analysis [9]. The PIC uses passive probes to capture traffic data and forward Probe Data Units (PDU) to the integrated xDR Platform (IXP). The IXP stores these traffic data and correlates them into detailed records. PIC provides applications that mine the detailed records to provide value-added services such as network performance analysis, call tracing and reporting [10]. For the purpose of this research, we used HTTP and TCP sessions defined on the Gn interface of the mobile network. Parameters and statistic sessions were defined by using the ProTraq application.

Furthermore, we used the HttpWatch as a software plugin that integrates with Internet Explorer and Mozilla Firefox browsers to decode HTTP and HTTPS traffic that is generated when a web page is accessed. In addition, it shows interactions between the browser and its cache. Each HTTP transaction can be examined to decode headers, cookies, query strings and other HTTP related data. HttpWatch has two components: a plug-in used to collect, view and save HTTP traffic within IE or Firefox, and a standalone log file viewer known as

HttpWatch Studio [11]. The user interface allows the use of different commands to control the HttpWatch, such as e.g. emptying the browser cache. The HttpWatch groups together requests within a heading for each page making it much easier to understand multi-page steps, e.g. log in, search and update pages. For the entire page, the HttpWatch shows the time when the browser first starts to display a downloaded page, when the page's DOM (Document Object Model) is loaded, when both the DOM and images have been loaded and the time for completion of all HTTP or HTTPS requests made by the page (HTTP Load parameter). In our experiment, we used the HttpWatch 9.4 Basic Edition and measured the *HTTP Load* parameter, renaming it *Web page download time*.

Finally, we used the NetEm as an enhancement of the Linux traffic control facilities that allow adding delay, loss, duplication and other characteristics as well to packets outgoing from a selected network interface. NetEm is built using the existing QoS and the Differentiated Services (DiffServ) facilities in the Linux kernel [12].

III. IN-SERVICE PARAMETERS

In this section, we analyze the fields of TDR 's collected from HTTP and TCP sessions on the Gn interface and describe which fields we used to define in-service parameters in Oracle PIC. Figures 2 and 3 show examples of how different times available in TCP and HTTP data record fields are measured, while Table 1 contains the parameters we defined in the Oracle PIC platform. The HTTP parameter *Success rate* is the ratio of the count of successful HTTP transactions (request/response) to the total count of HTTP requests. The *Cancellation rate* represents the ratio of the count of HTTP transactions with the status code *OK* but with incomplete transferred data, to the count of transactions that are either successful or with incomplete data regardless of the status code. Downlink/Uplink throughput is measured over 30-seconds interval, and the *Maximum Downlink/Uplink Throughput* parameters are the corresponding maximal values within the observed 5-minute intervals. The TCP *Success rate* parameter is the ratio of successfully established TCP connections to the total count of attempts to establish a TCP connection. The *Average Downlink/Uplink Throughput* are average values for the established connections, while the *Average Downlink/Uplink Maximum RTT* are the average values of maximal RTT for all TCP connections within a 5-minute test interval.

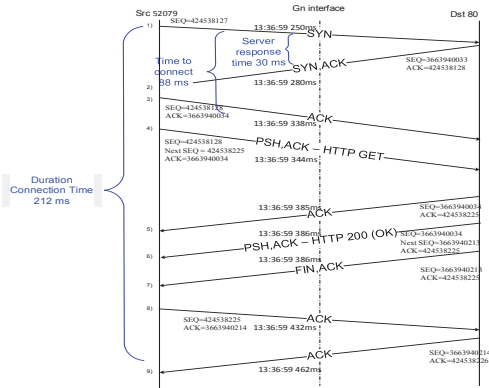


Fig. 2. Characteristic times for TCP record

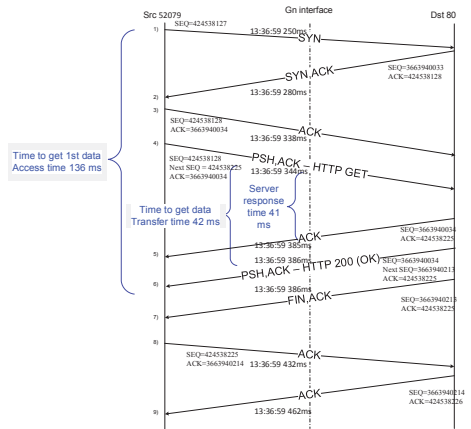


Fig. 3 Characteristic times for HTTP record

TABLE I. IN-SERVICE PARAMETERS

In-service parameters	
HTTP	TCP
<i>Success rate</i>	<i>Success rate</i>
<i>Cancellation rate</i>	<i>Average Server Response Time</i>
<i>Average Server Response Time</i>	<i>Average Time to connect</i>
<i>Average Time to get first data</i>	<i>Average Downlink Throughput</i>
<i>Maximum Downlink Throughput</i>	<i>Average Uplink Throughput</i>
<i>Maximum Uplink Throughput</i>	<i>Retransmission Downlink Packet ratio</i>
<i>Average Time to get data</i>	<i>Retransmission Uplink Packet ratio</i>
<i>Average Transaction Time</i>	<i>Average Downlink Maximum RTT</i>
<i>Average Transfer Time</i>	<i>Average Uplink Maximum RTT</i>
	<i>Average Duration Connection</i>

IV. TEST RESULTS AND ANALYSIS

Before analyzing the test results, let us review the statistical tests that we used to explore data. The collected data for *Web page download time* and in-service parameters are continuous, while the rating of user satisfaction (QoE), which was measured according to the usual 5-point Likert scale (Excellent; Good; Fair; Poor; Bad), is ordinal data. Numerous previous studies of QoE used parametric statistics to analyze user opinion scores (ordinal data) by means of the numerical scale of 1-5 instead of the Likert scale. However, the parametric tests are generally not appropriate for ordinal data because some crucial conditions for valid statistical analysis have not yet been verified [8], [9].

Therefore, in this research, we preferred to use non-parametric statistics which provides the relevant analysis for the subjective scores rated by the participants in the experiment.

To find out whether there is a relationship between two variables we used the Spearman's rank-order correlation (often abbreviated to Spearman's correlation), which calculates a correlation coefficient ( $\rho$ ) that is a measure of the strength and direction of the association/relationship between two continuous or ordinal random variables. We used the Wilcoxon signed-rank test to determine whether there is a median difference between paired or matched observations. This test can be considered as the nonparametric equivalent to the paired-samples t-test. The participants in our experiments used web browsing according to two scenarios, and we aimed to determine whether there are differences in their opinion score under the same network conditions.

A. Results of the first test scenario

1) Relationship between customer satisfaction and parameter "Web page download time."

To evaluate the association between customer satisfaction and the Web page download time, we used the nonparametric Spearman's rank correlation test, as all assumptions that required for Spearman's correlation could be considered applicable. QoE, which is an ordinal variable and Web page download time, which is a continuous variable, represent paired observations. The preliminary analysis showed that the relationship between QoE and the Web page download time was monotonic, as it is obvious from the scatter plot in Fig. 4. We state the null and the alternative hypotheses as it follows, respectively:

$H_0$ : There is no association between QoE and Web page download time.

$H_1$ : There is an association between QoE and Web page download time.

The results of Spearman's rank correlation test are presented in Table II. Apparently, there is a strong negative correlation between the customer satisfaction after loading each of the tested web pages and the page download duration. The correlation coefficients and the statistical significance ( $p$ -value) are:  $\rho = -0.834, p < 0.0005$  for *sport.sport.ba*,  $\rho = -0.854, p < 0.0005$  for *klix.ba*, and  $\rho = -0.739, p < 0.0005$  for *vijesti.ba*. We repeated the test for the overall QoE and the maximum time to load a page (after loading all three test pages), as well as for the overall QoE and the average time to load a page. The results of Spearman's rank correlation test are presented in Table II and Table III.

TABLE II. CORRELATION - THREE WEB PAGES

Spearman's Correlation Coefficient	Web page download time		
	<i>sport.sport.ba</i>	<i>klix.ba</i>	<i>Vijesti.ba</i>
QoE	-0.834	-0.854	-0.739

TABLE III. CORRELATION - MAX AND AVG TIME

Spearman's Correlation Coefficient	Web page download time	
	Maximum Web page download time	Average Web page download time
QoE	-0.834	-0.846

Obviously, strong negative correlation between the customer satisfaction and the maximum and average time to load pages was found by the test (with three test pages loaded), with the correlation coefficient values  $\rho = -0.834, p < 0.0005$  and  $\rho = -0.846, p < 0.0005$ , respectively. Here the value of  $p < 0.0005$  means that the correlation coefficient is significantly different from zero, and the null hypothesis can be rejected.

2) Relationship between the customer satisfaction and in-service parameters

One of the basic research questions is whether there is a statistically significant relationship between the customer satisfaction and in-service measured values of quality parameters defined from the traffic data on mobile network Gn interface?

We addressed this by testing the null hypothesis and the alternative one as it follows, respectively:

$H_0$ : There is no association between the in-service parameter and customer satisfaction;

$H_1$ : There is an association between the in-service parameter and customer satisfaction

The first two assumptions for using Spearman rank correlation test are met because we have one continuous and one ordinal variable representing the paired observations.

We derive our conclusions about the (non-)existence of monotonic relationship between the variables in question by visual examination of the scatter plots generated for each of the selected in-service parameters. In this case, however, no monotonic relationship between the QoE and HTTP parameter *Average Time to get data*, as well as between the QoE and the TCP parameters: *Average Server Response Time* and *Average Uplink Maximum RTT* could be confirmed. Regardless of that, we also run the Spearman correlation test for these parameters and found that it gives inconsistent results or correlation which is not significant. The results are shown in Tables IV and V.

B. Results of the second test scenario

To qualify the association between the customer satisfaction and in-service parameters using data related to the second scenario, we used the nonparametric Spearman's rank correlation test, carrying out the steps described in subsection A.2. The results are presented in Tables IV and V.

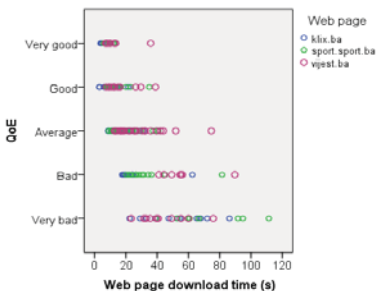


Fig. 4 Scatter plot of QoE vs. Web page download time

TABLE IV. CORRELATION QoE AND HTTP PARAMETERS

Spearman's Correlation Coefficient	QoE	
	First test scenario	Second test scenario
Success rate	0.422 <sup>b</sup>	0.168
Cancellation rate	-0.499 <sup>b</sup>	-0.320 <sup>a</sup>
Average Server Response Time	-0.116	-0.114
Average Time to get first data	-0.770 <sup>b</sup>	-0.790 <sup>b</sup>
Maximum Downlink Throughput	0.760 <sup>b</sup>	0.616 <sup>b</sup>
Maximum Uplink Throughput	0.579 <sup>b</sup>	0.575 <sup>b</sup>
Average Time to get data	-0.106	0.034
Average Transaction Time	-0.614 <sup>b</sup>	-0.510 <sup>b</sup>
Average Transfer Time	-0.688 <sup>b</sup>	-0.491 <sup>b</sup>

<sup>a</sup> Correlation is significant at the 0.05 level (2-tailed)

<sup>b</sup> Correlation is significant at the 0.01 level (2-tailed)

TABLE V. CORRELATION QoE AND TCP PARAMETERS

Spearman's Correlation Coefficient	QoE	
	First test scenario	Second test scenario
Success rate	0.269 <sup>c</sup>	0.239
Average Server Response Time	0.079	0.066
Average Time to connect	-0.791 <sup>d</sup>	-0.831 <sup>d</sup>
Average Downlink Throughput	0.761 <sup>d</sup>	0.482 <sup>d</sup>
Average Uplink Throughput	0.763 <sup>d</sup>	0.652 <sup>d</sup>
Retransmission Downlink Packet ratio	-0.805 <sup>d</sup>	-0.304 <sup>c</sup>
Retransmission Uplink Packet ratio	-0.766 <sup>d</sup>	-0.133
Average Downlink Maximum RTT	-0.678 <sup>d</sup>	-0.506 <sup>d</sup>
Average Uplink Maximum RTT	-0.054	0.129
Average Duration Connection	-0.568 <sup>d</sup>	-0.150

<sup>c</sup> Correlation is significant at the 0.05 level (2-tailed)

<sup>d</sup> Correlation is significant at the 0.01 level (2-tailed)

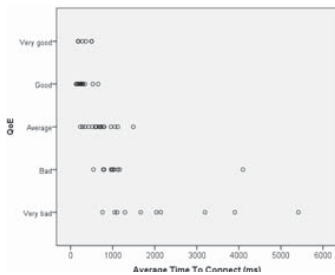


Fig. 6. Scatter plot of QoE vs. Average Time to Connect

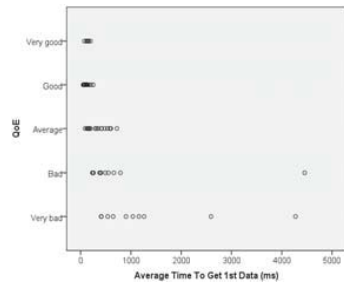


Fig. 7. Scatter plot of QoE vs. Average Time To Get 1st Data

C. Difference between the first and the second scenario QoE

It should be noted that testing web browsing according to the second test scenario was carried out immediately after testing according to the first test scenario, and under the same or similar network conditions. This has imposed an additional question in the research as it follows:

“Is there a statistically significant difference between the customer satisfaction when using web browsing according to the first test scenario and second test scenario under the same network conditions?”

As the assessment of customer satisfaction after testing web browsing according to the first test scenario and to the second one, can be considered as evaluating paired observations, we applied the Wilcoxon signed-rank test. Since customer satisfaction for web browsing is associated with loading time of web pages which, apart from the quality of the network, depends on the characteristics of the page (size, optimization, target servers etc.), we expected customer satisfaction to be significantly different between the first and the second scenario. So, let us set the null and the alternative hypotheses accordingly:

H0: The customer satisfaction rating is equal for both test scenarios

H1: The rating of customer satisfaction is not for both test scenarios

The necessary assumption for applying the Wilcoxon signed-rank test is that the distribution of the differences between the two related groups are symmetrical in shape, so we tested the null hypothesis that the distribution is centered around zero, by means of the histogram in Fig. 8.

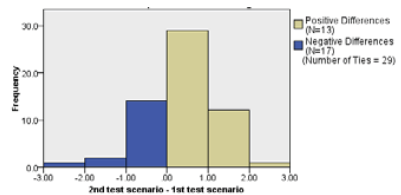


Fig. 8 Related-samples Wilcoxon Signed Rank Test

Visually, it came out that the assumption for the use of this test is fulfilled because the distribution of the difference is approximately symmetrical (it does not necessarily have to be perfect). The result of testing the null hypothesis is shown in Table VI. The significance level is 0.05, so we could not reject the null hypothesis on the basis of the conducted statistical test on the selected data sample.

TABLE VI. HYPOTHESIS TEST SUMMARY

Null Hypothesis	<i>Asymptotic Sig</i>	<i>Decision</i>
The median of differences between the first and the second scenario equals 0	0.342	Retain the null hypothesis

V. CONCLUSIONS

In this paper, we experimentally investigate to what extent is there a relationship between the in-service parameters, derived from the traffic data, and user satisfaction for web browsing service of the mobile network. With this regard, we found out that, specifically, the parameters *Average Time to get first data* and *Average Time to connect* had the strongest relationship with the QoE for web browsing, with the Spearman's correlation coefficients values between the *Average Time to get first data* and QoE equal to -0.770 for the first test scenario, and -0.790 for the second one, while the Spearman's correlation coefficient values between the *Average Time to connect* and QoE were -0.791 for the first test scenario, and -0.831 for the second one.

This indicates that *Average Time to get first data* and *Average Time to connect* parameters could justifiably be considered as the best candidates for estimation of web browsing QoE from network data.

Moreover, though we expected stronger relationship between the *Cancellation rate* parameter and the QoE, the correlation coefficient values obtained for the first and second test scenario, were just -0,499 and -0,320, respectively. So, for the second scenario, we could not reject the hypothesis that there was no relationship between the *Cancellation rate* and the QoE.

Furthermore, we found that the HTTP parameters – specifically: *Maximum Downlink/Uplink Throughput*, *Average Transaction Time*, *Average Transfer Time*, as well as the TCP parameters: *Average Downlink/Uplink Throughput*, *Average Downlink Maximum RTT* are also correlated with user satisfaction for both test scenarios.

Finally, our results derived from the experiments with three sample web pages are in line with previous research. There is a strong negative relationship between the *Web page download time* parameter and QoE. Our data show that the Spearman's correlation coefficients for the test web pages *sport.sport.ba*, *klix.ba* and *vijesti.ba* are -0.834, -0.835 and -0.739, respectively.

REFERENCES

- [1] Cisco and/or its affiliates, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019, 2015
- [2] Ericsson mobility report on the pulse of the networked society, 2015
- [3] ITU-T Recommendation G.1030, Estimating end-to-end performance in IP networks for data applications, 2005
- [4] E. Ibarrola, F. Liberal, I.Taboada, Web QoE Evaluation in Multi-Agent Networks: Validation of ITU-T G.1030, 2009
- [5] Peter Reichl, Bruno Tuffin, Raimund Schatz, “Logarithmic Laws in Service Quality Perception: Where Microeconomics Meets Psychophysics and Quality of Experience”
- [6] Star Khirman and Peter Henriksen, Narus Inc, “Relationship between Quality-of-Service and Quality-of-Experience for Public Internet Service”
- [7] Raimund Schatz, Sebastian Egger, “Vienna Surfing – Assessing Mobile Broadband Quality in the Field”, Telecommunications Research, Center Vienna – FTW
- [8] Ake Arvidsson, Ying Zhang, Neda Beheshti, Detecting “User Dissatisfaction from Passive Monitoring”
- [9] Oracle and/or its affiliates, Oracle Communication Performance Intelligent Center, Oracle data sheet, 2013 K. Elissa, “Title of paper if known,” unpublished.
- [10] Oracle, Oracle® Communications Performance Intelligence Center ProTrace User’s Guide, Release 10.1.5, E56987 Revision 1, 2015
- [11] Copyright 2002 - 2015 Simtec Limited, HTTPWatch Help, <http://help.httpwatch.com/#introduction.html>
- [12] Stephen Hemminger, Network Emulation with NetEm, 2005.

# OFDM Error Floor Based EVM Estimation

Adriana Lipovac  
 Dept. of Electrical Engineering and Computing  
 University of Dubrovnik  
 Dubrovnik, Croatia  
 adriana.lipovac@unidu.hr

Borivoj Modlic  
 Faculty of Electrical Engineering and Computing  
 University of Zagreb  
 Zagreb, Croatia  
 borivoj.modlic@fer.hr

Mislav Grgic  
 Faculty of Electrical Engineering and Computing  
 University of Zagreb  
 Zagreb, Croatia  
 mislav.grgic@fer.hr

**Abstract**— The residual BER – error floor, though useful and widely used metrics for the end-to-end digital radio transmission performance metrics, provides no insight into the error-generating analog impairments (e.g. modulation signal inaccuracy, power amplifier compression, carrier recovery phase error, or I-Q cross-talk), which, however, are easy to identify (so can be dealt with) by the VSA tools such as constellation. However, even this cannot help notice minor distortions that can still seriously degrade the performance, so the only option is the EVM analysis, which has become very popular figure of merit in this regard. So in this paper, after reviewing crucial advantages of EVM analysis with regard to what we get from BER, we consider estimating EVM from the residual BER assuming that the latter’s generally non-linear and non-additive “constituents” are substituted by the equivalent additive Gaussian noise source producing the same BER (and EVM) degradation.

**Keywords**—BER, OFDM, LTE, time-dispersive channel

## I. INTRODUCTION

The primary performance measure of a digital radio system is the *Bit-Error-Rate* – BER. However, the Long-Term Evolution (LTE) specifications express the physical layer performance in terms of *Block-Error-Rate* – BLER, rather than with BER. This is due to arising awareness that, in many practical situations, in-service detecting and counting negative receiver acknowledgements about the successfulness of the data block transfer, relative to the total acknowledgements, which is performed by the *Hybrid Automatic Repeat Request* (HARQ) error control protocol at the link layer, has many advantages over out-of-service BER measurements that presume transmission of pseudorandom binary sequences instead of real traffic.

However, in-service testing may provide inaccurate low BLER values (e.g.  $10^{-5}$ ), which can significantly increase the protocol-data-unit retransmission rate of higher-layer protocols (TCP) and so reduce the throughput of the information (“goodput”). Therefore, testing BER remains unavoidable in LTE digital radio products design and manufacturing, and in some cases (e.g. during equipment installation) in the field.

In an AWGN radio channel, as the *Signal-to-Noise-Ratio* (SNR) gets larger the BER will fall down to its irreducible lower limit called residual BER or the error floor, which remains constant regardless increasing the signal strength and presents the ‘normal’ operating performance of the data link.

In LTE terms, these errors give rise to the related remaining uncorrected block errors determining the related *residual channel* [1-3]. Specifically, the explicit prediction of the residual BER was proposed for the case of the Orthogonal Frequency Division Multiplexing (OFDM) signal transmission in a small-time-dispersion environment, indoor in particular [1]:

$$BER = \frac{1}{2\sqrt{\pi}} \cdot \sqrt{W^- E \left[ \left( \frac{\tau^-}{T_s} \right)^2 \Delta S_{n/n+1}^2 \right]} + \sqrt{W^+ E \left[ \left( \frac{\tau^+}{T_s} \right)^2 \Delta S_{n-1/n}^2 \right]} - \frac{1}{2\sqrt{\pi}} \cdot \frac{\sqrt{W^- E \left[ \left( \frac{\tau^-}{T_s} \right)^2 \Delta S_{n/n+1}^2 \right]} \cdot \sqrt{W^+ E \left[ \left( \frac{\tau^+}{T_s} \right)^2 \Delta S_{n-1/n}^2 \right]}}{\sqrt{W^- E \left[ \left( \frac{\tau^-}{T_s} \right)^2 \Delta S_{n/n+1}^2 \right]} + \sqrt{W^+ E \left[ \left( \frac{\tau^+}{T_s} \right)^2 \Delta S_{n-1/n}^2 \right]}} \quad (1)$$

where intersymbol interference due to multipath propagation is the dominant impairment. The common channel-dispersion parameters in (1) are: rms delay spreads  $E \left[ \left( \frac{\tau^-}{MT_s} \right)^2 \right]$  and

$E \left[ \left( \frac{\tau^+}{MT_s} \right)^2 \right]$ , normalized to the original symbol interval and

distinguished for the advanced and delayed multipath echoes with respect to the chosen sampling instant, respectively, as well as the corresponding composite powers  $W^-$  and  $W^+$ , respectively, relative to the total mean power of all echoes. The signal related parameters in (1) are the differences  $\Delta S_{n/n+1}$  and  $\Delta S_{n-1/n}$  between the  $n$ -th and its following  $(n+1)$ -th OFDM symbol in a row, and between the  $(n-1)$ -th and  $n$ -th OFDM symbol, respectively [1].



Nevertheless, although there is no doubt about the significance of the residual BER as a key end-to-end quality of service performance metric, still it only indicates a problem, but provides no useful clue (such as analog parameter value) on the cause, which would have pinpointed to a specific system component such as transmitter, modulator, oscillator, digital signal processor, transmission path, receiver, or demodulator.

The state-of-the-art vector signal analysis (VSA) [4] definitively provides a number of ways to handle these issues. The LTE lab schematic that we use for the VSA measurements and a typical VSA screen shot are presented in Fig. 1 and Fig. 2, respectively [2].

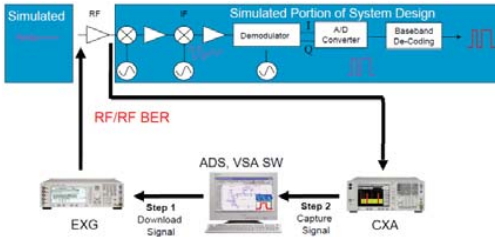


Fig. 1 RF-to-RF BER and VSA testing



Fig. 2 Typical modulation channel VSA screen shot

Specifically, the polar I vs Q vector diagram presents traces of carrier transitions from symbol to symbol at virtually all points in time, so e.g. providing the information to amplifier designers on the adequacy of bias and loading levels.

While the vector diagram is the best way to view the transition trajectories among states, it can be made to look like a constellation by synchronizing the time base of the analyzer to the symbol clock. The resulting constellation diagram provides the carrier amplitude and phase at the symbol clock

time only, and is an excellent modulation signal analysis tool, enabling visible insight not only about the additive noise level of the received signal, Fig.2, but also about many other qualitative signal characteristics, coming out of particular constellation shapes, Fig.1.

Consequently, imperfect constellation patterns pinpoint to the impairments that combined together determine the residual BER, and include e.g. modulator gain and phase imbalance (usually attributed to the modulator or IF section), power amplifier distortion, carrier recovery circuits or I-Q cross-talk, excessive phase noise in the oscillators, improper filtering, clock jitter etc. [4]. So, these can be easily identified and taken care of.

However, although such analog plots are very useful to identify large impairments, the distortions smaller than 10% of the peak values may be difficult to notice. In that case, specific VSA analysis based on Error Vector Magnitude (EVM) measurements is the best option.

Finally, having realized the benefits of the EVM analysis over the pure BER figure, it makes sense to estimate EVM from the residual BER (1), which is the goal of this paper.

In Section II, we review the EVM measurement concept, while in Section III, we focus relationship between the error floor and its belonging EVM and the according EVM estimation from given residual BER. Final conclusions are summarized in Section IV.

II. ERROR VECTOR MAGNITUDE MEASUREMENTS CONCEPT

EVM measurements are sometimes used as an alternative to BER testing, as it provides insight into the modulation quality, specifically with multi-symbol modulation methods such as M-ary Phase-Shift Keying (M-PSK) and M-ary Quadrature Amplitude Modulation (M-QAM) that are widely deployed in wireless local-area networks (WLAN), broadband wireless, and 4G cellular radio systems such as LTE, where M-QAM is combined with OFDM modulation.

As it is presented in the I/Q plane, the Error Vector (EV) is defined as the difference vector between the ideal (or reference) symbol vector and the actual vector assigned to that very symbol, Fig.3.

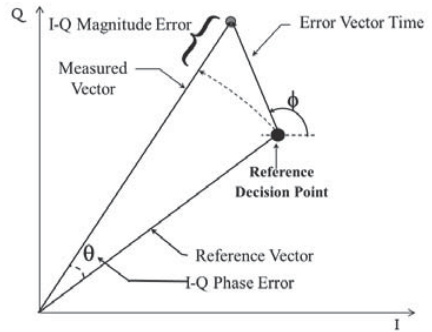


Fig.3 EV vs time (EVT)

With this regard, it is especially useful to measure the EV versus time (EVT), introduced long ago by Hewlett Packard [5]:

$$EVT = \frac{\sqrt{[I_{ref}(t) - I_{mea}(t)]^2 + [Q_{ref}(t) - Q_{mea}(t)]^2}}{\sqrt{I_{ref}^2(t) + Q_{ref}^2(t)}} \quad (2)$$

$$\Phi(t) = \tan^{-1} \left( \frac{Q_{meas} - Q_{ref}}{I_{meas} - I_{ref}} \right) \quad (3)$$

which references the error to the ideal decision points, so that the residual error at the symbols as well as between symbols are computed, Fig. 4.

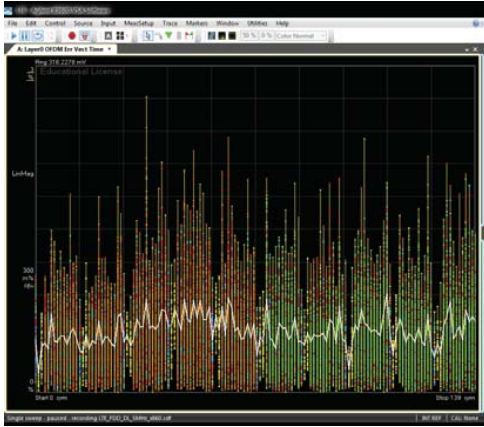


Fig. 4 EVT

The characteristics of these small deviations provides clear differentiation between many types of impairments and so enables assessment of multi-level, multi-phase modulation signals quality based on the measured amplitude and phase distortions that were too small to be visible in the constellation, vector, and eye traces.

Furthermore, averaging (2) along the data sequence provides the rms value of the EVM usually defined in relative terms, i.e. as the ratio of averages of the EV power ( $P_{error}$ ) and the ideal (reference) symbol vector power ( $P_{ref}$ ):

$$EVM = \sqrt{\frac{\frac{1}{N} \cdot \sum_{i=1}^N \{ [I_{ref,i}(t) - I_{mea,i}(t)]^2 + [Q_{ref,i}(t) - Q_{mea,i}(t)]^2 \}}{\frac{1}{N} \cdot \sum_{i=1}^N [I_{ref,i}^2(t) + Q_{ref,i}^2(t)]}} \quad (4)$$

where all squared I and Q components in (4) are properly normalized (so that the total power of any constellation equals unity), enabling that EVM values of different modulation

format (16QAM and 64 QAM that can coexist in LTE downlink) can be mutually compared.

The rms EVM is mostly expressed in dBs:

$$EVM \text{ (dB)} = 10 \log (P_{error}/P_{ref}) \quad (5)$$

or as a percentage:

$$EVM \text{ (%) } = \sqrt{P_{error}/P_{ref}} \times 100 \quad (6)$$

If the symbol errors were caused only by noise, EVM would be equal to SNR at each sample point. However, as it was already mentioned above, other sources of modulation errors exist that are neither additive nor linear, in which case the EVM has been accepted to be an appropriate overall single-number indicator of radio link health.

Typical VSA Error Summary is shown in Fig. 5, where the displayed results include the various EVM values, among them the overall rms, the peak and data only.

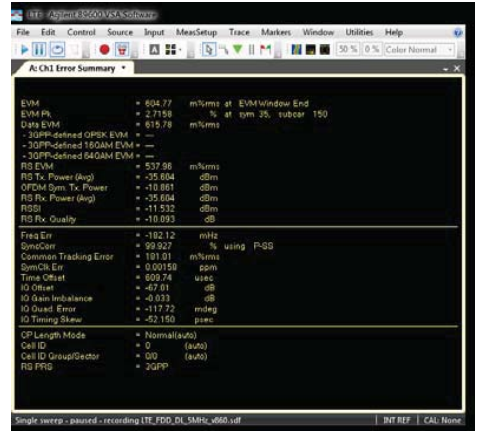


Fig. 5 Error Summary

### III. ESTIMATING EVM FROM RESIDUAL BER

Finally, having justified the advantages of EVM analysis with regard to what we get from BER, the residual one in particular, we consider useful to estimate EVM from BER.

Now let us assume that all residual BER (and the related EVM) “constituents” are substituted by the equivalent additive Gaussian noise source producing the same BER (and EVM) degradation [6].

With this regard, let us review the well-known BER expression for the M-QAM signal transmission over the AWGN channel with [7]:

$$BER = \frac{4}{\log_2 M} \cdot Q \left( \sqrt{\frac{3 \frac{E_b}{N_0} \cdot \log_2 M}{M-1}} \right) \quad (7)$$

where  $E_b$  and  $N_0$  are energy of bit and noise spectral density, respectively. The familiar very steep (“waterfall”) curves, shown in Fig. 6 visually reflect the threshold effect at the digital radio receiver.

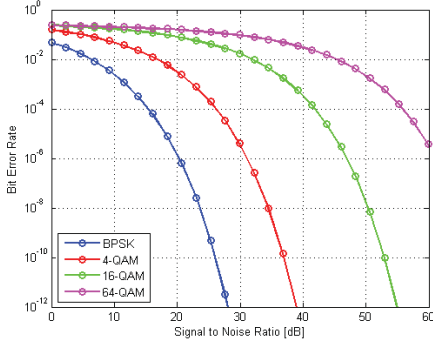


Fig. 6 Waterfall BER(SNR) curves (for Nyquist BW)

Taking into account that SNR can be expressed as:

$$\frac{E_s}{N_0} = SNR = \frac{E_b}{N_0} \cdot \log_2 M \quad (8)$$

where  $E_s$  is the energy of symbol, it follows that:

$$\frac{E_b}{N_0} = \frac{SNR}{\log_2 M} \quad (9)$$

Furthermore, substituting [8]:

$$SNR \approx \frac{1}{EVM^2} \quad (10)$$

into (7), it can be rewritten as:

$$BER = \frac{4}{\log_2 M} \cdot Q \left( \sqrt{\frac{3}{EVM^2 \cdot (M-1)}} \right) \quad (11)$$

From (11) it follows:

$$\frac{BER \cdot \log_2 M}{4} = Q \left( \sqrt{\frac{3}{EVM^2 \cdot (M-1)}} \right)$$

$$\frac{1}{EVM^2} = \frac{(M-1) \cdot \left[ Q^{-1} \left( \frac{BER \cdot \log_2 M}{4} \right) \right]^2}{3}$$

and finally:

$$EVM(BER) = \sqrt{\frac{3}{M-1}} \cdot \left[ \frac{1}{Q^{-1} \left( \frac{BER \cdot \log_2 M}{4} \right)} \right] \quad (12)$$

The plot EVM(BER) is given in Fig. 7.

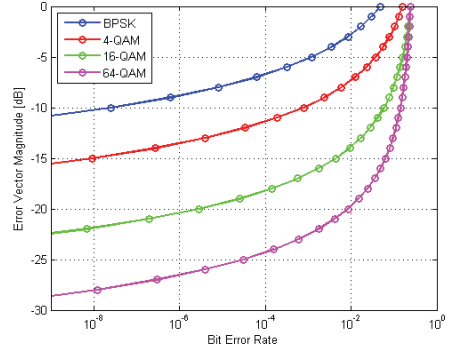


Fig. 7 EVM vs Residual BER

In order to verify (12), we conducted embedded coded BER/BLER tests on the LTE downlink channel (PDSCH 1 (UE 1) with high SNR [3]). The reference channels complied with TS 36.101, and the faded one to the definition in Annex B of TS 36.101, with no HARQ error control deployed.

Specifically, we focused the Extended Pedestrian A (EPA) channel delay profile model.

The obtained BER and EVM results for 4-QAM modulation are presented in Table I, showing very good matching with the values from the corresponding graph on Fig. 7.

Table I. LTE downlink residual BER and average data EVM

BER	4.1846E-3
Data_Avg_EVM %	8.9125

#### IV. CONCLUSION

The residual BER – error floor, though useful and widely used metrics for the end-to-end digital radio transmission performance metrics, provides no insight into the error-generating analog impairments that are easy to identify (and consequently be able to deal with) by the VSA tools such as polar I-Q vectors trajectories and constellation analysis. However, even this cannot help notice minor distortions that can still seriously degrade the performance, so the only option is the Error Vector Magnitude (EVM) analysis, which has become very popular figure of merit in this regard.

So, we considered estimating EVM from the residual BER assuming that the latter’s generally non-linear and non-additive “constituents” are substituted by the equivalent additive Gaussian noise source producing the same BER (and EVM) degradation. The resulting waterfall-shaped EVM(BER) curves were verified in the LTE lab to be a very

good first approximation of EVM from the available residual BER value, when VSA analysis tools are not available.

REFERENCES

- [1] A. Lipovac, „OFDM Error Floor Prediction in a Small-Time-Dispersion Channel“, *Wireless Personal Communications*, 75(1), 645-663, 2013.
- [2] A. Lipovac, V. Lipovac, M. Hamza, “Verification of OFDM Error Floor in Time-Dispersive LTE FDD DL Channel”, *Proc. of the 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, Aalborg, Denmark, 2014.
- [3] A. Lipovac, V. Lipovac, M. Hamza, " Verification of OFDM Error Floor Prediction in Time-Dispersive LTE FDD DL Channel", *Wireless Personal Communications*, DOI: 10.1007/s11277-014-2232-y.
- [4] Agilent Technologies, „3GPP LTE Modulation Analysis - 89600 Vector Signal Analysis Software“, 2010.
- [5] [http://rfmw.em.keysight.com/wireless/helpfiles/89600B/webhelp/subsystem/80216ofdm/Content/trc\\_error\\_vector\\_time.htm](http://rfmw.em.keysight.com/wireless/helpfiles/89600B/webhelp/subsystem/80216ofdm/Content/trc_error_vector_time.htm)<http://www.home.agilent.com/en/pc-1297131/systemvue-electronic-system-level-esl-design-software?&cc=HR&lc=eng>
- [6] Agilent Technologies, „Fundamentals of RF and Microwave Noise Figure Measurements“, 5952-8255E, 2010.
- [7] L. Hanzo, W. Webb, and T. Keller, „Single- and Multi Carrier Quadrature Amplitude Modulation“, Wiley, Chichester, 2nd edition, 2000.
- [8] H. A. Mahmoud, H. Arslan, „Error vector magnitude to SNR conversion for nondata-aided receivers“, *IEEE Transactions on Wireless Communications*, 8(5), 2694 - 2704, 2019.

# Integrated Microstrip Resonator Network for Multi-band Microwave Filter

Gavriloaia Bogdan, Halunga Simona, Fratu Octavian  
Telecommunications Dept.  
University Politehnica of Bucharest  
Bucharest, Romania  
bogdan.gavriloaia@gmail

Vizireanu Radu, Gavriloaia Gheorghie  
Electronics, Communications and Computers  
University of Pitesti  
Pitesti, Romania

**Abstract**— The rapid growth of wireless communications requires a new generation of multifunction devices operating simultaneously under multiple communication standards, in several bands, small, robust and low cost. Microstrip technology can provide these features. An original topological structure is presented in this paper. It integrates several microstrip lines and lumped components in an asymmetric network, and has three ports. A lot of resonance frequencies occur as a result of combination between normal and degenerate propagation modes. Dual-band and three-bands can be selected, depending on the ports used. The originality of this work is to investigate a pentagonal pattern microstrip and introduces two types of perturbations given by two capacitors and a microstrip line section between the corners of the pentagon. The electric field patterns and insertion loss are calculated and provide the possibility of implementing microstrip and larger flexibility for choosing different frequency bands for wireless applications.

**Keywords** - wireless communication, microwave resonators, computational electromagnetics, multi-band filter, transmission lines, degenerate modes

## I. INTRODUCTION

The amount of technical information is doubled to about 2 years, and traffic in telecommunication networks has a time much smaller, usually less than 6 months. Technical support for data transmission is ensured by the telecommunication equipment. The focus may be through their two principal clues: integrability and miniaturization. An increasing number of components, functions and standards define integrability, while miniaturization refers to reduction of size, weight, price, etc. As a result, new design techniques are developed and applied to improve technologies: micro-, nano- or pico-technologies in future close enough.

From the multitude of modern wireless communication systems, satellite and mobile communication systems are those in which a new generation of multifunctional devices simultaneously operating under more communication standards in multi-band microwave ranges, small size, compact, robust, and low cost must be implemented. Bandpass filter, BPF, is an analog component, a key element in RF/ $\mu$ W circuits, used at frequency channel separation for high instantaneous bands and/or moderate dynamic range. It becomes a dual-band, triple-band or multi-band filter when operates on different communication standards.

The BPF was mostly constructed using transmission lines with single-mode resonators. Microstrip resonators are preferred in practical applications because of important features such as: simpler structures, fine adjustment, reduced design complexity and fabrication uncertainty, small and compact sizes, low weight and easy fabrication, low cost, lower loss, and easy integration in integrated circuits. The dual-mode resonators, with a double resonant nature, have the advantages of compact size, high quality factor, and require half as multiple resonators as compared with traditional topology. They could work either by two degenerate modes or non-degenerate modes. The strength of coupling between the degenerate modes is determined by feed line position, size and shape of perturbation from the microstrip network. Simultaneously coupling of some resonant modes produces a shift to further locations of the other resonant modes. The two modes are orthogonal in the case of symmetric network, and no microwave power can be exchanged. The boundary conditions are changed for the asymmetric structures, and the two modes are present at slightly split frequencies.

The idea of using degenerate modes on a microstrip resonator for a double tuned circuit bandpass filter was first proposed in 1972 by Wolf [1]. A lot of research has been done in the recent years on different microstrip topologies: fractal geometries [2], dual-mode loop [3], meander resonators [4], capacitive loaded open-loop arms [5], circular disk, and square patches [6, 7], simple stepped-impedance ring resonator [8], ring resonators [9], etc.

Microstrip resonators could be realized as surface resonators, patches, or as transmission lines. When these resonators operate on dual mode, they could be used as microwave filters or antennas, depending on excitation way. The most used shapes are circular or triangular resonators. An exhaustive presentation of the microwave filters using dual-mode microstrip triangular patch resonators for microwave filters is done in [10]. The microwave planar filter operates on dual mode resulting from the rotation and superposition of the fundamental mode. A certain compromise has to be done between using one-dimensional transmission-line for dual-mode resonators such as rings and loops, and dual-mode resonators such as circular disks or square patches. The first is smaller in size than the two-dimensional patch, but the main drawbacks of the line-based resonators are a higher conductor loss and a lower power-handling capability.

The selectivity and applications of all sorts of resonant modes are the keys elements for designing a patch resonator filters, but the numerous higher order modes are not applied. Instead of it, a new technique that allows miniaturization of microstrip square open loop resonators to be reduced by more than 80% was presented in [11]. The method is based on the loading of resonators with a series surface mount capacitors. In this way, the microwave bandpass filters have a wider stop-band when compared with conventional filters. It is also noticed that the insertion loss of the integrated microstrip network is not modified too much, but in fact, it can be maintained or even enhanced by the integrated process. The explanation lies in the fact that the quality factor of the lumped capacitor is higher than the quality factor of the microstrip resonator. This facility will be used at designing of the proposed integrated microstrip network.

Another key factor of microwave telecommunication filter design is that the same component has to operate at two or three different frequency bands (dual/triple-narrowband bandpass filter). Such filters have complicated structures. For instance, in [11] have been presented multi-layered open and closed loop resonators with a size about  $64 \times 38 \text{ mm}^2$  on its substrate.

This paper focuses on a certain microstrip network integrating some transmission lines sections connected as a pentagon with two small capacitors. A short transmission line section connects two non-successive corners of the polygon.

## II. MATERIAL AND METHODS

During last years, theoretical aspects regarding filter circuit design have not changed significantly. Traditional design methods are based on image parameters or insertion loss. Important improvements have been done in practical implementation. They refer to the technology of new fabrication processes. Transmission line sections used for microwave low power filters are planar lines. They have two specific parameters: lengths and characteristic impedances. Details on design of microstrip lines, width and length, can be found in [11]. Lots of factors determine their choice. Single resonators can be transmission lines sections with half wavelength, which are open-circuited at both ends, and quarter wavelength, which have only one open end. The number of single resonators depends on order of frequency response. A cascade of two or more band pass resonators has to be used for multi-band bandpass filters with the main disadvantage of a large size.

During last years, many topologies like polygons or circles have proposed using dual-mode operation (the dominant mode and its degenerate modes). The degenerate modes arise as a result of various perturbation introduction producing specific effects: inductive, capacitive, mixt or conductive. Introduction of a conductive perturbation is our original idea. The structure perimeter determines the fundamental resonant frequency:

$$\lambda_g = \frac{c}{f_0 \sqrt{\epsilon_{eff}}} \quad (1)$$

where  $\lambda_g$  is the guided wavelength,  $c$  is the velocity of light in free space, and  $\epsilon_{eff}$  is the effective dielectric constant of the substrate. According to (1), for pentagon, the fundamental resonance occurs for:

$$\lambda_g = 5 a \quad (2)$$

where  $a$  is the size-length of one arms of pentagon. In order to minimize the size of the microwave filter, the electric permittivity should be increased. The center frequency of the low-pass band is related by using (2) and (1):

$$f_{01} = \frac{c}{\lambda_g \sqrt{\epsilon_{eff}}} = \frac{c}{5 a \sqrt{\epsilon_{eff}}} \quad (3)$$

The field distribution will be changed by cutting a part of the structure and introducing the two capacitors. On the other hand, the cross couplings between the degenerate modes will generate attenuation poles, and the resonant frequency of higher harmonic wave will be shifted, as well (positive effect for miniaturization).

In the case of TM mode in microstrip network, the electromagnetic field patterns have no amplitude variation along the transversal direction (thickness direction). So, the electric field has one component only,  $E_z$ , and two magnetic field components non-zero. The electric field longitudinal component is a solution of 2-D Helmholtz equation:

$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_{m,n}^2 \right) E_z = 0 \quad (4)$$

The magnetic field transversal components can be evaluated as:

$$H_x = \frac{j}{\omega \mu_0 \mu_r} \frac{\partial E_z}{\partial y} \quad (5a)$$

$$H_y = -\frac{j}{\omega \mu_0 \mu_r} \frac{\partial E_z}{\partial x} \quad (5b)$$

The above equations are important guidelines for microwave filter design. By solving them, important data can be got regarding the field pattern inside of the microstrip network, as for fundamental,  $f_{1,0}$ , as for superior harmonics,  $f_{m,n}$  (corresponding to  $m > 1$ , and  $n > 1$ ):

$$f_{m,n} = f_{1,0} \sqrt{m^2 + n^2 + mn} \quad (6)$$

The pattern of filter characteristic inside of passband is given by the nature of the lumped component: inductive or capacitive used as perturbation or coupling element. In the case

of the capacitive coupling, the filter will have an elliptic function characteristic because it creates the transmission zeros on both the upper and lower sides of the passband. The Chebyshev function characteristic appears by using the inductive coupling. The microstrip section connecting the two pentagon corners behaves as inductive or capacitive coupling, depends on the ratio between its length and the investigated frequency.

### III. SIMULATION RESULTS

The conventional design of the dual-mode bandpass filter is depicted in Fig. 1. The structure integrates a microstrip close-loop, a short galvanic connection between two corners of pentagon, and two lump capacitors ( $C_1=1$  pF,  $C_2=2.4$  pF). The entire topological structure is realized on a 31-mill-thickness square substrate with a dielectric constant of 3.9, and has no axis of symmetry. The input impedances of the  $P_1$ ,  $P_2$ , and  $P_3$  ports are  $50 \Omega$ . The filter is arranged in a square box with the wall impedance matched to the air impedance that surround the dielectric substrate in order to avoid reflections from them. Depending on how the ports are connected, different frequency characteristics of the filter can be obtained.

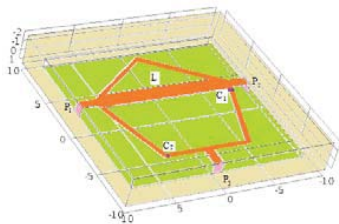


Fig. 1. Set up of the integrated microstrip network

In the first step, it is important to know what the values of the resonance frequencies are, and whether they are expressed by real or complex numbers. The complex numbers correspond to degenerate modes of propagation. Three such complex numbers had obtained for the set up. The real numbers show the critical frequencies for the embedded structure. A special attention has to be done because a part of them are for the outer box. To avoid this situation, the pattern of the electric field is evaluated. The electric field pattern for the fundamental resonant frequency is shown in Fig. 2.

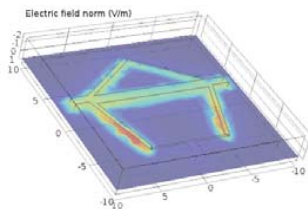


Fig. 2. Electric field pattern at the fundamental resonant frequency

As it can be seen, only  $P_1$  and  $P_2$  ports are interconnected, and  $P_3$  port is isolated. This means that between these two ports the microstrip network behaves as low-pass filter. The electric

field pattern at the first harmonic is depicted in the Fig. 3. In this case no port is connected, the entire energy is embedded inside of microstrip network, and the signal will be strongly rejected, the central frequency of the stopband filter.

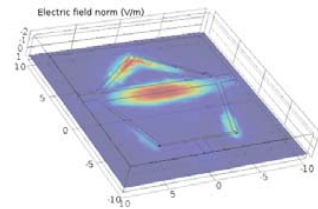


Fig. 3. Electric field pattern at the first harmonic frequency

The pattern electric field is shown in Fig. 4 in the case of the second resonance frequency. The microwave signal propagates from one port to other, and the network operates as passband filter. At the third resonance frequency it will be possible to transfer energy from  $P_2$  to the two ports  $P_1$  and  $P_3$ , Fig. 5. More electromagnetic energy can be transferred from  $P_3$  port to  $P_1$  and  $P_2$  ports at the forth resonance frequency, Fig. 6.

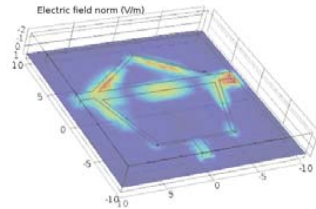


Fig. 4. Electric field pattern at the second resonance frequency

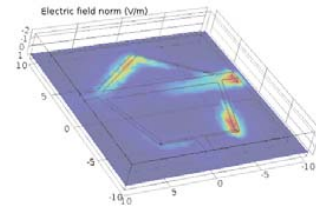


Fig. 5. Electric field pattern at the third resonance frequency

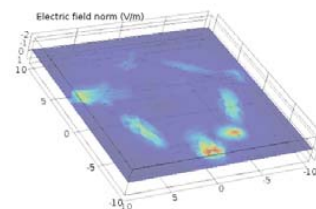


Fig. 6. Electric field pattern at the forth resonance frequency

The electric field pattern at the fifth resonance frequency, shown in Fig. 7, represents spatial distribution of the electromagnetic energy when the conductive connection operates between P1 and P2 ports at double resonance frequency. The P3 port is isolated. The image depicted in Fig. 8 corresponds to a similar situation as in Fig. 4, but for dual-mode. Both Chebyshev and elliptic function responses were obtained for the proposed device. They can be seen in Fig. 9.

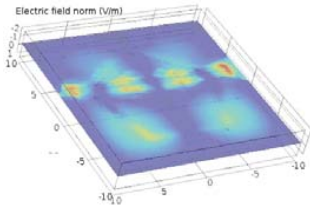


Fig. 7. Electric field pattern at the fifth resonance frequency

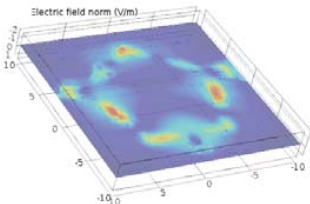


Fig. 8. Electric field pattern at the sixth resonance frequency

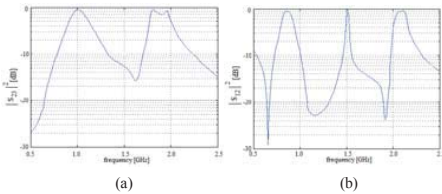


Fig. 9. The characteristic functions of the microstrip network

#### IV. CONCLUSIONS

An integrated microstrip resonator network consisting of pentagonal loop loaded with three perturbation elements has been proposed. Two operating modes in dual- and three- bands

were provided on the same physical structure. The microwave filter offers the needed features for application in modern wireless communication equipment working simultaneously on more standards.

#### ACKNOWLEDGMENT

The work has been funded by the EC through the FP7-IP project eWALL no. 610658, and by UEFISCDI through the FP7 eWALL support project no. 262/EU.

#### REFERENCES

- [1] I.Wolff, "Microstrip bandpass filter using degenerate modes of a microstrip ring resonator," IEEE Electronics Letters, Vol. 8, No. 12, 302–303, June 1972.
- [2] A. Alshurafa, A. Radwan, A. Emira, K. Salama, "RF MEMS fractal capacitors with self-resonant frequencies". Journal of Microelectromechanical Systems, 2012, vol. 21, no.1, p. 10-12.
- [3] W. Tu, and K. Chang, "Miniaturized dual-mode bandpass filter with harmonic control," IEEE Microwave and Wireless Components Letters, Vol. 15, No. 12, 838–840, Dec. 2005.
- [4] J.-S. Hong, and M. J. Lancaster, "Microstrip bandpass filter using degenerate modes of a novel meander loop resonator," IEEE Microwave and Guide Wave Letters, Vol. 5, No. 11, 371–372, Nov. 1995.
- [5] A. Guorur, C. Karpuz, and M. Akpınar, "A reduced-size dualmode bandpass filter with capacitively loaded open-loop arms," IEEE Microwave and Wireless Components Letters, Vol. 13, No. 9, 385–387, Sept. 2003.
- [6] J-S. Hong, and M. J. Lancaster, "Microstrip triangular patch resonator filters," IEEE MTT-S International Microwave Symposium Digest, 2000, V ol. 1, 331–334, June 2000.
- [7] M. Matsuo, H. Yabuki, and M. Makimoto, "Dual-mode stepped impedance ring resonator for bandpass filter applications," IEEE Trans. Microw. Theory Tech., vol. 49, no. 7, pp. 1235–1240, Jul. 2001.
- [8] E. E. Djoumessi and K. Wu, "Multilayer dual-mode dual-bandpass filter," IEEE Microw. Wireless Compon. Lett, vol. 19, no. 1, pp. 21–23, Jan. 2009.
- [9] C. Hua, C. Miao "Microstrip bandpass filters using dual-mode resonators with internal coupled lines," Progress In Electromagnetics Research, Vol. 21, pp.99-111, 2011.
- [10] J-S. Hong, L. Shuzhou, "Theory and Experiment of Dual-Mode Microstrip Triangular Patch Resonators and Filters", IEEE Transactions on Microwave Theory and Techniques, vol. 52, no. 4, april 2004 1237-1243 ,
- [11] J. Helszajn, and D. James, "Planar triangular resonators with magnetic walls", IEEE Trans. Microwave Theory, MTT-26, 95-100; 1978



# SUSTAINABLE ROUTING PROTOCOL TO IMPROVE LIFE TIME OF MANET

Manohar Chaudhari  
Faculty of  
Telecommunications  
Technical University of  
Sofia, Bulgaria  
mschaudhari20@gmail.com

Soumya Ranjan Samal  
Faculty of  
Telecommunications  
Technical University of  
Sofia, Bulgaria  
Samal.sr1981@gmail.com

Pavlina Koleva  
Faculty of  
Telecommunications  
Technical University of  
Sofia, Bulgaria  
p\_koleva@tu-sofia.bg

Vladimir Poulkov  
Faculty of  
Telecommunications  
Technical University of  
Sofia, Bulgaria  
vkp@tu-sofia.bg

**Abstract**—In today’s IOT scenario the connection of billions of devices will pave the way for the development of new services and applications in different domains. As a result huge data traffic in the network will be generated and the increase in the power consumption of the network will become an issue that is necessary to be considered. This will be a challenge especially for mobile ad-hoc networks (MANETs) since the nodes are battery powered and an efficient usage of the scarce battery power of the mobile nodes should be considered. This paper proposes a power efficient routing technique called Distance-power based DSR (DPDSR) to improve the existing DSR protocol by introducing a power aware approach in MANET. In the route discovery process, the DPDSR monitors the residual energy of nodes and the distance between intermediate nodes from the source node. Hence, an improvement of the network lifetime and prolongation of its survivability is achieved.

**Keywords**—DSR, Energy, MANET, Routing protocols, Wireless communication

## I. INTRODUCTION

A Mobile ad hoc network (MANET) is a collection of wireless nodes, which lacks any centralized authority to govern the communication activities within the network [1]. Up to now, extensive efforts have been dedicated to building a set of standard routing protocols. However, some of the released standard protocols, such as Ad Hoc On-Demand Distance Vector (AODV) and Dynamic Source Routing (DSR), have their pros and cons, but none of them is superior to the others in all contexts [2]. DSR (Dynamic Source Routing) is an on-demand routing protocol designed specifically for MANET [3,4]. The DSR protocol is oriented around two phases. In the route discovery phase, the source initiates a route discovery procedure if it does not have a route to the destination in its route cache. It broadcasts a special packet called Route Request packet (RREQ) to the neighboring nodes. Each intermediate node adds its ID in the packet and forwards it to the next intermediate node. If the route to the destination is known, a reply to the source node about the availability of such a route is sent. When the RREQ reaches the destination, it generates a Route Reply packet (RREP) and sends it back to the source node through the same path that it had arrived. There can be multiple path requests to the destination, but only the first one will be accepted, while all others will be dropped.

This paper proposes a new power efficient routing technique called Distance-power based DSR (DPDSR) to improve the existing DSR protocol by introducing a power aware approach for routing in MANETs. The rest of the paper is organized as follows: Section II focuses on related work, Section III explains the proposed algorithm, Section IV describes the simulation environment and elaborates on simulation results, Section V is the conclusion.

## II. RELATED WORK

There are multiple protocols popular in IoT/WLAN environments like AODV, DSDV, DSR, etc. All the protocols have different characteristics and are suitable for different applications. DSR is one of the most prominent routing protocols for MANETs.

In DSR as the entire route is contained in the packet header, there is no need of having a routing table to keep route for a given packet. It selects the path having the minimum hops count, as a result of which the number of intermediate nodes is smaller and the distance between each pair of nodes larger. With the increase of the distance between the nodes, the required transmission power for communication between any pair of nodes increases, more battery power is consumed and the overall network lifetime tends to decrease. The proposed mechanism in this paper introduces a modification of the DSR protocol mitigating this drawback.

Protocols representing energy efficient modifications for the DSR Protocol are discussed in [5]. In [6] energy-aware routing algorithms for wireless ad hoc networks, called reliable minimum energy cost routing (RMECR) and reliable minimum energy routing (RMER), are proposed. These algorithms consider reliability, energy efficiency and increase the network lifetime of the nodes. RMER does not consider the remaining battery energy of nodes. The routing metric in RMER and RMECR algorithms depends on the quality of the links. In [7] Shivashankar et al. proposes another efficient power aware routing (EPAR) that increases the network lifetime of MANET. Unlike traditional power aware algorithms, EPAR identifies the performance ability of a node not just by its remaining battery power, but also by the anticipated energy spent in the reliable data transfer over a specific route. In [8] a Local Energy Aware Routing (LEAR) Protocol is described,

where the “Willingness” of the nodes to participate in the routing process is based on the remaining battery power. But the route request is dropped if a single intermediate node in the route has a lower battery power level than its threshold value. If this occurs for every possible path, then the source will never receive a single route reply message even if there exist a path between source and destination. In [9] a so-called Energy Saving Dynamic Source Routing (ESDSR) Protocol is proposed aiming to prolong the network lifetime by focusing on a transmission power control and load balancing approach. But each node has to maintain a power table, which increases the overheads. In [10] the authors propose Energy Dependent DSR Routing (EEDSR) Protocol, where residual battery Power and power consumption per second (drain rate) are considered to calculate the predicted life time of a node. The major limitation of this algorithm is that the nodes having a high value of remaining power are injected with heavy traffic. In a dense network scenario, EEDSR improves the drain rate in terms of average node life times. In [11] an Energy Efficient DSR Protocol (E2DSR) is considered where a new energy field is inserted in the RREQ message, which is used to store energy information of each visited node. The route having maximum energy is selected considering link length, freshness and path energy.

The algorithms discussed above are proven to be better in different aspects in comparison with DSR. Still, there is scope to improve the functionality of DSR by modifying the basic mechanism of DSR as an energy efficient routing protocol. It can be observed that no approach is proposed in DSR routing protocol considering and related to the residual energy of a node and the distance between intermediate nodes.

This paper proposes new modified DSR routing technique based on the combination of residual energy of each forwarding node and the distance from the source, with the goal of improving energy efficiency and increasing overall network lifetime.

### III. PROPOSED ALGORITHM

The proposed Distance Power based DSR (DPDSR) algorithm aims at improving the energy efficiency, hence the survivability of the nodes and network lifetime. As energy is a critical issue for the survival of nodes in the network, the proposed algorithm is based on monitoring the energy levels of the nodes while building the on-demand path to the destination. Figure 1 shows the proposed algorithm flowchart.

To establish the path, the source node forwards a route request packet (RReq) to the intermediate nodes, to check for the residual energy and distance. This algorithm tends to select the intermediate nodes based on two parameters: a) residual energy of the node greater than a predefined threshold energy, and b) minimum distance  $D_{Si}$  between the source S and the  $i^{th}$  intermediate node i.

As DSR selects the path having the minimum possible hops count, so the distance between some of the pair of nodes usually is bigger which requires bigger transmission power for communication between the pair of nodes. The basic idea

behind the proposed algorithm is to select the intermediate nodes (neighbouring node) based on two criteria - minimum distance from the source and energy greater than a predefined threshold value  $E_{thr}$ . The DPDSR algorithm flow chart is shown in Figure 1 and is described below.

**STEP 1:** Initiate route discovery by broadcasting a route request packet (Rreq).

**STEP 2:** Define the Source (S) and Destination (D) node. Define  $E_{thr}$ .

**STEP 3:** Calculate the Distance between Source and Destination using the distance formula

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

where,  $(x_1, y_1)$  and  $(x_2, y_2)$  are the position coordinates of source and destination node respectively.

This parameter is used for comparison purpose. Any intermediate node having distance greater than this will not be selected for packet forwarding.

**STEP 4:** Calculate the residual energies of source node ( $E_S$ ) and destination node ( $E_D$ ). These values give the estimation if there is sufficient energy to transmit and receive the packet. In the proposed algorithm every node should have energy greater than the predefined threshold value  $E_{thr}$ .

**STEP 5:** Calculate the distance to energy ratio between source node and destination node,

$$DE_{SD} = D_{SD} / (E_S + E_D) \quad (2)$$

Here  $D_{SD}$  is the distance between the two nodes;  $E_S$  is the energy of the source node, and  $E_D$  is the energy of the target node.

**STEP 6:** For each next  $i^{th}$  intermediate neighboring node check if it is destination node or not.

**STEP 7:** Check If ( $i^{th}$  node ( $N_i$ ) = Destination node (D)). If yes, the target node is reached and a reply is send to the source node with the complete path.

**STEP 8:** Go to step 17 to send Route reply.

**STEP 9:** Calculate the residual energy of  $i^{th}$  neighbouring node ( $E_i$ )

**STEP 10:** Select the node for packet forwarding having sufficient energy greater than threshold energy. This step checks If ( residual energy of  $i^{th}$  neighbouring node  $E_i >$  threshold energy  $E_{thr}$  ) then select the node for further calculations, otherwise go for next node. In this step reliability of communication is ensured.

**STEP 11:** This step calculates the distance between source and  $i^{th}$  neighbouring node.

$$D_{Si} = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} \quad (3)$$

where  $x_i, y_i$  are coordinates of the  $i^{th}$  node. This parameter is useful to estimate the distance of the intermediate neighbouring nodes which should be less than  $D_{SD}$ .

**STEP 12:** Else Go to step 6 for next node selection.

**STEP 13:** Calculate distance to energy ratio for  $i^{th}$  node ( $DE_{Si}$ )

$$DE_{Si} = D_{Si} / (E_S + E_i) \quad (4)$$

where  $E_S, E_i$  energy of source node and intermediate node.

STEP 14: Check If  $(DE_{Si}) < DE_{SD}$  then go to step 15.

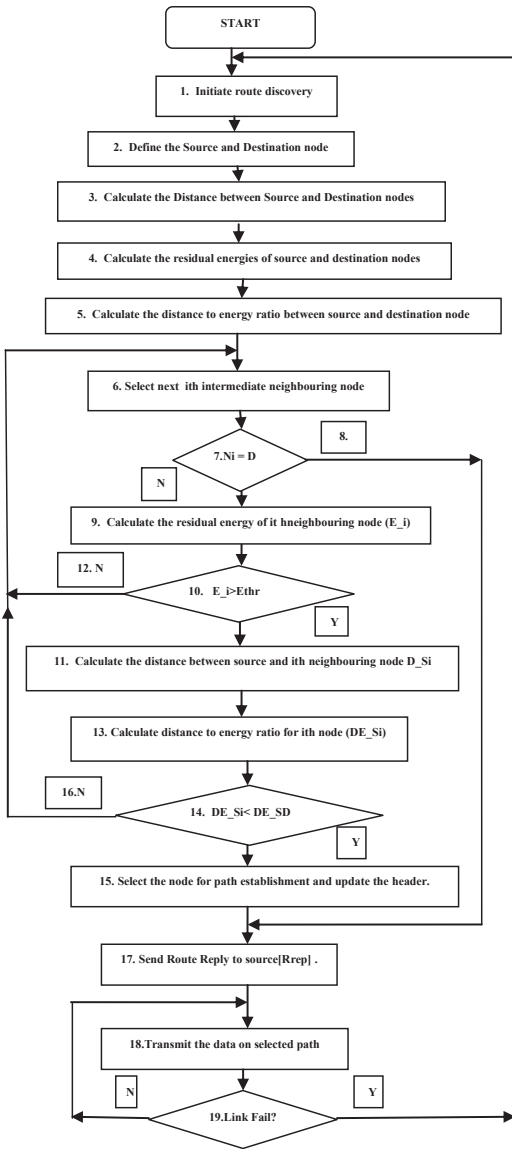


Fig. 1. DPDSR algorithm flow chart

The idea behind the DPDSR algorithm is the selection of a route based on the comparison of the two parameters  $DE_{Si}$  and  $DE_{SD}$ . An intermediate node for a route is selected if the value of the  $DE_{Si}$  parameter is smaller as compared to  $DE_{SD}$ . Here due to the ratio of the distance to the energy  $DE_{Si}$ , the probability of selection of a given node as part of a power aware route is increased. To satisfy the condition of a smaller ratio, the distance should be small and energy high. The proposed DPDSR algorithm ensures that an intermediate node that is having a distance  $D_{Si} < D_{SD}$  and residual energy bigger than the threshold energy will be selected as a part of the route. Due to small distance, the power required to transmit the data packet will be minimized.

STEP 15: if condition  $DE_{Si} < DE_{SD}$  satisfy, Select the node for path establishment and update the header.

The goal is to balance the power consumption within the network. Normally in the DSR routing algorithm, the shortest path is selected between the source (S) and the destination (D). In the first iteration when the packets are sent from S to D, few selected intermediate nodes are used for packet forwarding. The relative energy consumption of an intermediate node in the path is proportional to the number of packet transmission and forwarding. As a result, every time an intermediate node is selected in the route, this will lead to a decrease of its initial residual energy. Practically the energy decrease is proportional to the number of times a node is included in the routing process. If one and the same intermediate nodes are selected for packet transmission and forwarding more frequently, the result could be energy exhaustion and node outage, leading to shorter network lifetime and less sustainability.

STEP 16: if  $(DE_{Si} < DE_{SD})$  is not satisfied Go to step 6 to select next node and repeat the steps 6 to 14.

STEP 17: Send route reply (Rrep) to source for selected path.

STEP 18: Transmit the data on selected path.

STEP 19: If (Link break due to a node failure) then go to step 1 to initiate Route discovery else go to step 18 to send the data on the previous path.

To prolong the network life time and network sustainability the DPDSR algorithm selects an alternate path based on distance and the residual energy of the nodes. In this case due to the selected energy parameters the path may not be the shortest path, because in the communication process, the intermediate nodes lose their initial energy. If the residual energy of an intermediate node falls below a given threshold, it will not be selected for the route. The proposed algorithm selects another path based on distance and residual energy ratio. In this way for same S and D there are multiple paths available with nodes that have sufficient energy to forward packets. In this way, more nodes participate in the routing process and a better balance of the power consumption within the network is achieved.

The proposed algorithm monitors the distance between intermediate nodes along with their residual energy and tries to build a path with nodes having a smaller distance from each other and sufficient energy to forward the packets

successfully. As the DPDSR algorithm, considers the ratio of the distance over the sum of the energy of the corresponding nodes, this will ensure that nodes with more residual energy and shorter distance between them will be selected in the routing process.

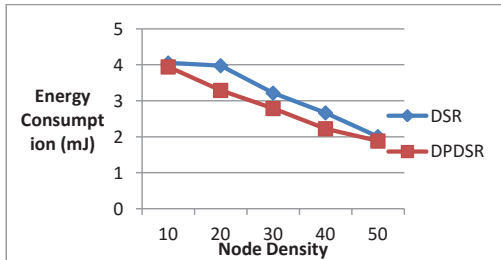
**IV. SIMULATION ENVIRONMENT AND RESULTS**

For simulation, we use NS2 as Network simulator, which is an open-source simulation platform [12]. The proposed algorithm DPDSR is compared with existing DSR algorithm in terms of Throughput, Energy, End to End Delay and Jitter. A standard scenario is created in which various network conditions are tested, with the simulation parameters shown in Table 1.

**Table 1: Simulation Environment for implementation of DPDSR**

Simulation Parameters	Values
Simulation Area	500m X 500m
Channel Type	Wireless Channel
MAC Type	IEEE 802.11
Traffic Type	CBR
Propagation Model	Two Ray Ground
Network Interface Type	Phy/Wireless PHY
Routing Protocols	DSR/DPDSR
Number of nodes	10-50
Mobility	2-10 m/s
Initial Energy of node	10 Joules
Reporting Rate	2-10 pkt/sec
Packet Size	100 bytes

After implementing the proposed DPDSR algorithm, the following results were observed.

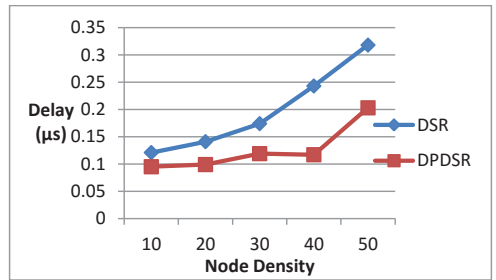


**Fig. 2. Energy consumption Vs. increasing number of nodes**

Figure 2 shows a comparison of the overall energy consumption versus the increasing number of nodes for the DSR and the DPDSR protocol. The node speed is kept constant; which is 2 m/s. It is seen that DPDSR performs better and reduces the overall energy consumption. Such performance is observed because of the consideration of the power parameters in the route discovery process. With the

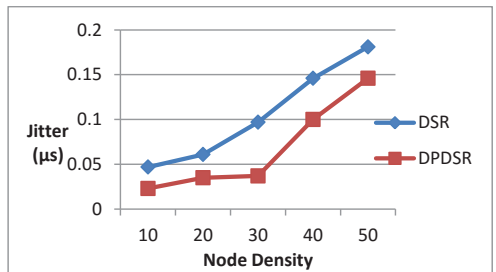
increase in the number of nodes i.e. the node density, the overall energy consumption decreases in both the cases since the average distance between the nodes becomes smaller which requires less energy for transmission. As the number of the nodes increases the routing performance is good, but there is a threshold for the density of nodes, because of probability for congestion in the network.

Figure 3 shows the comparison of the End-to-end delay in both cases with the increase of the number of nodes in the network. The speed of nodes for all readings is kept constant at 2 m/s. It could be seen that for the DSR algorithm delay increases as node density increases. This could be explained with the nature of the algorithm, as it always tries to find the shortest path irrespective of node load or its energy consumption and forwards the packets to a destination in one and the same route. In such a case the waiting time or the time it takes a packet to



**Fig. 3. Delay Vs. Increasing Number of Nodes**

be processed in the buffers of the nodes increases and so the overall delay and probability of network congestion. On the contrary the DPDSR algorithm selects the routing path based on the DE<sub>S</sub>i ratio, thus for packet forwarding are chosen the less loaded nodes that have sufficient energy and need less time for processing. The results indicate that in some scenarios with different distribution of the nodes the delay in the DPDSR could be up to 35% lower than in the DSR case.



**Fig. 4. Delay Jitter Vs. Increasing Number of Nodes**

Figure 4 shows the comparison of the delay jitter values encountered at increasing number of nodes. It could be seen

that the jitter values in the DPDSR implementation are less than those of the existing DSR. As in the delay case, this is because of the nature of the DPDSR algorithm related to the node selection based on the DE\_Si ratio. Delay jitter is directly dependent on the buffer occupancy of the nodes (even in most cases jitter reduction is achieved through buffering). Thus nodes with lower load tend to generate less packet delay jitter. Networks having lower jitter values are considered more suitable for reliable packet delivery.

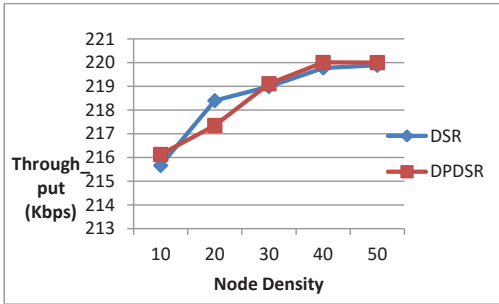


Fig. 5. Throughput Vs. Node density

Figure 5 shows the comparison of overall throughput for both protocols versus increasing number of nodes. In this case, again the speed of nodes is maintained as 2 m/s. It is observed that both protocols perform very similar and the modified implementation does not reduce the throughput. As the count of the nodes increases the throughput goes on increasing because of the higher the availability of multiple paths and thus lower probability of congestion and link breakage.

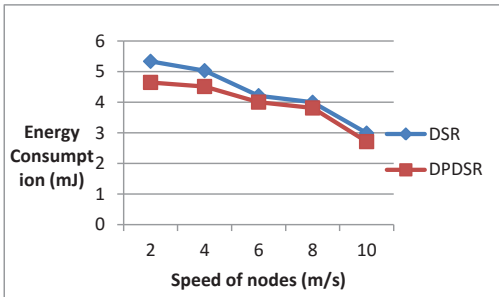


Fig. 6. Energy consumption Vs. Increasing speeds of nodes

Figure 6 shows the overall energy consumption Vs. Increasing mobility of nodes. The number of nodes is kept constant at 20. DPDSR outperforms the existing DSR protocols due to the power aware approach adopted by DPDSR. It could also be seen that the DPDSR energy consumption and its deviation is less affected by the mobility of nodes.

## V. CONCLUSION

Network survivability and network life time are one of the critical issues in MANET and are directly related to energy efficiency. There is a continuous demand to upgrade energy efficient routing techniques to save scarce battery power. This work proposes a power efficient routing technique called Distance-power based DSR (DPDSR) to improve the existing DSR protocol by introducing a power aware approach in the MANET. The proposed DPDSR algorithm was compared with the existing DSR regarding their energy consumption, throughput, delay and jitter. In the simulated scenario, the overall energy consumption is reduced and better performance is observed regarding throughput delay and jitter. This comes to show that due to the introduction of a power aware approach in the routing process, the overall energy consumption in the network, as well as other performance parameters such as delay and jitter could be significantly reduced. Such an approach could significantly improve the lifetime of the nodes and increase their survivability in MANET.

A potential drawback of the algorithm is some increase in the computational complexity due to the introduction of the additional metrics, which need to be calculated in the routing process. To improve this, as well as the overall routing performance, in the future work possible implementations of clustering or heuristic algorithms, together with power aware routing approaches will be analyzed and implemented.

## REFERENCES

- [1] Utkarsha P. Govindnar, ShubhamL.Paraskar, "An Overview of MANET and its Applications" International Journal of Engineering Applications and Technology, March-2013.
- [2] Marco Conti, Silvia Giordano, " Mobile Ad Hoc Networking: Milestones, Challenges, and New Research Directions" IEEE Communications Magazine • January 2014.
- [3] David B. Johnson, David A. Maltz and Yih-Chun Hu, The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR), <draft-ietf-manet-dsr-10.txt> Internet-draft, 19 July 2004.
- [4] D. B. Johnson, D. A. Maltz, and J. Broch, "DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad hoc Networks" IETF Draft, April 2003, work in progress. <http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-09.txt>
- [5] Vahid Nazari et al., "An Energy Efficient Flat Routing for Mobile Ad-hoc Networks," Poce. INFOS-ICT-22564, 2010.].
- [6] J. Vazifehdan, R. V. Prasad, I. Niemegeers, "Energy Efficient Reliable Routing Considering Residual Energy in Wireless Ad-Hoc Networks." IEEE Trans. on Mobile Computing, Volume 13, No.2, February 2014
- [7] Shiva Shankar, Golla Varaprasad, Hosahalli Narayan Gowda Suresh, "Importance of on-demand modified power aware dynamic source routing protocol for mobile ad-hoc networks," IET, Microwaves, Antenna, and Propagation-2014.
- [8] Stojmenovic I, Lin X., "Power-Aware Localized Routing in Wireless Networks," IEEE Trans. Parallel and Distributed Systems 2001; 12(11):1122-1133.
- [9] Tarique M., Tepe K., and Naserian M., "Energy Saving Dynamic Source Routing for Ad-hoc Wireless Networks," In WIOPT, 2005.
- [10] Garcia J.-E., Kyamakya A. K., Jobmann K., "A Novel DSR-based Energy-efficient Routing Algorithm for Mobile Ad Hoc Networks" IEEE Trans., Pages 2848-2854, 2003.
- [11] Vahid Nazari Talooki, Hugo Marques, Jonathan Rodrigue and Hugo Agua, "An Energy Efficient Flat Routing for Mobile Ad-hoc Networks," Poce. INFOS-ICT-22564, 2010.
- [12] The network simulator-ns 2.35. <http://www.isi.edu/nsnam/ns/>.

# The New Methodology for Vehicular Network

Krittika Kantawong and Rounsang Chairsicharoen

School of Information Technology  
Mae Fah Luang University  
Moung, Chiangrai, Thailand

Krittika.ka@up.ac.th, rounsang.cha@mfu.ac.th

**Abstract**— This work aims to construct a new methodology for the vehicular network. In the real world, route guidance is important for organizations. Cost of transportation can be saved by effective vehicle routes significantly. In this work, a new methodology has been proposed to solve the vehicular network in order to minimize cost. The proposed algorithm is tested on Solomon benchmark datasets and the computational results are compared with other algorithms. The finding from the experiment shows that the proposed algorithm produced satisfied results and can compete with other algorithms.

**Keywords**—vehicle routing problem; fuzzy membership function; route construction

## I. INTRODUCTION

Transportation is an important section in business and organizations. Efficient transportation systems provide economic and social opportunities and benefits. Transportation cost could be desired as the significant national expense in many countries such as the United Kingdom, France and Denmark [1], [2]. Moreover, its cost resulted in an increasing of goods price for 70% [3],[4],[19]. On the economic level, Rodrigue and Notteboom confirmed that it accounts from 6% to 12% of the GDP in many developed countries. In addition, they pointed out that 10%-15% of the household expenses are related to the transportation cost in average, while it accounts to the cost of each output unit in manufacturing around 4%. [5]. In 1989, 76% of the products have been transferred from places to places by vehicle transportation [6] which confirms the significance of the vehicle routing problem.

The Vehicle routing problem (VRP) was introduced by Dantzig and Ramser in 1960 [7], the high complexity and significance of the problem have been attracting researchers to study this discipline extensively. A massive number of existing literatures have been published by many researchers. In the real world problem, there are many variants of the VRP incorporating constraints and conditions. The Capacitated VRP (CVRP) vehicles have a limited freight capacity and the Vehicle Routing Problem with Time Windows (VRPTW) time interval of each customer is specified. Heuristics, Meta-heuristics and Hybrid Algorithms have been proposed to figure out the drawback of the exact method, they can solve the problem efficiently. The hybrid algorithm is the newest technique which has attracted researchers in the world to develop a proper method for solving the VRP problem.

This paper aims to solve the VRPTW by a new methodology where all of the related details are described in

the methodology section, the computational results and analysis section, and in the conclusion section.

### A. Vehicle Routing Problem with Time Windows

The VRPTW can be reviewed as a combination of vehicle routing and scheduling problem. A set of customers will be satisfied by a fleet of vehicles which start and end the service at the depot. Each customer will be served by a vehicle within their time windows in order to minimize the cost and total of customers demand in the route that does not exceed the maximum of vehicle capacity.

Toth and Vigo [30] define VRPTW as follows: it is a complete graph  $G=(V,E)$ . There is a set of vertices  $V=\{v_0, v_1, \dots, v_n\}$ ,  $v_0 \in V$  represents depot,  $N$  is a sub set of vertices represents customers,  $N=\{v_1, \dots, v_n\}$ . The location of customer  $v_i$  represents by  $(x_i, y_i)$  and the representation location of the depot,  $v_0$  is  $(x_0, y_0)$ . For each customer  $v_i \in N$ , there is a demand  $d_i$ , service time or unloading time  $t_i$  and a time window  $[e_i, l_i]$  where  $e_i$  is the earliest time to begin service and  $l_i$  is the latest time. Accordingly,  $s_i$  denotes the starting time service at customer  $v_i$ , the vehicle must wait with the waiting time  $w_i$  if the arriving time at customer  $v_i$  before time  $e_i$  and it's not allow to arrive after latest time  $l_i$ . Each pair of customers, there is  $t_{ij}$  represents a time that takes to go from customer  $v_i$  to  $v_j$ . Customer  $v_i$  is served by a vehicle at the beginning time  $s_i$  takes time  $t_i$  to unload or pickup goods. At the depot,  $v_0 \in V$ , there are  $m$  homogenous vehicles. Every vehicle starts and end service to customer with the specific quantity  $Q$  at the depot. The objective of the VRP is to find a set of  $K=\{k_1, k_2, \dots, k_m\}$  routes which serve all customers by  $m$  vehicles at the minimum cost. The cost can be the sum of travel distances, times, number of vehicles or other relevant factors of all routes. The overall distance is desired in this paper. The distance between a pair of customers  $v_i, v_j \in N$  is denoted by  $c_{ij}$  calculated by the Euclidean distance (1).

$$c_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

The objective of the VRPTW is to minimize cost which is referred to as the distance. The objective function is shown as follows:

$$\text{Min} \sum_{i=1, j=1}^N c_{ij} x_{ij} \quad (2)$$

In (2),  $N$  customers are waiting to be served by a fleet of vehicles,  $c_{ij}$  refer to distance between customer  $i$  and customer  $j$  ( $i \neq j$ ) and  $x_{ij}$  is a decision variable.  $x_{ij}$  is equal to 1 if there is a path between customer  $i$  and customer  $j$ , otherwise  $x_{ij}$  is equal to zero.

In VRPTW, there are  $m$  vehicles at the depot  $v_0$  to serve  $N$  customers,  $\sum_{k=1}^m \sum_{j=1}^n x_{ijk} \leq m$ , for  $i=0$ , each of them can service in a route by their capacity  $Q$ . In each route, each customer has their own demand  $d_i$  and the demand of all customers in the route must not exceed the vehicle capacity  $\sum_{i=1}^n q_i \sum_{j=0, j \neq i}^n x_{ijk} \leq Q$ , for  $k \in K$ . In the service, all vehicles start and end their routes at the depot,  $\sum_{j=1}^n x_{ijk} = \sum_{j=1}^n x_{jik} = 1$  for  $i=0$  and  $k \in K$  and each customer is served by only one vehicle  $\sum_{k=1}^m \sum_{j=0, j \neq i}^n x_{ijk} = 1$ , for  $i \in N$ ,  $\sum_{k=1}^m \sum_{i=0, i \neq j}^n x_{ijk} = 1$ , for  $j \in N$ . In addition, the summation of the arrival time at  $v_i$ , service time of  $v_i$ , travel time  $v_{ij}$  and the waiting time at  $v_j$  of each customer is not less than earliest time window  $e_j$  and not more than latest time window  $l_j$ ,  $e_j \leq s_i + t_i + t_{ij} + w_j \leq l_j$ .

### B. Fuzzy Membership Function

In the literature, fuzzy membership function has been applied to the problem in several papers. The fuzzy theory has been applied to the Chinese postman problem with time windows in 2002 by Wang and Wen [8]. They considered two cases in fuzzy time constrained of the problem. 1) the arrival time at a node may be fuzzy to describe the degrees of possibility and 2) the upper and lower bounds of time window constraints are uncertain. Their objectives are to minimize the total time to finish the tour and maximize the satisfaction level. Zheng and Liu [9] proposed a hybrid solution algorithm: they consider a vehicle routing problem in which the traveling times are assumed to be fuzzy variables. The triangular membership function based on the concept of fuzzy due time from Gupta et al. [10] is applied in this work.

### C. Evolutionary Algorithm (EA)

With regard to evolutionary algorithm (EA), the algorithm was first applied for the VRP by Homberger and Gehring [11]. After that, many EA have been brought to solve the problem and Genetic algorithm (GA) is the famous one. The basic

concepts of the algorithm are developed by Holland in 1975 [12]. The creation of a new generation of individuals involves primarily four major steps: initialization, selection, crossover and mutation. The two-phase hybrid of a genetic algorithm and an evolutionary algorithm was proposed by Bräysy et al. [13]. Genetic algorithm was applied in the first phase to find feasible solutions and then the best solution was produced by EA in the second phase. The hybridization of genetic algorithm and constructive heuristic was proposed by Berger et al. [14] in 1998. The nearest neighbor heuristic of Solomon [15] was used to create the initial population. Then, the genetic algorithm processes were applied to find the solution. The multi-criteria genetic algorithm was announced by Rahoual et al. [16] in 2001. The initial population is randomly generated as the 2-opt\* from Potvin and Rousseau [17], which is then applied in a crossover operation while random reinsertions of customers between routes is applied in mutation stage. Based on the rank of individual, the probability selection is done. The best found individual will be chosen to replace the worse one.

In this paper, the new methodology is applied to solve the VRPTW. The fuzzy technique is brought to deal with the initialization stage and then the evolutionary algorithm has been brought to deal with the improving stage of the solution. The remainder of this work is organized as follows: the methodology is demonstrated in section 2, the computational result is shown in section 3 and section 4 presents the conclusion.

## II. METHODOLOGY

This section aims explain the proposed algorithm in details. Each essential part of the proposed algorithm is described in the following sections.

### A. Solution Representation

Suppose  $n$  customers are visited by  $m$  vehicle routes.  $K$  is a set of vehicles;  $K = \{k_1, k_2, \dots, k_m\}$ . There are vectors of customer  $V_j^{k_i}; i = 1, 2, \dots, m; j = 1, 2, \dots, n$  and there are  $m$  number of  $v_0$  (depot) in the vector representing each solution as a representation of a VRPTW instance shown in Fig.1.

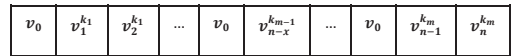


Fig. 1. Solution representation

The order of customers in each route ( $v_i$ ) after  $v_0$  is the sequence of customers served by a vehicle  $k_i$  as the following example.

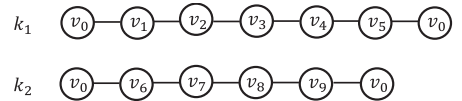


Fig. 2. Sequence of customers' representation

Each vehicle serves customers in its route in the same direction as shown in Fig. 2. The example in Fig. 2 represents customers in  $k_1$ . In this figure, number of customers  $n = 9$  and number of vehicle routes  $m = 2$ . Customer  $v_1, v_2, v_3, v_4$  and customer  $v_5$  are served by vehicle  $k_1$ . Customer  $v_1$  is the first customer visited by vehicle  $k_1$  and customer  $v_5$  is the last customer who is visited in this route before returning to the depot  $v_0$ . In vehicle  $k_2$ , there are customers  $v_6, v_7, v_8$  and  $v_9$  served by the vehicle sequentially.

### B. The Proposed Method

This work intends to combine an evolutionary algorithm with a fuzzy technique. Under this scheme, the evolutionary algorithm searches for a good ordering of customers, while the feasible solution construction is handled by the fuzzy membership function. After the initial population is produced in the first stage, the evolutionary algorithm is applied in order to improve the solution in the second stage. Each stage is described in details in this section.

In the initialization stage, the fuzzy technique has been introduced to construct effective feasible solutions or feasible initial route. The fuzzy theory of Zadeh [18] is considered to apply in order to produce the route. The triangular membership function has been brought to decide the position of the customer in the route. The details are described as follows:

In this work, the time service of customer is desired as the uncertainty variable. The service time satisfaction of customers will be desired by the concept of Gupta et al. [10] as the following formula.

$$\theta_i(s) = \begin{cases} 1; & e_i < s < l_i \\ 0; & \text{otherwise} \end{cases} \quad (3)$$

After that the triangular membership function (formula 4) is applied to choose the customer to the appropriate route.

$$f(x; a, b, c) = \begin{cases} 0, & x < a ; x > c \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \end{cases} \quad (4)$$

In the second stage or improvement stage, it aims to produce an efficient final solution. The main process in this stage is in an evolutionary form where the solution will continually improve itself; iteration by iteration. This work provides the specific selection method to choose the position for the operator. The important processes in this stage are described in the following work flow.

#### 1) Selection method: Select the position for operator

- Select the first point randomly.
- Apply the fuzzy method between the selected customer from a) and all other remaining customers.

- The customer who provide the best membership value will be selected to be the second point.

#### 2) Evolutionary process

In this step, the evolutionary concept is introduced in order to improve the population iteration by iteration.

The processes of the algorithm starting with the initialization stage whereas the initial population is produced. Each individual in the population will be evaluated and continued to take part in the evolutionary process where the parent will be selected by the roulette wheel probability selection method. Based on the results, the higher quality of the individual can get a better opportunity to be selected.

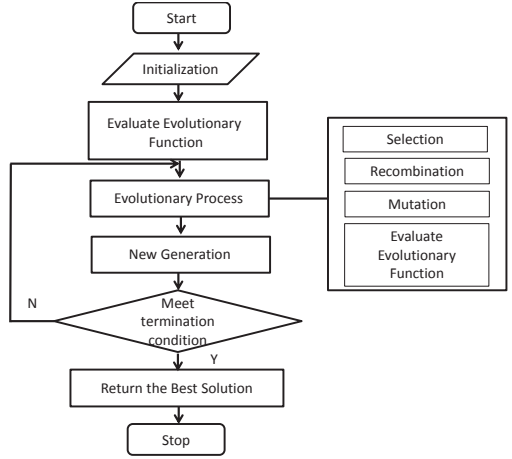


Fig. 3. The algorithm work flow.

However, the low quality individuals still have a chance to be chosen. This can help the improvement by not being stuck in the local optima. In the recombination process, the experiment applied a local search operator between the two selection parent, and in the mutation process which works in the internal selected parent. The new generation will be selected based on their quality. All processes will be continued until the termination condition is met and the best found solution will be returned.

### III. COMPUTATIONAL RESULTS AND ANALYSIS

The experiment is designed for the two stages of the proposed algorithm. In the first part, the initialization stage is tested on 6 problem sets of Solomon benchmark datasets [15] including the problem set in which their geographical points are generated randomly as R10x and R20x; problem set RC10x and R20x are the locations of customers that are mixed between the random and classify position. The last problem sets are C10x and C20x, the geographical of customers are classified by their geographical position.

The first part of the experiment aims to evaluate the performance of the proposed algorithm by comparing it to



other heuristics algorithms. The results from the comparative algorithms are taken from [19]. The cultivation number of vehicle, distance and computational time are compared to the other three route constructive heuristics; this is illustrated in Table I.

In Table I, the cumulative number of vehicles and total distance from Solomon benchmark datasets of the proposed algorithm are demonstrated in comparison with other 3 algorithms which are taken from Bräysy and Gendreau [19]. The result in Table I shows that the proposed algorithm provide the best cumulative number of vehicles. For cumulative number of total distance, the proposed algorithm provides a better result than [15] and [20]. It seems that the algorithm from [20] is not competitive with other approaches that are in this table.

TABLE I. THE CUMULATIVE NUMBER OF VEHICLES AND TOTAL DISTANCE OF 56 PROBLEM INSTANCES FROM SOLOMON BENCHMARK.

Problem Set	Author			
	[15]	[20]	[21]	Propose
R1	13.58	13.33	12.67	11.92
	1436.7	1509.04	1370	1210.86
R2	3.27	3.09	3.09	3.09
	1402.4	1386.67	1310	1351.91
C1	10.00	10.67	10.0	10
	951.9	1343.69	865	966.22
C2	3.13	3.38	3.13	3.13
	692.7	797.59	662	699.86
RC1	13.50	13.38	12.50	13.38
	1596.5	1723.72	1512	1717.04
RC2	3.88	3.63	3.50	3.63
	1682.1	1651.05	1483	1631.35
CNV/CDT	453	453	429	428
	73004	78834	67891	70483

In the second part of the experiment, the proposed algorithm is tested on Solomon problem set R10x (R101-R112). The computational results are in comparison with the other 9 evolutionary algorithms that are demonstrated in Table II.

TABLE II. THE CUMULATIVE DISTANCE OF PROBLEM SET R101-R112 FROM SOLOMON BENCHMARK.

Author	CNV	CDT	Time (mins)
[22]	12.75	1300.25	2
[23]	12.58	1296.83	25
[24]	12.58	1261.58	1-10
[25]	12.58	1272.34	17
[11]	11.92	1228.06	13
[11]	12	1226.38	19
[26]	12.42	1213.86	15
[27]	12	1217.73	-
[28]	12.08	1210.14	-
Propose	11.92	1210.86	1

Table II provides the results from the proposed algorithm and other 9 algorithms which are taken from [29]. In terms of the cumulative number of vehicle, it shows that the proposed method produce the best result. For the cumulative distance, which is compared to the other 9 algorithms, it has been found that the proposed algorithm returns with a better result than 8 algorithms.

In terms of time consumption (Time), rounded numbers of the time consumption in column 4 are taken from [29]. To find the results, the time consumption from [27] and [28] are not reported. The table shows that almost algorithms take more than 10 minutes to get the results. The genetic algorithm from [22] takes only 2 minutes while the proposed algorithm provides the best result (1 minute).

IV. CONCLUSION

This work aims to propose a new methodology to solve the vehicular network problem. The methodology contains two important stages: 1) the initialization stage and 2) the improvement stage.

In the initialization stage, the fuzzy technique was applied to deal with the uncertainty time windows constraint. In the improvement stage, there are two important processes known as the selection method and the evolutionary process. It aims to improve the initial population to be optimal. The proposed methods in this stage are tested with the Solomon benchmark datasets. The results found that the proposed method can produce an effective initial solution. After showing an improvement with the initial population, the improvement method produces better results than [15] and [20]: this is shown in Table I. The computational results in Table II show that the proposed method is very effective for problem set R10x (R101-R112). To make a comparison with the other 9 algorithms, the proposed algorithm produced the best cumulative number of vehicles. In terms of the cumulative number of distance, the algorithm from [28] provides the best value, however there was a lack of significant difference from the cumulative distance that was produced by the proposed algorithm. More significantly, the proposed algorithm spend only 1 minute to find the result. To compare the time consuming with other algorithms, the results in Table II illustrates that the proposed algorithm provides the best result.

ACKNOWLEDGMENT

The authors would like to acknowledge the School of Information Technology, Mae Fah Luang University, University of Phayao and the Ministry of Science and Technology for funding and all resources. The author is highly grateful to Dr. Ravee Phoehawm for the technical assistance.

REFERENCES

- [1] T.G. Crainic and G. Laporte, Planning Models for Freight Transportation, European Journal of Operational Research, 1997, Vol. 97, pp. 409-438.
- [2] J. Larsen, Parallelization of the Vehicle Routing Problem with Time Windows, Ph.D. thesis, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, 1999.
- [3] B.D. Backer, V. Furnon, P. Prosser, P. Kilby and P. Shaw, Local Search in Constraint Programming, Application to the Vehicle Routing

- Problem, in Proceedings of CP-97 Workshop on Industrial Constraint-Directed Scheduling, 1-15, Schloss Hagenberg, Austria, 1997.
- [4] B.I. Golden and E.A. Wasil, Computerized Vehicle Routing in the Soft Drink Industry, Operations Research, Vol. 35, pp. 6-17.
- [5] J-P. Rodrigue and T. Notteboom, The Geography of Transport Systems, New York: Routledge, forth edition, ISBN 978-1138669574.
- [6] K. Halse, Modeling and Solving Complex Vehicle Routing Problems, Ph. D. thesis, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, 1992.
- [7] G. B. Dantzig and R. H. Ramser, The Truck Dispatching Problem, Management Science, 1959, Vol 6, pp. 80-91.
- [8] H.F. Wang, and Y.P. Wen, Time-constrained Chinese postman problems. Computer Mathematic Application, 2002, Vol.44, pp. 375-387.
- [9] Y. Zheng, and B. Liu, Fuzzy vehicle routing model with credibility measure and its hybrid intelligent algorithm. Computer Mathematic Application, 2006, Vol. 176, pp. 673-683.
- [10] R. Gupta, B. Singh, and D. Pandey, Multi-Objective Fuzzy Vehicle Routing Problem: A Case Study. International Journal Contemp, Math. Sciences, 2010, Vol. 5(29), pp. 143-1454.
- [11] J. Homberger, and H. Gehring, Two evolutionary meta-heuristics for the vehicle routing problem with time windows, INFORMS Journal on Computing, 1999, Vol. 37(3), pp. 297-318.
- [12] J.H. Holland, Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, 1975.
- [13] O. Bräysy, J. Berger, and M. Barkaoui, A new hybrid evolutionary algorithm for the vehicle routing problem with time windows. Presented at the Route 2000-Workshop, Skodsborg, Denmark, 2000.
- [14] J. Berger, M. Salois, and R. Begin, A hybrid genetic algorithm for the vehicle routing problem with time windows. In Proceedings of the 12th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Springer-Verlag, Berlin, 1998, pp. 114-127.
- [15] M.M. Solomon, Algorithms for the vehicle routing and scheduling problems with time window constraints, Operations Research, 1987, Vol. 35(2), pp. 254-265.
- [16] M. Rahoual, B. Kitoun, M.H. Mabed, V. Bachelet, and F. Benameur, Multicriteria Genetic Algorithms for the Vehicle Routing Problem with Time Windows. Proceedings of the 4th Metaheuristics International Conference MIC'2001, Porto, Portugal.
- [17] J.Y. Potvin, and J.M. Rousseau, A parallel route building algorithm for the vehicle routing and scheduling problem with time windows, European Journal of Operational Research, 1993, Vol. 66(3), pp. 331-340.
- [18] L. A. Zadeh, Fuzzy sets. Information and Control, Vol. 8, 1965, pp. 338--353.
- [19] O. Bräysy and M. Gendreau, Vehicle Routing Problem with Time Windows, Parel: Route Construction and Local Search Algorithms, Transportation Science, 2005, Vol.39(1), pp. 104-118, DOI: 10.1287/trsc.1030.0056.
- [20] J.Y. Potvin and J.M. Rousseau, A parallel route building algorithm for the vehicle routing and scheduling problem with time windows, European Journal of Operational Research, 1993, Vol. 66(3), pp. 331-340.
- [21] G. Ioannou, M. Kritikos and G. Prastacos, A Greedy Look-Ahead Heuristic for the Vehicle Routing Problem with Time Windows, Journal of the Operational Research Society, 2001, Vol. 53, pp. 523-537.
- [22] S. Thangiah, Vehicle routing with time windows using genetic algorithms, In Application Handbook of Genetic Algorithms: New Frontiers, 1995, Vol. II, pp. 253-277, CRC Press, Boca Raton.
- [23] J.Y. Potvin and B. Samy, The vehicle routing problem with time windows part II: genetic search, Journal on Computing, 1996, Vol 8(2), pp. 165-172.
- [24] J. Berger, M. Salois and R. Begin, A hybrid genetic algorithm for the vehicle routing problem with time windows, In Proceedings of the 12th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Springer-Verlag, Berlin, 1998, pp. 114-127.
- [25] O. Bräysy, A new algorithm for the vehicle routing problem with time windows based on the hybridization of a genetic algorithm and route construction heuristics, Proceedings of the University of Vaasa, Research papers 227, Vaasa, Finland, 1999.
- [26] O.Bräysy, J. Berger and M. Barkaoui, A new hybrid evolutionary algorithm for the vehicle routing problem with time windows, Presented at the Route 2000-Workshop, Skodsborg, Denmark, August 2000.
- [27] L.M. Gambardella, E. Taillard and G. Agazzi. MACS-VRPTW: a multiple ant colony system for vehicle routing problems with time windows. In New Ideas in Optimization, pages 63-76, McGraw-Hill, London, 1999.
- [28] J.F. Cordeau, G. Laporte and A. Mercier. Unified tabu search heuristic for vehicle routing problems with time windows. Publication CRT-2000-03, University of Montreal, Montreal, Canada, 2000.
- [29] O. Bräysy , Genetic Algorithms for the Vehicle Routing Problem with Time Windows, Special issue on Bioinformatics and Genetic Algorithms, Dept. of Mathematics and Statistics, University of Vaasa, 2001.
- [30] P. Toth and D. Vigo, The Vehicle Routing Problem, Siam Monographs on Discrete Mathematics and Applications, Philadelphia: Siam, 2001.

# Primary Evaluation of a Software-defined Security Architecture for an IoT Environment

Ștefan-Ciprian Arseni, Alexandru Stancu, Alexandru Vulpe, Mădălina Oproiu, Simona Halunga, Octavian Fratu

Telecommunications Department  
University Politehnica of Bucharest  
Bucharest, Romania

{stefan.arseni, alex.stancu, alex.vulpe}@radio.pub.ro, {shalunga, ofratu}@elcom.pub.ro

**Abstract**—Communications have become an important enabler of Human-to-Human, Human-to-Machine or Machine-to-Machine interactions. With the rise of the Internet of Things (IoT) as a trending concept that enables everyone to interact with objects not only directly, but also from distance, while empowering smart objects to also communicate with each other, communication security has seen a growing interest from the industry and the scientific community. In order to provide support for the massive data transmission that will exist between IoT objects, another concept, Software-defined Networking (SDN), ensures a performance increase, throughout the network. From a security point of view, in an SDN-enabled IoT system, security bundles can be replaced with software applications integrated in the central SDN controller or with security protocols/standards, such as Software-defined Perimeter (SDP). In this article we address the implementation of two security elements (an Intrusion Prevention System application and an implementation of SDP), while assessing the first security element in the scenario of a DoS attack. A simple IoT topology has been implemented in mininet and several basic tests have been performed to observe the behavior of the IoT network in case of an attack.

**Keywords**—Security, Internet of Things, Software-define networks, Software-defined perimeter

## I. INTRODUCTION

Considered two of the most widely adopted and intriguing paradigms of the past few years, the Internet of Things (IoT) and Software-Defined Networking (SDN) concepts have become more and more intertwined, reaching a point where IoT seems to be efficiently enabled only by the use of various new networking concepts, such as SDN. These two concepts can be briefly described as:

- **IoT** represents an interconnection between physical elements from the surrounding environment and virtual services through which the control of those physical elements is enabled, therefore enabling users to interact with smart objects, even from a distance;
- **SDN** emerges as the new way of defining networks, by decoupling the control plane from the data plane, thus minimizing the requirement of different networking equipment, while empowering network administrators

to dynamically define and alter the network structure, without interfering with the work done by the users.

SDN introduces a novel approach on how rigid networks can be defined, making them more agile, by moving the entire decision making processes in a central controller, and also easier to manage, through different applications that can be implemented on top of the controller. Given its capability to dynamically change its structure, a software-defined network seems to be a key enabler for other concepts such as IoT or Cloud Computing [1].

The communication between the SDN controller and the forwarding elements, namely sensors, with regards to the IoT concept, is done through a standardized SBI (Southbound Interface). Several such SBI protocols exist and are accepted by the community, such as OpenFlow, NETCONF (Network Configuration Protocol), SNMP (Simple Network Management Protocol), or REST APIs. This implies that the network nodes of an IoT network need to support at least one of these interfaces, in order to consider that IoT architecture to be SDN-enabled.

Still, SDN brings some major security benefits when deployed in an IoT environment, such as eliminating the need of additional specialized equipment that not only need to be placed in specific points/boundaries of the IoT network (IPS, IDS, firewalls), but also introduce both an additional cost for the overall implementation of the network and a hindering of management activities executed by an administrator. To compensate the elimination of external security devices, SDN enables administrators to implement software applications in the central controller to mimic the functions of a security bundle, providing the same or even better performance, given that the controller offers an overview of the network and the ability to act in any node.

In [2] the authors present the implementation of an IPS application on an SDN controller that monitors the traffic over a switch and blocks any abnormal exceeding of a defined threshold. However, only DNS packets are monitored, and the topology used for simulations is very simple, containing only a switch and two hosts. Also it does not take into account the characteristics of an IoT network.

There are some examples, such as [3] and [4], which propose secure IoT architectures based on SDN that

interconnect multiple communication domains (wired, wireless or ad-hoc networks), while ensuring their security by setting up a distribution scheme for security rules to ensure that each domain will not be compromised. Another security perspective is introduced in [5] as a Black SDN architecture for IoT, in which both meta-data and payload are secured at each layer of an IoT communication protocol, while the central SDN controller is utilized as a trusted third party for enabling secure routing. Also, it demonstrates, through a simulation, the feasibility of the proposed architecture, primarily examining its behavior based upon the IEEE 802.15.4 standard for Low-Rate Wireless Personal Area Networks.

The main contribution of this paper is to define and validate a novel security architecture for IoT based on SDN, using Software-Defined Perimeter (SDP) flows. Starting from the common points that these two concepts have, we set out to securely integrate them into a combined security solution. Therefore, we present a new perspective on how security can be ensured for an IoT infrastructure based on SDN.

The paper is organized as follows: Section II presents a general IoT architecture based on SDN, while Section III introduces the security elements that are used to protect the previously mentioned IoT system. Section IV focuses on some performance tests conducted on the simulated IoT system and, finally, Section V briefly discusses the obtained results and concludes the paper.

## II. SECURE SDN-ENABLED IOT ARCHITECTURE

Multiple IoT architecture proposals have been made for complying with the demand of the industry, each one of them enabling the use of another concept, such as SDN. For example, in [6] authors make use of the dynamicity of software-defined networks to extend MINA (Multi-network Information Architecture [9]) with a layered SDN controller for IoT, in order to provide differentiated quality levels to tasks coming from a heterogeneous IoT architecture. Another point of view is mentioned in [7] where authors present the challenges that the centralized methods of software-defined networks introduce when providing global optimizations for an IoT architecture.

If we take into consideration that an IoT system would require the use of a Cloud system to manage and store data, a new concept is being revealed, Fog Computing, as a middle layer ensuring the connectivity between the numerous elements found at the edge of an IoT architecture and the central Cloud management system. In this idea, paper [8] introduces the Fog Computing concept, while SDN is presented as an enabler for it.

From the security perspective, both SDN and IoT present some vulnerabilities, but the advantages that these two concepts bring ease the process of creating a safe implementation of an IoT system. The major disadvantage of SDN is its centralized controller that acts as the logic manager of the entire network, becoming the major point-of-entrance for an attacker. Yet, there are other security threats for SDN, such as forged or faked traffic flows under a DOS attack. Regarding IoT, it brings multiple security risks that were emphasized, assessed and treated in papers such as [11] – [14].

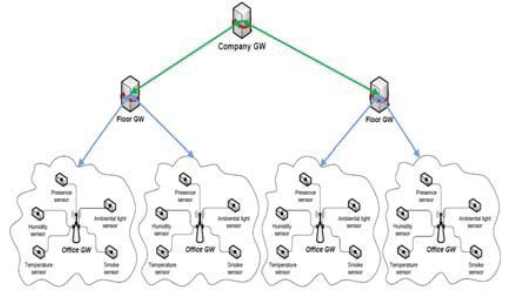


Fig. 1. IoT simple topology evaluated

The IoT architecture taken into consideration for the scope of testing the security elements has a simple topology, as presented in Fig. 1. It resembles an enterprise IoT system for controlling the environment in each office, while aggregating Office flows of data in Floor Gateways that, in turn, connect directly to the central Company Gateway. This topology was chosen because it allows an individual assessment of the state of each Office, while each Floor GW can integrate some security rules so that the data transmitted to the central GW will contain only valuable information. Each sensor is considered to be represented by a host, while the Floor Gateways are considered as switches, representing Open Virtual Switches in a real environment. For the central SDN controller, ONOS was chosen for several reasons, as described in [15].

For securing the considered IoT topology, two security elements were used. First, a simple IPS application was developed, starting from a sample application for ONOS, in which a new *deviceListener* was implemented, with the purpose of listening to the *PORT\_STATS\_UPDATED* event. Every five seconds, the controller polls, through the OpenFlow protocol, the port stats for every device that it controls and if a delta is present, meaning that traffic has passed through a specific device, the controller will generate the event of interest. The IPS application will then compute, for each of the Floor Gateway switches' ports connected to a host the amount of bandwidth it receives from the host, in kbps, and will compare it with a predefined threshold. If the threshold of 225 kbps is exceeded for a device, a flow rule is installed on the nearest switch, so that all traffic from that device is being dumped for a configurable amount of time. The threshold of 225 kbps was chosen based on an approximation of the amount of data normally routed in a sensor network, in which sensors report their status on a predefined schedule or only when important modifications arise.

## III. SOFTWARE DEFINED PERIMETER APPROACH

The analysis conducted in the remaining sections of the paper focuses mainly on the security of communication from the lower layers of the topology. Yet, the remaining layers are the ones that can be seen as the primary vulnerable breaching point in the company's IoT system from an outside attack. Having the possibility of customizing the security solutions

implemented at every layer of the topology, given the fact that it is software-defined, the upper layers of the network can be secured using the software defined perimeter (SDP) protocol.

SDP allows users to define air-gapped network, by replacing the hardware equipment with logical components that can be closely managed and configured, according to the need of separating services or network domains from the unsecured areas, such as the Internet, as described in [16]. Even though SDP is only a protocol, while SDN is a broader concept, the two of them have multiple common architectural elements that allow them to be easily integrated in a combined solution, such as in the case of the Company IoT topology presented in section II. The informational flow behind the SDP protocol [17] is presented in Fig. 2.

For the adaptation of the general workflow schematics to the presented IoT system, several equivalences need to be made:

1. SDP Controller can be emulated on the SDN controller that defines the entire network topology. In this way, there is no need of extra equipment to act as a hardware base for the SDP Controller, and, also, SDP complements the SDN by seamlessly integrating and taking advantage of the network topology.
2. SDP Hosts are, in this case, the Floor GWs, more specifically the egress communication enabled when transmitting data to the central node, the Company GW.
3. SDP Server is represented, in this scenario, by the Company GW, but it can consist of more than one node, as depicted in Figure 1.

By setting up the correct roles for each one of the intended equipment, the network can go “dark”, by disrupting any communication request that has not been previously authorized by the SDP Controller. This disruptive characteristic is what defines the defensive feature of the SDP protocol. The principle from which SDP was derived is fairly simple, as described in [18]: instead of allowing devices to access the network and send requests to servers, that can be identified using DNS, in a SDP, devices first need to be authorized and authenticated before initiating any request to a server.

Also, in an SDP environment, the Servers remain hidden, passively waiting for any incoming request, while the Hosts do

not have access to the network topology, until after the authentication and authorization stage, when the Host is informed of the active Servers that it can initiate communication with.

By applying this mixture of SDN and SDP in the upper layers of the IoT infrastructure, an increased level of persistence regarding any type of attacks is ensured. In this manner, the entire topology is protected, while allowing a certain degree of freedom for implementing various security methodologies in the lower layers of the topology, without causing any disturbance or creating any security breach.

IV. SYSTEM SETUP AND TESTING METHODOLOGY

We implemented the topology in Fig. 1 in the mininet network emulator. The links between the sensors and the Office Gateway were modeled having a bandwidth of 250 kbps, a delay of 5 ms and a packet loss of 3%. The links between the Office Gateway and the Floor Gateways were configured to have a bandwidth of 10 Mbps, a delay of 3 ms and a packet loss of 2%. Finally, between the Floor Gateway and the Company Gateway links with 100 Mbps bandwidth, a delay of 1 ms and a packet loss of 1% were allocated. Also, another host was created, and attached to the Company Gateway, representing a server that would collect the data provided by the sensors.

Another argument for choosing ONOS as an SDN controller for our simulation is its capability to allow administrators to dynamically load applications at runtime. This allows us to continuously adapt and re-integrate the IPS application, without having to reboot the SDN controller, therefore without interrupting the connectivity inside the network. For generating traffic inside the network, an Iperf3 [10] server was started on the host representing the cloud server and an iperf3 client was ran on each of the hosts representing the IoT sensors. UDP traffic was generated for 60 seconds, with a bandwidth controlled by the client. We considered that this specific period of time is enough for good statistical results. After that, when the client closes the connection, the iperf3 server provides statistics for the terminated connection, such as packet loss, jitter and bandwidth. For normal traffic conditions, a bandwidth of 125 kbps was considered for each host. For simulating a faulty or malicious sensor, that could cause a DoS attack on the server, a higher bandwidth was configured in the iperf3 client, namely 250 kbps.

In order to have a better understanding of how the network performance is modifying in different stages of an attack, three specific simulation scenarios were put in place:

- An evaluation of network performance under normal load;
- An evaluation of network performance when an attack takes place, without having the IPS application integrated in the SDN controller;
- An evaluation of network performance when an attack is identified and denied by the IPS application integrated in the SDN controller.

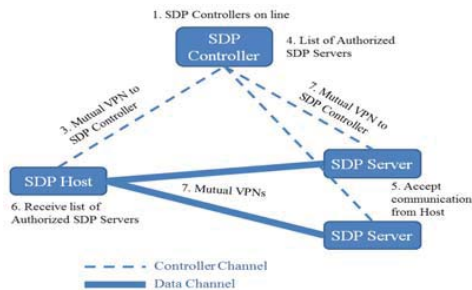


Fig. 2. Software Defined Perimeter Informational Flow

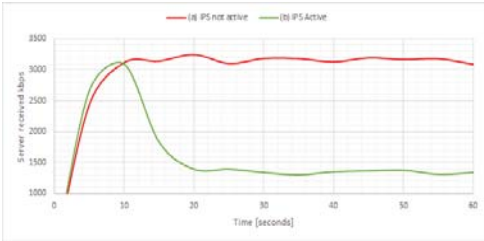


Fig. 3. Bandwidth received at the Server

The period of time needed by the application to detect the malicious traffic and react to it by blocking the respective traffic was also measured, in two ways: first by using ping from the compromised nodes to the Server, that sends an ICMP packet every one second, and then seeing when the connectivity between the sensor and the Server is lost. The second method was polling the Port Stats for the Company Gateway switch and seeing when the amount of traffic sent to the Server drops.

The results from the ping tests reveal the following information. Since we simulated a number of 20 sensors, from which 8 were considered malicious, we have the measured parameters from a set of 12 sensors in order to compute some statistics. Therefore, when comparing the average RTT and jitter measured in the second simulation with the same values that were used to build a benchmark in the first simulation, we observe that only in 25% of the sensors, the values increase slightly by approximately 10%. Thus, doubling the bandwidth used by the malicious sensors, as presented in Fig. 3, does not affect very much the flow of the ICMP packets. Nevertheless, when enabling the IPS application, these values decrease again to values close to the ones from the first simulation, meaning normal behavior. Since the ICMP traffic is not very affected, the variation in time of the jitter reported by the ping utility will also not be influenced in such a manner that could induce errors in the results.

After enabling the IPS application on the SDN controller, a noticeable improvement was observed for each characteristic that was tested in the previous two scenarios. The percentage of

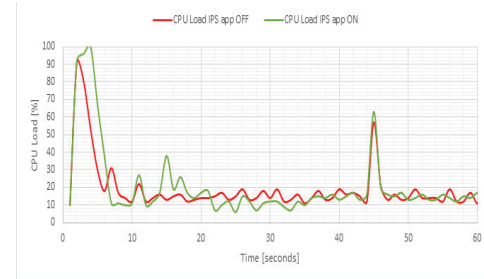


Fig. 4. CPU Load on the Server

sensors that are affected by increased packet loss decrease from 66% to 33%.

The results of the second method of detecting the period of time needed by the IPS application to block the malicious traffic are illustrated in Fig. 3, which depicts the bandwidth measured at the port of the Company Gateway connected to the Server. We observe that after approximately 10-11 seconds the traffic drops to normal values, meaning that the faulty sensors were removed from the network.

As a final evaluation criterion, the CPU load was measured on the Server. This was done using the *vmstat* tool offered by Linux. The CPU idle time percentage is offered every second by the tool. The values were collected in two situations: in the presence of malicious traffic with the IPS application disabled or enabled. The graphic of the CPU load variation in time is illustrated in Fig. 4.

### V. CONCLUSIONS AND FURTHER WORK

Even though the topology emulated was not very complex and the application implemented for the SDN controller was rather simplistic, only taking into consideration the bandwidth received at the ingress ports of the network, the experimental results allowed us to make some interesting observations.

First of all, we noticed that not all the sensors are influenced by the malicious traffic and not all the network parameters that were measured vary with respect to the presence of the malicious traffic in the network.

Also, since the variation of the RTT in time is rather the same, regardless of the increase in traffic sent by compromised nodes, it seems that the ICMP flows are not really affected by malicious traffic.

The amount of time needed by the application to identify and block the malicious traffic is adequate, because a DoS attack cannot be completed in such short time. Also, the application is based on the supposition that the sensors, in a normal operating regime, do not send more traffic than the threshold configured in the application. This may not be always be true, and false positives could appear, meaning that sensors that are neither malicious nor faulty are blocked from accessing the network.

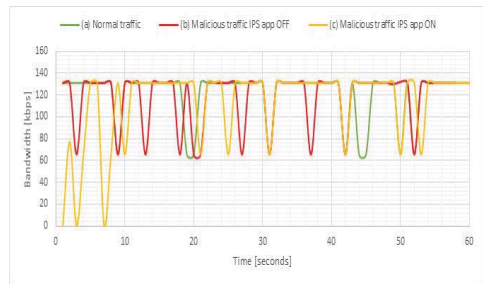


Fig. 5. Bandwidth variation in time for a sensor

Another research direction in this context would be creating an IPS application that would monitor not only the ingress Gateways, but also the other switches in the network. This would imply differentiating the traffic on the ports of the switches that are not ingress, based probably on the source IP address, and after deciding that the traffic is malicious, applying a flow rule on that port, which drops only the traffic with the specific source IP address.

ACKNOWLEDGMENT

This work has been funded by University Politehnica of Bucharest, through the "Excellence Research Grants" Program, UPB – GEX. Identifier: UPB–EXCELENȚĂ–2016 project Platform for Studying Security in IoT, contract number 96/2016 (PaSS-IoT), by UEFISCDI Romania under grants no. 20/2012 "Scalable Radio Transceiver for Instrumental Wireless Sensor Networks - SaRaT-IWSN" and 262EU/2014 eWALL support project and by the European Commission by FP7 IP project no. 610658/2013 "eWALL for Active Long Living - eWALL"

REFERENCES

[1] R. Vilata, R. Munoz, R. Casellas, R. Martinez, "Enabling Internet of Things with Software Defined Networking", CTTC, 2015

[2] Gökhan Akin, Enis Karaarslan, Ozan Bük, Erdem Uçar, "SDN architecture fundamentals and DOS prevention basics: a case study with openflow", International Scientific Conference "UNITECH 2015", Gabrovo

[3] Flauzac, Olivier; Gonzalez, Carlos; Hachani, Abdelhak; Nolot, Florent "SDN Based Architecture for IoT and Improvement of the Security", Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on, Gwangju, 2015, pp. 688 – 693, DOI 10.1109/WAINA.2015.110

[4] Flauzac, O.; Gonzalez, C.; Nolot, F. "Original secure architecture for IoT based on SDN", Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), 2015 International Conference on, Paris, 2015, pp. 1 – 6, DOI: 10.1109/NOTERE.2015.7293481

[5] Chakrabarty, Shaibal; Engels, Daniel W.; Thatapudi, Selina "Black SDN for the Internet of Things", Mobile Ad Hoc and Sensor Systems (MASS), 2015 IEEE 12th International Conference on, Dallas TX, 2015, pp. 190 – 198, DOI: 10.1109/MASS.2015.100

[6] Z. Qin, G. Denker, C. Giannelli, P. Bellavista and N. Venkatasubramanian, "A Software Defined Networking architecture for the Internet-of-Things," 2014 IEEE Network Operations and Management Symposium (NOMS), Krakow, 2014, pp. 1-9, DOI: 10.1109/NOMS.2014.6838365

[7] P. Thubert, M. R. Palattella and T. Engel, "6TiSCH centralized scheduling: When SDN meet IoT," Standards for Communications and Networking (CSCN), 2015 IEEE Conference on, Tokyo, 2015, pp. 42-47, DOI: 10.1109/CSCN.2015.7390418

[8] Flavio Bonomi, Rodolfo Milito, Preethi Natarajan, Jiang Zhu, "Fog Computing: A Platform for Internet of Things and Analytics", Big Data and Internet of Things: A Roadmap for Smart Environments, Springer, vol. 546 of the series "Studies in Computational Intelligence", 2014, pp. 169-186, DOI: 10.1007/978-3-319-05029-4\_7

[9] Z. Qin, L. Iannario, C. Giannelli, P. Bellavista, G. Denker and N. Venkatasubramanian, "MINA: A reflective middleware for managing dynamic multi-network environments," 2014 IEEE Network Operations and Management Symposium (NOMS), Krakow, 2014, pp. 1-4, DOI: 10.1109/NOMS.2014.6838332

[10] iPerf - The TCP, UDP and SCTP network bandwidth measurement tool [Online] Available at <https://iperf.fr/>

[11] Qi Jing, Athanasios V. Vasilakos, Jiafu Wen, Jingwei Lu, Dechao Qiu, "Security of the Internet of Things: perspectives and challenges", Wireless Networks (Springer), vol. 20, issue 8, November 2014, pp. 2481-2501.

[12] S. Sicaria, A. Rizzardina, L.A. Griecob, A. Coen-Porinisia, "Security, privacy and trust in Internet of Things: The road ahead", Computer Networks (Elsevier), vol. 76, January 2015, pp. 146-164

[13] Zheng Yan, Peng Zhang, Athanasios V. Vasilakos, "A survey on trust management for Internet of Things", Journal of Network and Computer Applications (Elsevier), vol. 42, June 2014, pp. 120-134.

[14] Kim Thuat Nguyen, Maryline Laurent, Nouha Oualha, "Survey on secure communication protocols for the Internet of Things", Ad Hoc Networks (Elsevier), vol. 32, September 2015, pp. 17-31

[15] Stancu, A.; Halunga, S.; Vulpe, A.; Suci, G.; Fratu, O.; Popovici, E.C., "A Comparison between several Software Defined Networking Controllers," 12th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS 2015), Niš, Serbia, October 14-17, 2015, pp. 223-226

[16] Cloud Security Alliance, "Software Defined Perimeter Working Group - SDP Specification 1.0", April 2014.

[17] Jim Reavis, "Software Defined Perimeter: Securing the Cloud to the Internet of Things", RSA Conference, Japan, 2014, Session ID: CDS-T08.

[18] Cloud Security Alliance, "Software Defined Perimeter Working Group – Software Defined Perimeter", December 2013

# DCDT: Degree-based Multi-hop Clustering Scheme with Directional Packets Transmission for EH-WSNs

Dan Liu\*, Yan Sun†, and Changchuan Yin\*

\* Beijing Key Laboratory of Network System Architecture and Convergence  
Beijing University of Posts and Telecommunications, China

†School of Electronic Engineering and Computer Science

Queen Mary University of London, London, UK

Emails: liudan10213188@bupt.edu.cn, yan.sun@qmul.ac.uk, ccyin@ieee.org

**Abstract**— Cluster-based scheme has been proved to be one of the efficient approaches to improve network performance for large scale sensor networks. In this paper, DCDT is proposed as a multi-hop clustering scheme, which operates in a directional packets forwarding manner in intra-cluster for energy harvesting enabled wireless sensor networks (EH-WSNs). DCDT moderates the boundary nodes effect without introducing extra overhead. With the attempt to balance the energy usage among CH and its CMs, adaptive transmission power allocation scheme is adopted for data forwarding among clusters. Verified on Opnet simulation platform, DCDT outperforms the compared schemes in terms of data transmission and energy consumption.

**Keywords**—cluster; energy efficient; wireless sensor network; node degree; cross layer.

## I. INTRODUCTION

In WSNs, routing mechanisms can be generally divided into two categories: flat routing and hierarchical routing. For the flat routing, all the nodes are deployed equally with the same functions, as discussed in [1]. The frequent route breakages and unpredicted topology changes make the flat routing less scalable as well as consuming more control overhead [2]. Cluster-based routing mechanisms, as the typical scheme under hierarchical routing, is recognized as the effective solution in terms of network scalability and energy efficiency [3], [4]. A wide variety of the clustering algorithms have been surveyed with various cluster constructions presented. In general, there are four common considerations in the cluster construction: the cluster head (CH) selection criterions, CHs rotation, the way of intra-cluster communication, and the inter-cluster data exchange. The most common metrics considered in CH selection are nodes' distribution density, distance, and residual energy [5], [6]. These three metrics are also considered for CHs rotation [13]. The clustered WSNs are classified by the number of hops from cluster member (CM) to its CH as single-hop and multi-hop. M-LEACH [7] is a typical multi-hop clustering algorithm which has a higher energy efficiency than LEACH from several aspects. Along with the further researches about multi-hop

clustering, intra-cluster routing schemes are investigated and explored in many aspect [8], [9], [10]. Instead of one hop direct transmission from CH to sink node, the routing schemes for multi-hop inter-cluster scenarios are presented and explained in work [11], [12]. However, the overhead caused by the signaling exchange for CHs is overlooked by the previous work. To our knowledge, a few related work have been done from this aspect. Besides, the boundary nodes effect, single-neighbor nodes acting as CHs without CM, is another issue worthy considering to improve the system efficiency.

Recently, EH-WSNs have attracted great interests due to its prominent data collection and gathering paradigm together with a couple of advanced features, such as the low-cost, portability and replenishing energy capability. Introducing energy harvesting enabled nodes into WSNs, the whole network lifetime is extended due to prolonged nodes' lifetime. However, the complex ambient environment causes big fluctuation and uncertainty in terms of the energy harvesting rate, which results in high indeterminacy for the nodes' available energy. Although work [16]-[19] have addressed the effects of the transmission power for the wireless network capacity and power saving in WSNs, the energy harvesting feature has been left out. Therefore, in this work, we propose a Degree-based Multi-hop Clustering Scheme with Directional Packets Transmission (DCDT) for EH-WSNs to overcome the overlooked issues as mentioned above.

## II. RELATED WORKS

Authors in [14] investigated and analyzed the effect of cluster structure on the average routing overhead for the wireless ad hoc network where nodes are distributed along a line. The numerical results show the tradeoff between smaller clusters and fewer clusters. Unfortunately, compared with the previous clustering schemes the boundary node effects result in quite a few independent CHs. For overcoming the boundary node effects, authors in [15] proposed a RGCA clustering scheme, which combines the relative neighborhood graph (RNG) with green cluster algorithm (GCA). To build the RNG, the adaptive transmission power is applied to the whole network based on the network topology built in RNG. After the RNG is built, GCA is conducted for constructing a cluster network topology with

---

This work was supported in part by the NSFC under Grants 61271257, 61629101, and 61671086, and in part by the Beijing Municipal Natural Science Foundation under Grant 4122034.



balancing the cluster numbers (CM) among clusters. M-LEACH in [7] was proposed to achieve higher energy efficiency. Instead of direct communication among CH and CM via high transmission power, the multi-hop communication for intra-cluster seems to be more cost effective and reliable. Authors in [20] have pointed out that the one-hop long transmissions would cause more clusters, more CHs, more control message and more energy consumption for large scale networks. Again, the boundary nodes effect is not described and addressed in these works.

In this work, the nodes' degree based cluster construction scheme (DCDT) with adaptive transmission power control among clusters is proposed to ease the boundary node effects and enhance the intra-cluster communications to improve network performance. With the knowledge of neighbors' degree, packets can be sent directionally to the boundary node and forwarded to the CHs without extra control messages involved. As for the intra- and inter-cluster communication, transmission power control is executed based on the point-to-point distance and network's propagation loss, intended to diminish the energy consumed in the cluster reconstruction. The details will be presented in the following sections.

### III. PRELIMINARIES

#### A. Network model

The total number of the sensor nodes in network is  $n$ , and all nodes in the considered network are configured with same capabilities for data collecting, energy harvesting, data processing, packets transmission, and the energy and packet buffer. The node IDs are always uniquely allocated for every node in network, denoted by the numbers  $1, 2, \dots, n$ . Moreover, all the nodes have the same transmission range at first. And they can adapt the transmission power later once they get instruction from their CHs. For a node  $v$ , the node degree of  $v$ ,  $deg(v, t)$  is the number of nodes that  $v$  can directly communicate with at time  $t$ . The node with  $deg(v, t)=1$  is called the boundary node. Fig. 1 shows a network with 24 nodes and 9 boundary nodes. Moreover, the directional transmission can be categorized to downwards and upwards transmission. Downwards transmission means that the message exchange from higher degree node with to lower degree node till no smaller degree is found. Similarly, if the transmission started from lower degree node with to higher degree node it is upwards transmission.

#### B. Node Model

To model the sensor node accurately in EH-WSNs, a time slotted system designed in [21] is adopted in this paper.

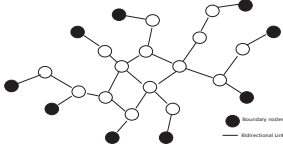


Fig. 1. Example for Network Model

Meanwhile, the common stochastic process is employed to model the energy harvesting amount, and the independent and identically (i.i.d) Bernoulli distribution for packets and energy creation is applied in this work. The probability for the packets and energy creation are  $p_l$  and  $p_e$  respectively. Packets creation and energy harvesting are two independent process. For a sensor node, there are two energy sources, the initial energy and the harvesting energy. Energy harvesting is an accumulation process which has a strong relationship with the energy source nearby.

Energy buffer and packet buffer are used to store available energy (residual energy and harvested energy) and packets (receiving packets and generated packets) respectively. To transmitting a packet, energy is scheduled and allocated in the node for corresponding transmission. The energy needed for transmitting or receiving a packet with size  $L$  is modeled as in [21]:

$$E_{tx} = L \times E_{ele} + L \times \varepsilon \times d^2, \quad (1)$$

$$E_{rx} = L \times E_{ele}. \quad (2)$$

Where  $d$  is the distance between the two sensor nodes and  $E_{ele}$  is the energy consumed by the node to perform the basic tasks, such as modulation, data processing etc.  $\varepsilon$  is a constant value according to acceptable signal-to-noise ratio (SNR) at the receiving node. And we also assume that the energy stored for receiving packets is always adequate.

The next section will give details about how the proposed scheme operates.

### IV. LOCAL CLUSTER CONSTRUCTION

For constructing the cluster efficiently, three different types of HELLO packets are defined in this part: the normal HELLO packet which contain its node ID and position, HELLO packet with the node's degree and the HELLO packet carrying cluster header ID. The functions of the three types of HELLO packets will be explained in the following section step by step.

#### A. CHs Selection

Two steps are designed to select CH:

##### 1. Step One: Degree number Collection

In this step, the normal hello is adopted. Along with frequently HELLO packets' exchanging, all the node in the network will generate its neighbor list by checking the source ID in normal hello packets sent by its neighbors and calculate the total number of its neighbors.  $Deg(i, t)$  is employed here to denote the neighbor number of node  $i$  at time  $t$ .

##### 2. Step Two: CHs Declaration

For finding the node with largest node degree,  $Deg(i, t)$  is inserted into HELLO to form the 2<sup>nd</sup> type of hello message. When the nodes receive the 2<sup>nd</sup> type of hello messages, it will compare all neighbors' degree with its own degree value. If its own degree is larger or equal to all its neighbors'  $Deg(i, t)$ , it will declare itself as a CH. For example in the following scenario, shown in Fig. 2,

7 CHs are claimed as filled black nodes and numbered from 1 to 7. For CH 6, its two neighbors have the same degree as 2, so it announces itself as a CH. Same principle is applied to the other 6 CHs.

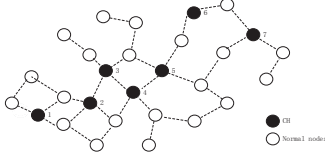


Fig. 2. Example for CHs Declaration

When the CHs are declared respectively, the next phase is to organize the corresponding cluster.

### B. CMs Notification

After CHs are self-declared, the cluster member notification packets (CMN) will be generated by CHs and sent its members. The node that receives the CMN from a CH will mark itself as a cluster member associating to its CH and forward received CMN through downwards transmission. In order to build up intra-cluster upwards transmission, the CH ID is carried in CMN packets together with the last hop ID inserted by the node who processed the CMN. For instance, the arrows shown in Fig. 3 indicate the forwarding directions of the CMN of CH 3 and CH 5.

If a node receives more than one CMNs, it flags itself as a shared CM. **Algorithm 1** is designed for the shared CMs to select only one CH from multiple potential CHs. The nodes with shadow in Fig. 4 are the shared CMs. For simplicity, only the first node along the path who receives multiple CMNs is marked as shared CM. All the down streaming nodes of shared CM will follow the same CH as its up streaming node.

#### Algorithm 1 Multiple Potential CHs Selection

**Define:** Hop number between CM to CHs:  $HN_{CHID}$ .

Energy harvesting Rate:  $ER_{CHID}$ ;

Cluster member number for CHi:  $NN_{CHIDi}$

**Input:**  $HN_{CHID1}, HN_{CHID2}; NN_{CHID1}, NN_{CHID2};$   
 $ER_{CHID1}, ER_{CHID2};$

**Begin**

1. **IF**  $HN_{CHID1} \neq HN_{CHID2}$
2. Select the CH with small  $HN_{CHID}$  ;
3. **Else**
4. **IF**  $\frac{ER_{CHID1}}{NN_{CHID1}} = \frac{ER_{CHID2}}{NN_{CHID2}}$
5. Select the CH with larger  $\frac{ER_{CHID}}{NN_{CHID}}$
6. **Else**
7. Select the CH whose CMN received first;
8. **End else;**
9. **End if;**
10. **End else;**
11. **End if;**
12. **Return** CHID;



Fig. 3. Example for CMNs Spreading

### C. CH Notification and Intimate Node Assignment

When the cluster boundary nodes receive the CMN packets, it will generate a CH inform (CMI) message, which carrying on its own ID as well as neighbor CH ID (if its neighbor belonging to other CH), then send it to its CH via upwards transmission. Every node along the upwards transmission will add its ID to the CMI before forwarding the CMI message back to the CH. Besides, CH's neighbor nodes will also add their location and energy information into the CMI before forwarding to CH. When the CH receive the CMI, it will add the information carried by CMI into CM table. Meanwhile, certain non-cluster boundary nodes of the CHs are marked as intimate nodes. The function of intimate nodes is to act as the relay nodes of CH to communicate with other clusters when the energy of the CH decrease to the threshold value  $E_{th}$ . Till now, each CH will have a full picture of its cluster structure.

Three tables are created in each cluster CH, including the CM table, which contains upwards transmission routes, the neighbor cluster table recording neighbor CHs and the intimate nodes table. The AODV routing scheme is applied to the inter-cluster communication. The data exchange within a cluster always follows the upwards or downwards transmission. The data transmission among CHs is handled by the allocated intimate nodes of CH when energy of CH is less than the  $E_{th}$ . Meanwhile, the adaptive transmission power is applied to the inter-cluster data transmission with the knowledge of the transmission distance.

### D. Analysis of the Cluster Maintenance Overheads

For an EH-WSN which has  $n$  sensor nodes and  $j$  clusters, when the exchange packet length is  $M$ , the energy required for the intra- and inter-cluster communication is formulated below. Since the inter-cluster communication maintenance overhead under AODV scheme is zero, the total energy for DCDT cluster maintenance is

$$E_{total} = E_{intra}, \quad (3)$$

and

$$E_{intra} = E_{CH} + E_{CM}, \quad (4)$$

$$E_{CH} = \sum_{i=1}^j (M \times E_{CH(i)}) + \sum_{i=1}^{n-m_b} [m_i \times M \times E_{unit}], \quad (5)$$

(i is a CH) (i is the non boundary CM)

$$E_{CM} = \sum_{i=1}^n (i \text{ is a CM}) (m_i \times M \times E_{unit}). \quad (6)$$

where  $E_{CH}$  is the total energy consumed for CHs declaration and CMs notification, and  $E_{CM}$  is energy used for CHs notification,  $m_b$  is the number of the boundary nodes for the network and

$E_{unit}$  is the unit energy required for one-hop communication among nodes.

#### E. Adaptive Transmission Power Allocation

For the packet exchanging among clusters, the adaptive transmission is allocated for the energy saving. Friis free space propagation model is applied by authors in [15] to build the RNG with adaptive transmission range. In this work, Friis free space propagation model is still adopted as the basic propagation model. The received power can be obtained by (7), in which  $P_t$  is the transmission power,  $G_t$  and  $G_r$  are antenna gains for the transmitter and receiver respectively.  $\lambda$  is the wavelength of electromagnetic waves, and  $d$  is the distance between transmitter and receiver. It is the key point to find the distance between the transmitter and receiver when the required received signal strength is learnt.

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \quad (7)$$

In this scheme,  $P_r(d)$  and distances of the inter-cluster data forwarding nodes are included to the control packets in the routing construction stage for realizing the adaptive transmission power allocation.

### V. SIMULATION RESULTS

In this work, OPNET is adopted as the simulation platform to evaluate the performance of DCDT. As RGCA in [15] is also targeting to solve the boundary node effects and adopting adaptive transmission power, it is implemented to compare with DCDT. MLEACH in [7] is also simulated for performance comparison. For RGCA, the transmission range is adjustable for all types of message after the RNG is built. While the adaptive transmission power is applied for routes construction and data exchange under DCDT. The simulation parameters are listed in the **Table I**. Configurations are the same for all these three schemes unless explicitly addressed. Ten random seeds are used for each simulation scenario and each call lasts for 5 minutes. For clarity, only the mean values are shown in this section.

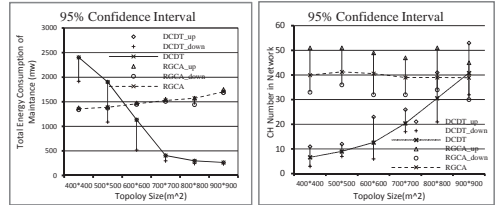
**Table I** Simulation Parameter Settings

Parameters	Settings
Nodes Location	Random
Number of nodes(n)	100
DCDT Transmission Range in intra-cluster (m)	70
DCDT Transmission Power in intra-cluster	0.1
RGCA Construction Transmission Power(mw)	0.5
Propagation Loss(dB)	124.5
Packet length $l$ (bits)	400
$p_e, p_t$	0.5, 0.5

#### A. Energy Consumption for Cluster Maintenance

In the first scenario, 100 nodes with  $3w$  initial energy each are distributed randomly in different sizes of the network topologies to investigate the energy consumption in cluster maintenance. The

network size increases from  $400 * 400m^2$  to  $900 * 900m^2$ . The total power consumption for the cluster maintenance is studied. Since periodic re-clustering is adopted by MLEACH and the CH number can be control by adjusting the priori (the desired percentage of cluster heads), to conduct a fair comparison, only DCDT and RGCA are simulated for energy consumption and total CH number comparison. Results are shown as 95% confidence interval in Fig. 4.



**Fig. 4.** Energy consumption and CH number for cluster maintaining

Due to the increasing of network size, the distance among nodes is greater while the nodes density is decreasing. The energy used for RGCA cluster construction increases slightly for all topologies scenarios as shown in Fig. 4. For DCDT, when the network size increases, the number of clusters has to increase to cover the wider area. Therefore the corresponding CMN and CMI packets are dropping due to less CM nodes in each cluster, resulting in less energy consumption and large variation. Regarding the CH number generated in the network, both scheme have big variations in large size networks. However, DCDT is less variable in the small size network topology, which is due to the high node density.

#### B. Energy Consumption in Data transmission

The energy consumption of the data transmission is studied under 10 random network topologies with 100 nodes distributed within  $500 * 500m^2$  network. The initial energy value for nodes remains as  $3w$ . Considering the network topology in Fig. 5, three schemes are executed respectively to construct the clusters. Every node in the scenario will send data to sink node, node\_47, if packets are created. The data created by node\_0 is measured at the sink node. Similarly, every scenario is repeated 10 times with different seeds set. The routes and final clusters of three schemes are also illustrated in Fig. 5. The shortest route selection algorithm is applied for inter-cluster data exchange under RGCA and DCDT schemes. For the MLEACH, it re-clusters every 150 seconds and the one-hop transmission is adopted by CHs to communicate with the sink node.

Fig. 6 and Fig. 7 show the performance of three schemes in terms of the data transmission from three different aspects. Three schemes achieve almost the same throughput. However, in terms of the end-to-end delay, RGCA is much higher than DCDT due to more hops along the route for data transmission. Since the one-hop transmission is applied to the data transmission, MLEACH shows the least end-to-end delay value. As for the energy

consumption during data transmission, DCDT keeps in a low and stable state comparing to the other two schemes.

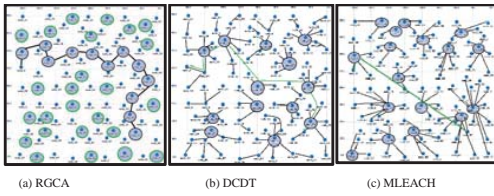


Fig. 5. 500 \* 500m<sup>2</sup> topology for network performance

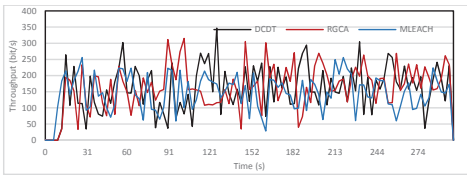


Fig. 6. Data received from node\_0 by sink node

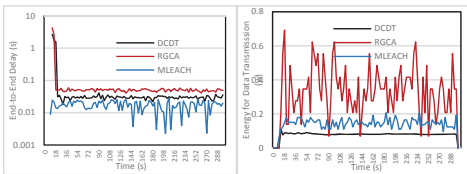


Fig. 7. End-to-End delay and Energy consumed in node\_0 of data transmission

VI. CONCLUSIONS

This paper proposes one energy efficient multi-hop cluster construction scheme, DCDT, for energy harvesting enabled WSNs. In DCDT, CHs are selected strictly by the node degree to carry on the directional transmission within the cluster. Besides, for diminishing the energy used for periodical re-clustering, the intimate nodes act under the instruction of its CHs to transmit data to outside clusters. Meanwhile, the adaptive transmission is applied for the inter-cluster communications. To evaluate the new scheme, the RGCA and MLEACH are implemented in this paper to compare the performance of the transmission quality and energy usage. Simulation results indicate that the DCDT outperforms others.

REFERENCES

[1] M. Zhang and P. H. J. Chong, "Performance comparison of flat and cluster-based hierarchical ad hoc routing with entity and group mobility," in *Proc. Wireless Communications & Networking Conference (WCNC)*, Budapest, pp. 2450-2455, Apr. 2009.

[2] R. Pasquini, F. L. Verdi, M. F. Magalhães, and A. Welin, "Bloom filters in alandmark-based flat routing," in *Proc. International Conference on Communications (ICC)*, Cape Town, South Africa, pp. 1-5, May, 2010.

[3] Chang S.Y, Wu H.-C., Wu Y. , and Chao H.-C., "Theoretical analysis for tree-like networks using random geometry," *Communications, IET*, vol.5, no.15, pp.2167-2176, Oct. 2011.

[4] X. Niu, Z.Tao, G. Wu, C. Huang and L. Cui, "Hybrid cluster routing : an efficient routing protocol for mobile ad hoc networks", *International Conference on Communications(ICC)*, Istanbul, pp. 3554-3559, Jun. 2006.

[5] Q. Liang, "Cluster head election for mobile ad hoc wireless network," in *Proc. Personal, Indoor and Mobile Radio Communications( PIMRC)*, pp. 1623-1628, Sept. 2003.

[6] I. Gupta, D. Riordan and S. Sampalli, "Cluster-head election using fuzzy logic for wireless sensor networks," *Communication Networks and Services Research Conference (CNSR)*, pp. 255-260, May 2005.

[7] V. Mhatre and C. Rosenberg, "Homogeneous vs heterogeneous clustered sensor networks: a comparative study," in *Proc. International Conference on Communications (ICC)*, pp. 3646-3651, Jun. 2004.

[8] L. Cheng, S. K.Das, M. D. Francesco, C. Chen and J. Ma, " Scalable and Energy-Efficient Broadcasting in Multi-hop Cluster-Based Wireless Sensor Networks," in *Proc. International Conference on Communications (ICC)*, Kyoto, pp. 1-5, Jun. 2011.

[9] L. Xu, M. J. O'Grady, G. M. P. O'Hare and R. Collier, "Reliable multihop intra-cluster communication for Wireless Sensor Networks," in *Proc. Computing, Networking and Communications (ICNC)*, Honolulu, HI, pp. 858-863, 2014.

[10] B. Guizani, B. Ayeb and A. Koukam, "A Stable K-hop clustering algorithm for routing in mobile ad hoc networks," in *Proc. International Wireless Communications and Mobile Computing Conference (IWCMC)*, Dubrovnik, pp. 659-664, Aug. 2015.

[11] C. H. Hsu and K. T. Feng, "On-demand routing-based clustering protocol for mobile ad hoc networks " *Personal, Indoor and Mobile Radio Communications( PIMRC)*, Athens, pp. 1-5, Sept. 2007.

[12] L. Conceição, D. Palma and M. Curado, " A novel stable and low-maintenance clustering scheme," *ACM Symposium on Applied Computing (SAC)*, New York, NY, USA, pp. 699-705, 2010.

[13] D. Kumar, T. C. Aseri, and R. B. Patel, " EEHC: energy efficient heterogeneous clustered scheme for wireless sensor networks," *Computer Communications*, vol. 32, no. 4, pp. 662-667, 2009.

[14] K. Abboud and W. Zhuang, " Impact of node clustering on routing overhead in wireless networks", *Global Telecommunications Conference (GLOBECOM)*, Houston, TX, pp. 1-5, Dec. 2011.

[15] C. Tseng, L. Chang, H. Wang, F. Kuo, K. Ting, and D. Deng, "Analysis of energy consumption for the clustered topology constructed by combining RNG and GCA," in *Proc. International Conference on Communications(ICC)*, Ottawa, ON, pp. 5245-5258, Jun. 2012.

[16] P. Gupta and P. R. Kumar, " The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. IT-46, pp. 388-404, 2000.

[17] A. Castagnetti, A. Pegatoquet, T. N. Le, and M. Auguin, "A joint duty-cycle and transmission power management for energy harvesting WSN," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 928-936, May 2014.

[18] J. Deng, Y. Han, P. Varshney, "Optimal transmission range for wireless ad hoc networks based on energy efficiency," *Communications, IEEE Transactions on*, vol. 55, no.9, pp. 1772-1782, Sept. 2007.

[19] J. Gomez and A. T. Campbell, "Variable-Range Transmission Power Control in Wireless Ad Hoc Networks," in *IEEE Transactions on Mobile Computing*, vol. 6, no. 1, pp. 87-99, Jan. 2007.

[20] V. Mhatre, and C. Rosenberg, "Homogeneous vs heterogeneous clustered sensor networks: a comparative study," in *Proc. International Conference on Communications(ICC)*, pp.3646-3651, Jun. 2004.

[21] D. Liu, Y. Sun, and C. Yin, "QoS-Aware Transmission Management for Energy Harvesting Wireless Sensor Networks," in press. International Conference on Telecommunications(ICT), May, 2016.

[22] S. Peng, and C. P. Low, "Energy neutral routing for energy harvesting wireless sensor networks," in *Proc. Wireless Communications and Networking Conference (WCNC)*, Shanghai, pp. 2063-2067, Apr. 2013

# Effect of GPS errors on Emission model

Anders Lehmann\*, Allan Gross†

\*Aarhus School of Engineering University of Aarhus  
anders@ase.au.dk

†Department of Business Development and Technology  
University of Aarhus  
agr@btech.au.dk

**Abstract**—In this paper we will show how Global Positioning Services (GPS) data obtained from smartphones can be used to model air quality in urban settings. The paper examines the uncertainty of smartphone location utilising GPS, and ties this location uncertainty to air quality models. The results presented in this paper indicates that the location error from using smartphones is within the accuracy needed to use the location data in air quality modelling. The nature of smartphone location data enables more accurate and near real time air quality modelling and monitoring. The location data is harvested from user data gathered in the wild.

## I. INTRODUCTION

Air quality can have serious health effects, both immediate and longterm. In many countries limit values for different pollutants are imposed, to keep the health consequences to a minimum. To monitor that the air pollution is well below these limit values, and to be able to alert the affected communities when concentrations are above the limits, extensive air quality models have been developed. These models take different emission sources like long range transport of pollution from regional or transnational sources, urban pollution sources, and emission from traffic into account. All these pollution sources are considered to estimate the influence on the street level air quality. The effects of the weather is also factored in as wind patterns in street canyons will increase the pollution on the lee side of the street and temperature and precipitation will change the rate of chemical reactions and deposition of pollutants.

Air quality in cities is dependent on a number of factors. Pollution can travel for long distances, and thus affect the air quality in a distant city, but often more important are sources local to the city. The local sources of pollution affecting the city air quality are for instance local residential and district heating, local processing industry and traffic. We will in this paper focus on how we can improve the modelling of traffic induced emissions, and thus the traffic related effects on air quality, through data obtained from smartphones used by users of the urban traffic system. Figure 1 shows the different levels of the pollution model.

To model the impact of traffic on air quality the existing models use traffic counts to estimate the traffic. The traffic counts are typically points measures done rather seldom and for intervals of a few weeks or months. Ordinarily traffic counts only gives information on the number of vehicles passing through the measuring gates, and thus the speeds

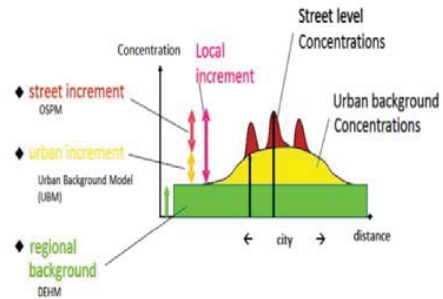


Fig. 1. The pollution in a street is a combined from regional, urban and local sources

of vehicles are not measured. Only a few streets in a city is exposed to traffic counts, and the traffic in the remaining streets are estimated from traffic models.

In this paper we consider the quality of data collected from smartphones of people moving in an urban transport system, and if this data can be used as a dynamic supplement to the more static traffic counts. We identify two issues to investigate. First the uncertainty of location measurements based on Global Positioning Services (GPS), which will influence how to count vehicles on the streets. The second issue is the uncertainty on the speed measurement from smartphones, which will influence the outcome of the level of emissions estimated from the model.

The data used in this paper is collected in deployments of the project EcoSense, which have provided smartphone apps for different projects promoting sustainable energy, energy savings, sustainable transport and climate change mitigation efforts.

The paper is first considering the available literature on the subjects of air quality and location accuracy in section II. The experiments are presented in section III. Section IV presents the results of the experiments. The paper ends with a discussion of the contributions and results, as well as an outlook towards future work.

## II. RELATED WORK

Air quality modelling in urban areas has been a scientific field since the privatisation of urban transport became abundant. The large concentration of cars in city centres, leading to serious health issues, has spurred the evolution of methods to monitor the level of pollution. The active measurements of pollution led to an understanding of the sources of the pollution, but was too costly to be deployed in more than a few places in each large city. Modelling approaches were considered to be put in place of active measurements for the rest of the urban areas. One of the more thorough models for urban air quality is the Operational Street Pollution Model [1], [2]. OSPM combines results from Long Range Transport models (i.e. Danish Eulerian Hemisphere model [3]) for regional pollution sources, with Urban background models for urban pollution sources and a street canyon model for local sources and variations due to turbulent winds, to calculate the pollution concentration at the street level. The model is used to monitor the adherence to the limit values set up by national and international regulating bodies [4].

The OSPM model uses an extensive amount of data from diverse sources in order to estimate the emission of pollutants, the mixture and reactions of pollutants, and the transport of pollutants. The data sources stems from traffic counts, traffic surveys, traffic models, fleet information, and meteorological data.

Since the advent and ubiquity of smartphones, many projects have utilised the new opportunities for data collection in the wild. The big problem, from an academic point of view, is that it can be hard to persuade users to submit their data to a research community. Some research projects have used the smartphone location data indirectly, by using location data included in status updates for social media [5], [6]. Other projects have used cell tower data from network operators, to coarsely trace movements of individual mobile phone users [7]. Others, like in this project, have succeeded in using data directly from smartphones to perform their research studies [8]–[10].

The GPS system is well documented and very useful [11]. The location and speed accuracy of non-military commercial receivers has been investigated in different settings. In Witte and Wilson [12] GPS accuracy is measured by a bike driving on an athletics running track. The GPS receiver used had an external antenna placed on the bikers helmet, which is different from a smartphone with a builtin antenna placed in a pocket or bag. The paper uses the placement of the antenna to explain the drop in location accuracy during curves, since the bikers head will be at another place than the bike, due to the leaning necessary for keeping the balance on the bicycle while turning.

In Modsching et al. [13] the GPS location accuracy from four different GPS receiving devices is investigated in an urban environment. The authors report high error rates due to buildings shadowing the radio signals from the GPS satellites, and provide map-matching as a tool to improve the location accuracy. Furthermore the paper finds that there is no corre-

lation between actual location error and device reported error level.

A thorough investigation of error statistics of GPS location for GIS applications, is given in Zandbergen et al. [14]. The error statistics are measured by a mapping grade GPS receiver in a fixed outdoor position over eight hours. The mean error is reported to be 2.4 meters.

To improve the accuracy of GPS location and speed estimation from smartphone data, the paper by Nitsche et al. [15] employ a Kalman filter on the raw GPS data to reduce the location uncertainty. The presented experiments show how the Kalman filter removes outliers, but also shows residual inaccuracy when the filtered GPS data is drawn on a map.

## III. EXPERIMENTS

A series of experiments was conducted, to investigate the error behaviour of GPS in smartphones.

The first experiment concerns stationary indoor accuracy to test experienced system accuracy, without introducing possible distortions from mobility.

The second experiment is about experienced accuracy of moving GPS receivers in smartphones. Data has been collected from smartphones of different users and the reported uncertainty of the GPS location were evaluated.

Lastly an experiment was made to investigate the sensitivity of the Air Quality model to the average speed of the modelled traffic.

### A. Stationary GPS accuracy experiment

The GPS system involves 27 satellites (24 is the designed number of satellites, the rest are in reserve but operational) in six different medium earth orbits at 20.000 km above earth. The orbit is chosen in such a way that each satellite will trace the same area of the earth two times each day. The recurring events of satellites appearing and disappearing over the horizon, could lead to recurring errors in the localisation of a GPS receiver.

To test this hypothesis we recorded the location and error estimate of a smartphone placed in a fixed indoor location, over a period of 5 weeks. The fixed location was chosen to minimise the effects of changing environments, which could cause location errors from multi-path radio propagation, shadowing or other degradation of the received radio signals. The indoor location was chosen to ensure the integrity of the smartphone and to keep it powered during the experiment. The indoor location may also cause some extra damping of the radio signals from the satellites, which can influence the accuracy of the location estimation. The location was chosen to be a single story home in a residential area. The location and horizontal and vertical error were sampled every second.

The experiment was made in two variants - one with Wifi connected to the local Wifi network, and one with the wifi radio turned off. As the smartphone reported location accuracy is determined from the combined localisation the GSM network, Wifi networks and GPS localisation, the two variants were made to test if the smartphone connected to a

TABLE I  
STATISTICS FOR THE "HERNING CYKLER" DATA SET

	Number of trips	Number of location points
Herning Cykler data	900 thousand	250 million
In Herning proper	150 thousand	3 million

local wifi network would influence the location accuracy. The GSM network was disabled in both variants, by removing the SIM card.

*B. Mobile GPS accuracy experiment*

To assess the accuracy of mobile smartphone localisation data from EcoSense deployments were used. Data on reported GPS accuracy and localisation was extracted from the "Herning cykler" data set. As the data set is collected from users who have downloaded an app from an appstore, the app has been used in many different places, and for different purposes (including boating). As we wish to focus on accuracy in urban environments, points located outside a bounding box containing the city of Herning were removed. The resulting data contains 250 million location points from 900 thousand different trips.

*C. Emission modelling speed sensitivity*

The OSPM [1] modelling program was used to investigate the potential influence of speed inaccuracy on the results for the modelled air quality. OSPM models the air quality at the street level by combining long range pollution transport models, with urban background pollution model and traffic based street level pollution contributions. In our experiment we choose a street with heavy traffic in a built up area of a city, where the surrounding buildings create a street canyon. In the street canyon, turbulence from cross winds will concentrate the pollution on the lee-side of the street, and thus exacerbate the effects of the traffic emissions. The traffic load was chosen from information on traffic counts performed by the municipality. The builtin meteorological data for the city and fleet data for Denmark was used. A number of simulations was performed where the average speed of the traffic was varied. Numbers for the air quality measures (the concentration of the different pollutants at the height of an virtual sensor) was retrieved from the resulting reports to show the impact of different average speeds.

IV. RESULTS

We present in this section a number of results from our investigation. As the experiments span from single phone stationary experiments, over multiple mobile phones to computer simulations, there are quite diverse results.

*A. Stationary GPS accuracy experiment results*

The experiment concerning indoor stationary GPS accuracy, was conducted by having an iPhone 4s, without a SIM card, running the app "SensorLog" for five weeks. In the first three weeks the wifi was enabled and connected to a residential wifi

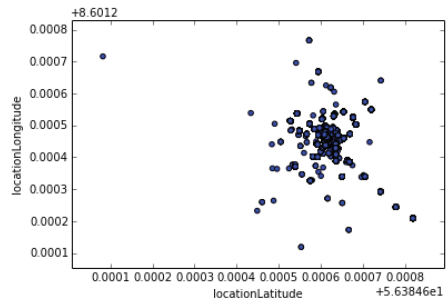


Fig. 2. Reported locations in both stationary experiments

network. In the last two weeks the experiment was repeated with the wifi radio disabled.

In Figure 2 a scatter plot for 1.2 million position measurements. To be able to see if there are structures apparent only points with a reported location accuracy below 200 meters has been shown. The figure shows that most points are grouped close to each other. The largest distance between points are 67 m. The central grouping shows some signs of preferred directions, and do not resemble a gaussian distribution of points from a fixed centre.

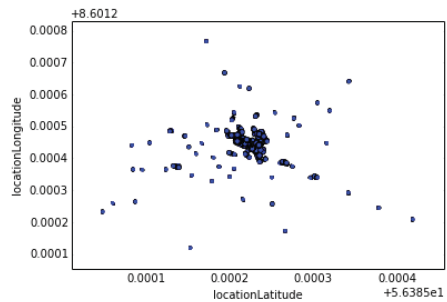


Fig. 3. reported locations in stationary experiment Wifi enabled.

In figure 3 locations from the first three weeks, where the smartphone was connected to the wifi network, and where the wifi radio could be used to aid the location system, by triangulation between neighbouring wifi networks. The single outlier from the previous figure is gone, but the preferred directions are more clearly discernible. When looking at a map of the residential area of the experiment, it is a quite suggestive to match the preferred directions to directions from the smartphone toward the neighbouring wifi access points. There app. 800.000 location points in this data set. The largest difference between the locations is 54 meters.

Figure 4 shows the location data from the second experiment

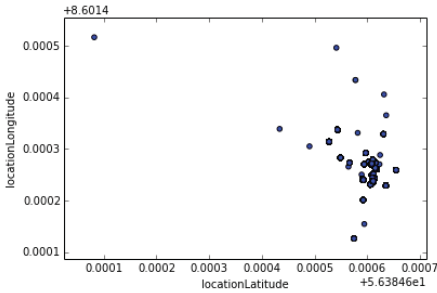


Fig. 4. reported locations in stationary experiment Wifi disabled

where the wifi has been disabled. The figure shows the single outlier from figure 2 and a very tight center with a few strays around the center. The preferred directions are less pronounced in this figure and the largest difference between points is 68 meters, and if the outlier is removed the largest distance between points is decreased to 33 meters. There are almost 430000 location points in this experiment.

From the examination of these two experiments we can conclude that using wifi networks as part of a localisation solution, does have an effect, but maybe it is better to solely rely on GPS for stationary use.

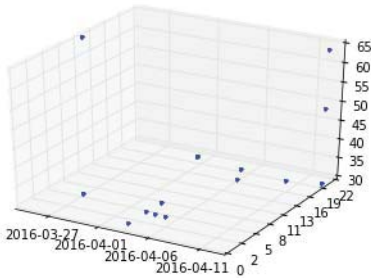


Fig. 5. Large reported errors > 10 meters as a function of time. GPS and Wifi.

To see if there are any recurring accuracy events, which could relate to the recurring constellations of GPS satellites, we show in figure 5 the reported accuracy and the date (x-axis) and time of day (y-axis). The measurements which shows low location uncertainty (less than ten meters) has been removed to not clutter the diagram. Figure 5 shows the 30 location measurements for the 3 weeks where both GPS and wifi was used for location estimation, where the reported accuracy was above 10 meters. There are no apparent recurring daily events so we cannot conclude that satellite constellations seriously affect the accuracy of stationary location measurements. There

seems to a small tendency towards larger uncertainty at five in the morning and at ten in the evening.

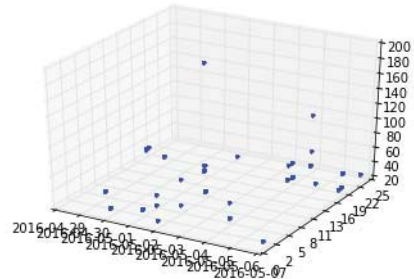


Fig. 6. Uncertainty histogram

In figure 6 we show the same diagram as in figure 5, but for the two weeks with only GPS as localisation device. In these two weeks there were 330 incidents of reported errors larger than ten meters, but there are no apparent recurring events of large localisation uncertainty, thus we cannot find support for the hypothesis that satellites constellations influence GPS accuracy. In this GPS-only experiment it seem that not relying on wifi for localisation introduces more incidents of large accuracy uncertainty.

From the above analysis a few observations can be made. First it can be seen from the figures 2, 3 and 4 that the accuracy of GPS location without wifi based location is better than GPS with wifi based location, except for a single point. From the figures 5 and 6 it seems that the variability of the reported errors is higher when the wifi is switched off, even though that the scatter plots suggested otherwise. It seems that the self reported error is overly relying on the access to known wifi access points.

**B. Mobile GPS accuracy experiment results**

In this section we report the results of our investigation into reported accuracy from smartphones as used in the wild. The data used is collected through the "Herring cykler" deployment. Even though the app was only promoted in Herring, Denmark, the deployment has generated data from other places in Denmark, and outside Denmark, the last part probably because of some of the informants going on holiday, and bringing their smartphones with them. The dataset contains data from 900 thousand trips, of varying lengths, with 250 million location measurements.

In figure 7 a histogram for the reported uncertainties. The histogram is abbreviated to uncertainties below 200 meters, and the bins are ten meters wide. It is clear to see that the most of the reported uncertainties are ten meters and below. But there is still almost half of the measurements with larger uncertainties.

The figure 8 shows some of the long tail of the uncertainty distribution. The small uncertainties (below 20 meters), has



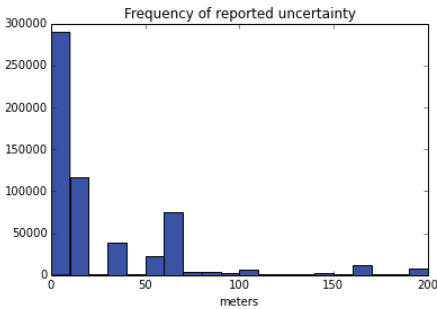


Fig. 7. Uncertainty histogram

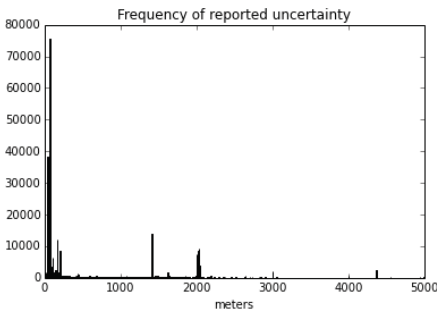


Fig. 8. Uncertainty displays a wide spread. The counts of low uncertainty (< 10) has been removed

ben removed to enable the view of the smaller counts of the larger uncertainties. In the figure 8 one can see that even though most of the location measurements have uncertainties below 100 meters, there is still a considerable amount of measurements with quite large uncertainties above 500 meters.

C. Emission modelling speed sensitivity results

The results of the experiment to quantise the sensitivity of the air quality model towards the speed of the traffic.

The figure 9 shows the concentration the gas  $NO_2$  at a simulated measurement point in a street canyon. The figure shows that the concentration of  $NO_2$  drops as the average speed increases. The drop in concentration is congruent with the speed dependency of emission factors [16], for low speeds the emission per meter is high due to idling or low utilisation of the motor. At higher speeds the motor output is used more effectively, thus the emission factor is lower. As the speed increases towards highway speeds the emission factor grows as the motor has to overcome increasing wind forces. We have only made the simulations for speeds less and equal to 120 km/t, as higher speeds are not deemed relevant for traffic in an urban street canyon.

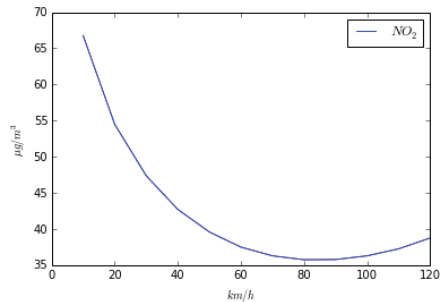


Fig. 9. Modelled  $NO_2$  concentration for different average speeds

From the figure 9 it can be seen that the speed sensitivity of the emission model is largest for low vehicle speeds. To accurately model the street level pollution it is important that the traffic data is accurate for low speeds.

V. APPLICATION OF GPS LOCATION DATA IN STREET LEVEL AIR QUALITY MODEL

To improve street air quality models like OSPM, with data obtained from smartphones of urban travellers, two essential pieces of information can be supplied. The location of travellers can be used to account for emissions from a trip to the locations traveled through. The speed at different locations is the second important information that the model needs in order to estimate the emission amount.

The smartphones could also supply information on the type and motor size of the vehicle, which could further improve the model estimate, but currently this information has to be gained by asking the user of the smartphone. Research has been done to automatically detect transportation mode [17], but currently it seems impossible to detect motor size and fuel type. One exception is to detect electric vehicles versus internal combustion engine vehicles [18].

VI. DISCUSSION

From the experiment with the OSPM model, it can be seen that the speed of vehicles are very important to the air quality at the street level. But it is also important to be able to attribute the emissions from vehicles surveilled by applications like ours to the right streets. That means that the accuracy of the location of vehicles monitored by smartphone apps must be high enough to assign the location to a single street.

To evaluate the influence of GPS accuracy an air quality modelling we have to determine the connection between location accuracy and speed accuracy and possible methods to overcome inaccuracy problems.

From the GPS design documents the best case error estimate is below six meters, which should be achievable in 95% of the time [11]. The speed is (falsely) reported to be larger than zero in 29 cases in the stationary experiment. As the reported

locations as seen in figure 2,3,4 should result in more reported nonzero speed events, if only the raw GPS locations were used to calculate the reported speed. The largest reported speed in the 29 cases of non zero speed is 2 km/h.

Map matching algorithms can reduce the inaccuracy from the GPS data, as seen in [10]. A median filter might also provide a stable location trace.

The errors as seen in the real life data set from "Herning Cykler", points toward the conclusion that we can use smartphone data to supplement our other data sources, for air quality modelling at the street level. This possibility enables more dynamic and up to date air quality monitoring, and possibly also new kinds of health alerts and urban transport planning overview.

Since the projects that have enabled the data collection has been targeted at special audiences, it is likely that there is a bias towards people who cares about sustainability, environment and man made climate change. The validity of the analysis of GPS accuracy does not suffer from this bias.

## VII. FUTURE WORK

In this paper we have only considered the collected data to be from person transport in cars. To further the quality of the collected data, methods to discern between different modes of transportation, to correctly adjust for the emissions from the vehicles, is needed.

## VIII. CONCLUSION

We have measured the accuracy of GPS data from smartphones, and investigated if the data is As traffic models become more ubiquitous, and are able to model the largest cities [19], urban transport planning get more tools for simulation of urban policy initiatives and new infrastructure initiatives, we hope that the work in this paper will lead to urban planning to take air quality into account.

## ACKNOWLEDGMENT

We want to thank Henrik Blunck, and Niels Olof Bouvin for their generous support.

This work has been supported by The Danish Council for Strategic Research as part of the EcoSense project (11-115331), the TUD COST Action TU1302 Satellite Positioning Performance Assessment for Road Transport (SaPPART), and EU Interreg project Biogas2020.

## REFERENCES

- [1] R. Berkowicz, "OSPMa parameterised street pollution model," in *Urban Air Quality: Measurement, Modelling and Management*. Springer, 2000, pp. 323–331.
- [2] K. E. Kakosimos, O. Hertel, M. Ketzel, and R. Berkowicz, "Operational Street Pollution Model (OSPM) - A review of performed application and validation studies, and future prospects," *Environmental Chemistry*, vol. 7, no. 6, pp. 485–503, 2010.
- [3] J. Christensen, "The Danish Eulerian hemispheric modelA three-dimensional air pollution model used for the Arctic," *Atmospheric Environment*, vol. 31, no. 24, pp. 4169–4191, 1997.
- [4] C. Directive, "Council Directive 1999/30/EC of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air," *Journal of the European Communities*, vol. 50, no. 163, pp. 41–60, 1999.
- [5] J.-m. Kang and B.-h. Hong, "A Study on the SNS ( Social Network Service ) Based on Location Model Combining Mobile Context-Awareness and Real-Time AR ( Augmented Reality ) via Smartphone," in *Future Information Technology*. Springer, 2011, pp. 299–307.
- [6] C. M. Walton, "Urban Travel Demand Analysis for Austin TX USA using Location-based Social Networking Data," *TRB Annual Meeting*, pp. 1–17, 2013.
- [7] Y. Zhang, X. Qin, S. Dong, and B. Ran, "Daily O-D Matrix Estimation Using Cellular Probe Data," *Transportation Research Board 89th Annual Meeting*, 2010.
- [8] J. Gone and R. J. F. Rossetti, "Smartphone Sensor Platform to Study Traffic Conditions and Assess Driving Performance," in *IEEE 17th International Conference on Intelligent Transportation Systems*, 2014, pp. 2596–2601.
- [9] S. Hemminki, P. Nurmi, and S. Tarkoma, "Accelerometer-based transportation mode detection on smartphones," *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13*, pp. 1–14, 2013.
- [10] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "{PEIR}, the Personal Environmental Impact Report, As a Platform for Participatory Sensing Systems Research," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, ser. {MobiSys} '09. New York, {NY}, {USA}: ACM, 2009, pp. 55–68.
- [11] DoD, "GLOBAL POSITIONING SYSTEM STANDARD POSITIONING SERVICE," 2008.
- [12] T. H. Witte and A. M. Wilson, "Accuracy of non-differential GPS for the determination of speed over ground," *Journal of Biomechanics*, vol. 37, pp. 1891–1898, 2004.
- [13] M. Modsching, R. Kramer, and K. ten Hagen, "Field trial on GPS Accuracy in a medium size city: the influence of built-up," *3rd Workshop on Positioning, Navigation and Communication 2006*, vol. 2006, pp. 1–10, 2006.
- [14] A. Zandbergen, "Positional Accuracy of Spatial Data : Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy," *Transaction in GIS*, vol. 12, no. 1, pp. 103–130, 2008.
- [15] P. Nitsche, P. Widhalm, S. Breuss, N. Brändle, and P. Maurer, "Supporting large-scale travel surveys with smartphones A practical approach," *Transportation Research Part C*, vol. 43, pp. 212–221, 2014.
- [16] O.-K. Nielsen, M. Plejdrup, M. Winther, M. Nielsen, S. Gyldenkerne, M. Mikkelsen, R. Albrektsen, M. Thomsen, K. Hjelgaard, L. Hoffmann, P. Fauser, H. Bruun, V. Johannsen, T. Nord-Larsen, L. Vesterdal, I. Møller, O. Caspersen, E. Rasmussen, S. Petersen, L. Baunbæk, and M. Hansen, "Denmark's National Inventory Report 2014," University of Aarhus, Tech. Rep. 101, 2014.
- [17] Z. Sun and X. J. Ban, "Vehicle classification using {GPS} data," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 102–117, dec 2013.
- [18] M. Wüstenberg, H. Blunck, K. Grønbaek, and M. B. Kjærgård, "Distinguishing Electric Vehicles from Fossil-Fueled Vehicles with Mobile Sensing," in *2014 IEEE 15th International Conference on Mobile Data Management*. IEEE, jul 2014, pp. 211–220.
- [19] A. Lehmann and A. Gross, "Using Crowd Sensed Data as Input to Congestion Model," in *IEEE International Conference on Pervasive Computing and Communication Workshops*. IEEE, 2016, p. 6.

# Design and Performance Analysis of RTS/NCTS Based MAC Protocol in IEEE 802.11 Systems

Md. Emdadul Haque<sup>‡</sup>, Faisal Tariq<sup>†</sup>, Ghazanfar A Safdar<sup>§</sup>, Yue Chen<sup>†</sup>

<sup>‡</sup>Department of Information and Communication Engineering, Rajshahi University, Bangladesh.

<sup>†</sup>School of Electronic Engineering and Computer Science, Queen Mary University of London, UK.

<sup>§</sup>Department of Computer Science and Technology, University of Bedfordshire, UK.

Email: haque-ice@ru.ac.bd, f.tariq@qmul.ac.uk, ghazanfar.safdar@beds.ac.uk, yue.chen@qmul.ac.uk

**Abstract**—IEEE 802.11 Medium Access Control (MAC) is one of the major standard for most of the latest and next generation Wireless Local Area Networks (WLAN). Despite its widespread use, the performance of the IEEE 802.11 MAC in terms of throughput and delay degrades noticeably with the increasing number of nodes in a certain area. Since the number of contention increases with the increasing number of nodes, probability of accessing channel reduces, resulting in increased queue length which eventually becomes full rapidly. When the queue becomes full, any message arrived afterwards gets dropped causing severe performance degradation. This paper proposes a novel technique for enhancing the IEEE 802.11 MAC layer which prohibits queue overflow and thus eliminates the packet drop due to the queue overflow. Extensive simulation is carried out to demonstrate the effectiveness of the enhanced MAC protocol and its impact on the higher layer protocol especially Transmission Control Protocol (TCP). The simulation results shows significant performance improvement in terms of throughput, delay and message drop rate.

**Keywords**—IEEE 802.11, MAC, packet drop rate, queue management, overflow.

## I. INTRODUCTION

IEEE 802.11 MAC [1] standard and its' variants for Wireless Local Area Networks (WLANs) is expected to be widely used in the next generation networks. The WLANs based on IEEE 802.11 MAC is becoming increasingly popular in recent years. As a result, number of supported devices have also increased significantly. The MAC protocol is also used in wireless sensor networks [2], wireless testbeds and simulation packages for wireless ad hoc networks.

For channel access in WLAN scenario, Distributed Coordination Function (DCF) is used in most cases. It works based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. Apart from this, Point Coordination Function (PCF) [3] and Hybrid Coordination Function (HCF) are also used as channel access mechanisms in certain scenarios. In CSMA/CA, retransmission of the collided packet is managed according to its binary exponential backoff algorithm. MAC performance degrades dramatically with the increase in number of active stations (STA) or devices [4]. A STA drops a packet either for queue overflow or after exceeding the maximum retransmission limit. In large networks, the queues of the STAs frequently overflow for resource limited STAs and results in packet drop. This is exacerbated by the fact that even if the queue of an STA with IEEE 802.11 MAC is full, it still accepts packet from a sending STA if other properties in [1] is

satisfied. After receiving the packet the station drops the packet due to lack of space in the queue buffer. This leads to severe inefficiency as the message transmission process occupies the channel for the complete packet transmission period which could have been used by other STAs nearby.

A congested STA continues to receive redundant MAC packets from sending STA until it is directed otherwise by the top layer protocol. Usually TCP instructs sending STA to reduce traffic rate towards the affected STA in order to reduce congestion. However, this also degrades the network throughput to a certain extent.

A redundant message transmission is used in the MAC layer to maintain the network dynamics. A STA keeps the dynamics using its binary exponential backoff algorithm [5]. In this algorithm when a station senses a channel as free and after the specified period it starts the packet transmission. If another station starts transmission at the same time then the packet collides. The station increases the backoff interval using the algorithm and retries after a relatively longer interval. On the other hand if the transmitter STA can send a packet to receiver STA successfully, the backoff interval for next transmission starts from the initial value. Thus the sending station can send packet to a destination station with lower backoff interval. Although the redundant message transmission indirectly helps a node to access the medium more frequently and increase network dynamics, the network becomes unstable. Moreover the unnecessary message transmission will cause faster battery power drainage which is detrimental for power limited devices.

Several enhancements on IEEE 802.11 MAC could be found in recent years to improve the performance of the MAC layer. The existing protocols highlight different issues to improve the network throughput. The protocols indirectly reduce message drop. For example, an effective backoff algorithm based on linear increment of contention window is proposed in [6]. In this algorithm each node goes back to the previous backoff stage after each successful packet transmission. The algorithm gives better performance in terms of throughput, packet delay and fairness. There are some other backoff algorithms could be found in [7], [8], [9], to improve the performances. These algorithms adjust the contention window by increasing or decreasing the window. The protocols indirectly increase throughput and decrease message drop without specifically addressing the message drop issue. A numbers of researchers propose the MAC layer enhancement based on cooperation among the wireless nodes [10], [11]. These protocols also do not pay attention to the direct message drop

reduction issue.

In the existing works explicit message drop reducing policy is largely ignored especially for buffer limited wireless nodes. Hence a scope to prevent the redundant message transmission keeping the network dynamics unchanged is still open. The simplest way to eliminate the redundant message transmission is to ensure that a node does not accept any new message when the buffer is full. Though message drop problem can be eliminated this way, it will affect the overall network dynamics adversely since the sender increases the backoff interval in this process.

To solve the problem this paper proposes a novel idea to maintain the network dynamics. So far the authors knowledge this is the first time the direct packet drop elimination policy is introduced in the MAC layer. In this technique, if the receiver STA is full, it sends a specialized reply message, defined as Not Clear To Send (NCTS), to the sender STA in response to the data send request message. after receiving the NCTSs, the sender initiate the backoff interval with the initial value as if the message has been sent. To demonstrate the overall effectiveness of the proposed MAC enhancement technique, the system is simulated with widely used TCP protocol. The results demonstrates that the proposed MAC enhancement mechanism achieves a significant performance gain in terms of throughput, packet delay and amount of message drop.

The rest of the paper is organized as follows. Section II presents a brief description of recent development of MAC layer enhancement protocols. The proposed mechanism is simulated and the performance is analysed in section III. Finally, the conclusion is given in section IV.

II. PROPOSED IEEE802.11 MAC ENHANCEMENT

In this section the IEEE 802.11 DCF is briefly discussed followed by detailed discussion on the proposed MAC enhancement algorithm.

A. IEEE 802.11 Distributed Coordination Function

This paper concerns with the Distributed Coordination Function (DCF) of IEEE 802.11, while other coordination function are out of the scope of the paper. DCF operates on two-way handshaking mode known as basic access mechanism and the four-way handshaking mode known as RTS/CTS access mechanism. The limit the size this paper concerns only on RTS/CTS mechanism. The four-way handshaking mode with Request To Send (RTS) and Clear To Send (CTS) is shown in Figure 1(a) for description refer to standard [1].

In the DCF protocol, a STA starts the transmission process when it senses the channel as free. It selects a random backoff interval uniformly from zero to current Contention Window (CW) size. The station then decreases the backoff timer by one at each time slot when the medium is sensed as idle. If the channel is sensed as busy, the station suspends its backoff timer until the completion of the current packet transmission (followed by a Distributed Inter Frame Space (DIFS) on successful completion of the packet or Extended Inter Frame Space (EIFS) on unsuccessful packet transmission). Transmission starts when the backoff timer reaches to zero.

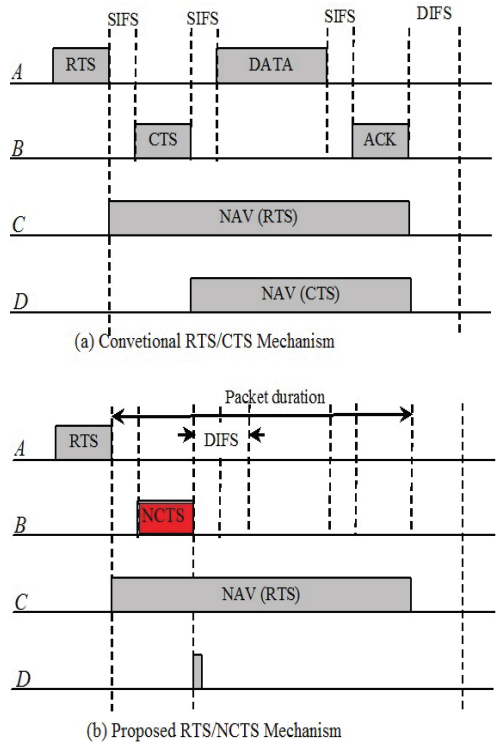


Fig. 1. Comparison of conventional access mechanism with proposed mechanism and their working procedure.

When the collision happens or the transmission fails the station calls its backoff procedure. In the backoff algorithm contention window starts with an initial value  $CW_{min}$  and doubles after each failure of successful transmission which has an upper limit  $CW_{max}$ . Once the  $CW$  reaches to upper bound maximum after each time the procedure is invoked the  $CW$  remain to the value until the maximum try limit is reached. If the transmission fails within the maximum try limit then retransmission stops by discarding the packet and  $CW$  initiate with the  $CW_{min}$ .

B. Proposed IEEE 802.11 MAC Enhancement

In the proposed enhancement, in addition to conventional RTS/CTS mode, each STA operates in another mode which is referred to as RTS/NCTS mode. A station receives packet according to RTS/CTS access mechanism until it overflows the queue as explained in Figure 1(a).

The proposed enhancement of the MAC protocol is triggered by a destination STA when its' queue becomes full. A destination STA B stops receiving any new packet from

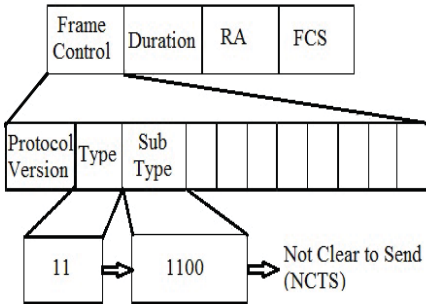


Fig. 2. NCTS Frame Structure.

Type value (b3 b2) bits	Subtype value (b7 b6 b5 b4) bits	Subtype description
01	1011	RTS
01	1100	CTS
01	1101	ACK
11 (reserve)	1100	NCTS

TABLE I. TYPE AND SUBTYPE FIELD OF IEEE 802.11

a sender STA A, since the queue has no empty space. B refuses a data sending request (i.e., RTS) with a specialized response message defined as *Not Clear To Send (NCTS)*. RTS and NCTS exchanging process of the proposed mechanism is shown in Figure 1(b). To define the NCTS frame this paper uses a reserve values of Type and Subtype field of IEEE 802.11 CTS frame as shown in Figure 2. The type and subtype field for RTS, CTS and ACK are described in Table I.

When A receives the NCTS in response to the RTS message, it discards the packet from sending message queue and initiates the *CW* by  $CW_{min}$ . STA A waits for DIFS time and again tries to access the medium. When STA B receives the RTS, it updates its Network Allocation Vector (NAV) by the period specified in the RTS. On the other hand another station, D received the NCTS and do nothing as no packet transmission will take place following the NCTS.

III. SIMULATION RESULTS AND ANALYSIS

In order to demonstrate the effectiveness of the proposed MAC protocol enhancement mechanism, two different network scenarios were simulated. The network is initially simulated for the enhanced MAC layer only and the performances is analysed in saturation condition where all the stations always have packets to send. Then the network is simulated with top layer protocol especially, TCP. Since TCP control the rate of transmission in the network, the node traffic condition is non-saturated.

To evaluate the performance of the proposed protocol a C program is written to implement IEEE 802.11 DCF. The MAC parameters are shown in Table II. In both saturation and non-saturation conditions, the networks are simulated for 60sec.

A. Performance Analysis in Saturation Condition

As mentioned before, in saturation condition every STA will always have sufficient amount packets to transmit for

Parameters	Values
MAC header	272 bits
PHY header	128 bits
RTS	288 bits
CTS	240 bits
ACK	240 bits
Packet Size	1000 bits
Channel bit rate	2Mbps
Propagation delay	2 $\mu$ S
SIFS	28 $\mu$ S
DIFS	128 $\mu$ S
Queue Length	20

TABLE II. DIFFERENT PARAMETERS AND THEIR CORRESPONDING VALUES

the entire duration of the simulation. To demonstrate the performance of the protocol, at first a network with only MAC layer and static network layer is simulated. To introduce the limited queue of each station a network layer protocol is required. For this purpose, shortest path routing is used for messages forwarding to the gateway. The network is simulated for different number of STAs. For each of these cases, the same random network is simulated for RTS/CTS and RTS/NCTS. Note that RTS/CTS means the performance related to conventional four-way handshaking mode of IEEE 802.11 DCF. On the other hand RTS/NCTS means the RTS/CTS along with the proposed RTS/NCTS.

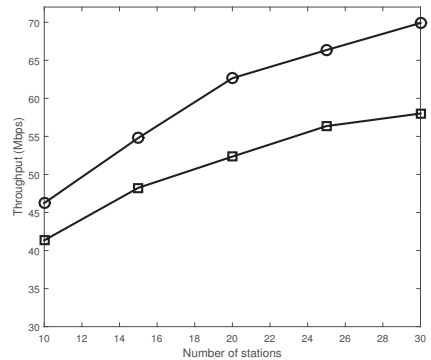


Fig. 3. Throughput performance for varying number of stations in saturation condition.

Figure 3 shows the throughput performance of the network for RTS/CTS and RTS/NCTS. It can be seen that the proposed enhanced MAC always outperforms the existing MAC in terms of throughput. It is also evident that the throughput gradually increases with the increase in number of STAs in both cases. When the traffic increases the number of messages drop saved with RTS/NCTS increases and hence throughput improvement increases as in the figure. However, the rate of improvement is much higher in the NCTS case compared to the traditional case. When the number of STA is 30, the proposed mechanism achieves 13MB of additional throughput which is approximately 20% higher than the conventional systems.

Since the aim of the work was to reduce unnecessary packet transmission and thereby reducing the amount of packet

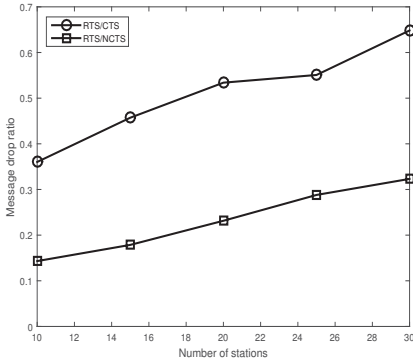


Fig. 4. Message drop saving performance in saturation condition.

drop after reception due to buffer overflow, the performance of the enhanced MAC can be best evaluated by measuring the amount of message drop saving which is demonstrated in Figure 4. The message drop ratio refers to the amount of message dropped by the sender before sending the message as stated in Section II(B) compared to the amount of message transmitted. Intuitively, with the increase in number of STA, the message drop increases. However, the RTS/NCTS technique drops approximately 50% less messages compared to the conventional system. This is mainly due to the fact that NCTS mostly stops transmitting redundant packets.

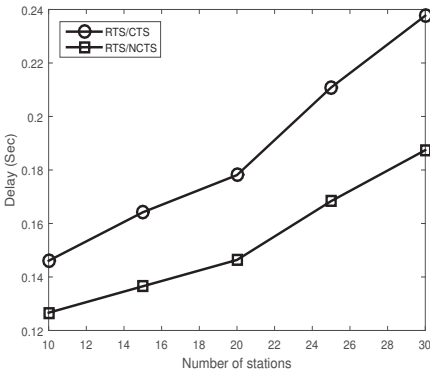


Fig. 5. Average End-to-End Delay Performance in Saturation Condition.

Another important measure of the performance of the protocol is average end-to-end delay. The end-to-end delay in this case is defined as the time difference between the message generation time and when the message successfully arrives at the gateway. The Figure 5 compares the end-to-end delay performance of the conventional system and the

proposed enhanced MAC system. Like other two performance indicators, the proposed RTS/NCTS outperforms the conventional RTS/CTS mechanism in end-to-end delay performance too. When the number of active STA is low (e.g. 10), the end-to-end delay for RTS/NCTS case is approximately 12% less compared to RTS/CTS case. The difference increases to 21% when the number of node is 30. This is due to the fact that at higher STA density, amount of message drop increases in conventional system which is offset by the proposed technique.

*B. Performance Analysis in Non Saturation Condition*

To evaluate the performance in the non-saturated condition, the most widely used transport layer protocol, TCP is implemented. In the non-saturation condition, not every STA will have enough packet to transmit (according to TCP protocol) and therefore, not every STA will be able to continuously transmit packets (unlike the saturation condition). So, the values of the performance indicators are lower compared to the saturation condition performance. The performance of the network in non-saturation condition is discussed below in detail.

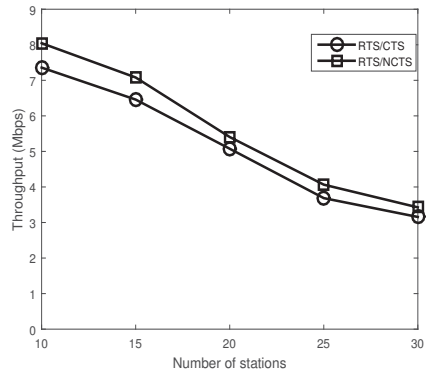


Fig. 6. Throughput Performance in non saturated condition.

Figure 6 shows the throughput performance of the network with varying number of stations for two different scenarios. The same network is simulated for both RTS/CTS and RTS/NCTS cases. The figure shows that the throughput almost constantly decreasing over number of stations. Since TCP throughput increases with increased amount of traffic [12], curve declines in the figure. As the figure about 1.5 MB more throughput is achieved when the number of stations is 30.

Figure 7 shows the message drop ratio performance comparison between proposed system and the conventional system. It can be clearly seen from the figure that the proposed RTS/NCTS drops less messages compared to the RTS/CTS. However, the difference between the two systems is lower in non saturation case compared to the saturation case as evident from Figure 7 and Figure 4. Since TCP controls the packet flow of the network, the number of messages in MAC layer is less than that of saturated network. So, the amount of message drop saved is also less than the saturated network.

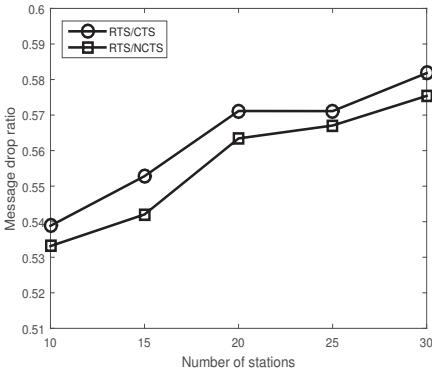


Fig. 7. Total message save in non saturated condition.

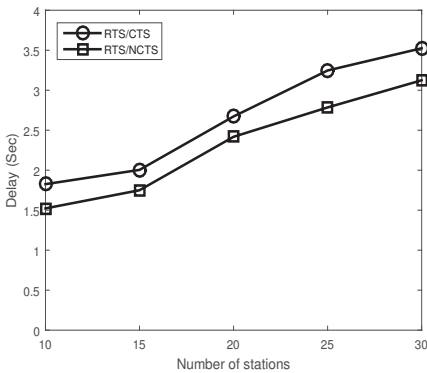


Fig. 8. Average End-to-End Delay Performance in non saturated Condition.

Finally, Figure 8 shows the average end-to-end delay performance of the simulated network scenario. The RTS/NCTS mechanism consistently demonstrates better delay performance compared to the conventional system with increasing number of the STA. Again, the difference of delay between two mechanism is lower in case of non-saturation compared to the saturation which is due to the similar reason mentioned above.

In summary, the simulation results demonstrate superior performance of the newly designed RTS/NCTS mechanism when compared with the conventional RTS/CTS mechanism in the MAC layer.

#### IV. CONCLUSION

In this paper, a novel MAC enhancement mechanism is demonstrated. The detail design and working mechanism is described with the aid of diagram and table which also demonstrates the difference of the proposed RTS/NCTS system with

the conventional RTS/CTS system. The proposed RTS/NCTS can be implemented with the existing frame structure and does not require any additional bits to be transmitted. Extensive simulation is carried out to demonstrate the effectiveness of the proposed mechanism where two different scenarios were considered. The first scenario, which is referred to as saturated condition focuses on the performance of the MAC layer only. The second scenario, referred to as non saturation condition was implemented with most widely used transport layer protocol, TCP and hence the cross layer performance gain has been captured. The performance of the proposed RTS/NCTS technique is evaluated against some key performance indicators such as throughput, end-to-end delay and amount of message drop saved. In all cases, the novel RTS/NCTS system consistently outperforms conventional system by a significant margin.

#### REFERENCES

- [1] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE 802.11 Std, 2012.
- [2] P. Huang, L. Xiao, S. Soltani, M. W. Mutka and N. Xi; "The evolution of MAC protocols in wireless sensor networks: A survey," IEEE Communications Surveys & Tutorials, 15(1), pp.101-120, 2012. .
- [3] G. A. Safdar and W. G. Scanlon; "Low power transmission efficient medium access control protocol for IEEE 802.11 infrastructure networks", Irish Signals and Systems Conference, Belfast, Ireland, pp. 357-362, July, 2004.
- [4] G. A. Safdar and W. G. Scanlon, "Pointer controlled power saving medium access control protocol for IEEE 802.11 infrastructure networks," 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 915-919, Vol.2, 2004.
- [5] X. Sun and L. Dai; "Backoff design for IEEE 802.11 DCF networks: Fundamental tradeoff and design criterion," IEEE/ACM Transactions on Networking (TON), Vol. 23, No.1, pp.300-316, 2015.
- [6] P. Patel and D. K. Lobiyal. A Simple but Effective Contention Aware and Adaptive Back-off Mechanism for Improving the Performance of IEEE 802.11 DCF. Wireless Personal Communications, vol. 83, no.3, pp.1801-1841, 2015.
- [7] M. Al-Hubaishi, T. Alahdal, R. Alsaqour, A. Berqia, M. Abdelhaq and O. Alsaqouret. Enhanced binary exponential backoff algorithm for fair channel access in the IEEE 802.11 medium access control protocol. Int. J. Commun. Syst.Vol. 27, No. 12, pp. 4166-4184, 2014.
- [8] K. Maraslis, P. Chatzimisios and A. Boucouvalas, IEEE 802.11aa: Improvements on video transmission over wireless LANs, IEEE International Conference on Communications (ICC), Ottawa, ON, pp. 115-119, 2012.
- [9] C. W. Park, D. Hwang and T. J. Lee, Enhancement of IEEE 802.11ah MAC for M2M Communications, in IEEE Communications Letters, vol. 18, no. 7, pp. 1151-1154, July 2014.
- [10] S. Kumar, H. Lim and H. Kim, Hierarchical MAC protocol with multi-channel allocation for enhancing IEEE 802.11ah relay networks, International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, pp. 1458-1463, 2015.
- [11] C. Cetinkaya, Multi-channel cooperative MAC protocol for wireless LANs, Ad Hoc Networks, vol. 28, pp.17-37, May 2015.
- [12] B Raffaele, M Conti, and E Gregori. Throughput analysis and measurements in IEEE 802.11 WLANs with TCP and UDP traffic flows, IEEE Transactions on Mobile Computing, Vol.7 No.2, pp. 171-186, 2008.

# Detection of offensive lateral movements using finite-state-machine-based patterns

Alessandro Greco<sup>1</sup>, Giuseppe Bianchi<sup>1,2</sup>

<sup>1</sup> University of Rome “Tor Vergata”, Italy – <sup>2</sup> CNIT, Italy  
name.surname@uniroma2.it

**Abstract** – The ever increasing sophistication of Advanced Persistent Threats (APTs) brings about the need for new detection approaches. The ability of modern offensive operations to acquire a foothold and then locally expand an infection inside the victim’s local area network, usually referred to as lateral movements’ activity, is particularly critical. Not only a distributed monitoring infrastructure is necessary to overcome the lack of a single network point for detection (opposed to the traditional network perimeter defense relying on outbound network intrusion detection systems), but also new signatures appear necessary to model the stealthy and complex behavior of offensive lateral movement activities. In this paper we focus on this latter issue, and we propose a formal model for stateful signatures, based on finite state machines. With reference to real world lateral movement attacks (e.g. PSEXEC), we show how the relevant detection signature can be gathered and formally modeled. Numerical results on real world traces show the effectiveness of the proposed approach in avoiding false positives.

**Keywords** – APT detection, lateral movements, finite-state-machine-based signature, PSEXEC.

## I. INTRODUCTION

In the cyber threat landscape, espionage campaigns, sometimes indicated as APT, are becoming one of the most insidious threat for industrial and governmental infrastructures. Despite the crucial attention and intense research effort spent by the cyber-security community, the average time it takes to detect APTs is still measured in days or even months, with an average reported in [1] of as much as six months. The main reason for this detection gap is the sophistication of infection and evasion techniques used by attackers, in most cases operating without specific automatic tools (easily detectable, as traditional virus).

An aspect that makes hard to defend against APTs is the stealth way in which such attacks spread. Most often, APTs first limit to acquire an internal foothold in a target networks, by, typically, finding a single weak spot and exploiting it to gain access to the network [2]. Once inside, APTs start to spread all around the network, collecting and exfiltrating sensitive data. In front of such fundamentally different behavior with respect to more traditional threats, network operators and target companies are still protecting their networks using defense techniques primarily targeting the protection of the network perimeter, e.g. via outbound Network Intrusion Detection Systems (NIDS). Indeed, a report from the security company Fireeye [2] suggests that companies still spend more than \$5 billion on traditional security

measures, whereas, as reported in [3], they often completely neglect internal network protection.

Unfortunately, the detection of such internal insidious activities is arduous for several reasons such as the granularity of controls to be performed and, above all, the complexity of indicators of compromise to be implemented, because of the use of administration-like operations and the consequent high false positive rate.

We posit that an effective internal network defense, capable of detecting the spread of infections via lateral movements, requires significant advances in *two complementary directions*. The first direction is related to the monitoring infrastructure. Rather than relying on centralized outbound NIDS, new generation monitoring systems should rely on the analysis of *network-level patterns* via a *distributed monitoring architecture*, employing multiple (and ideally collaborative) probes in the form of software agents widely spread in the network, in principle down to one single probe per device. Note that we foster an approach based on the analysis of network-level events, as this holds the promise to more promptly detect anomalies and end-point misbehavior with respect to the current log-based host analyses (e.g. [4,5]). The second direction, and the main contribution of this paper, consists in identifying *new attack signatures which are amenable to network-level analysis and support scalable operation*. In fact, the obvious shortcoming on a network-level analysis, and to a greater extent of a widespread distributed infrastructure with multiple vantage points, is the sheer amount of traffic to inspect and the inherent complexity in identifying malicious patterns. Rule-based approaches (Snort-like [6]), appear too simple to properly model and characterize complex behaviors such as those involved in APT’s internal infection spreads. Indeed, current literature approaches which rely on Snort-like IDSs [7] usually take a two-step approach and thus lose the property of being able to detect and respond to threats in (near) real time: first log, in real time, network events gathered from Snort probes, and only later on parse offline such network logs to identify complex patterns.

The crucial question, addressed in this paper, is thus: *how to formally model realistic complex patterns, such as those involved in offensive lateral movements, in a way that they are amenable to be detected via a packet-by-packet analysis?* Our answer resides in the notion of *finite-state-machine-based signatures* (section III) i.e. signatures which are still based on the analysis of low-level packet level events (which can be thus expressed via rules), but which also permit to explicitly



model the different behavior in time of an attack through the notion of “attack state” and the formalization of the relevant state evolution. It was proven in [8] that a stateful signature can be scalably supported by a probe engine called “StreaMon” which operates in real time over a stream of packets and without the need to actually log data; what was however missing in [8] was a set of compelling and realistic use cases showing that real world attacks could be modeled using a finite-state-machine-based signature.

This paper fills this gap, by taking as use case the well known lateral movement technique, called PSEXEC, and by step-by-step showing (section IV) how it can be formally modeled in a stateful composition of rules which can then be executed over the StreaMon probe [8]. While the PSEXEC analysis is carried out into fine grained details, we posit that the same approach can be also applied to many other use case scenarios. A discussion of related works is then provided in section VI and conclusive remarks and further possible evolutions of this work are summarized in Section VII.

II. BACKGROUND AND SCENARIO

A new widely-used approach to model cyber threats is called *Intrusion Kill Chain*. Firstly introduced by Lockheed Martin [9], it provides a way to characterize each phase of an intrusion event, in order to reconstruct the entire framework and then drive defensive actions. According to this model, an intrusion event is modeled in seven steps, from the reconnaissance phase (1), needed to gather information about the target, to the choice (weaponization – 2) and delivery (3) of the adequate malware that, after the exploitation (4) and installation (5) phases, is remotely controlled by a command and control (6) link. Once achieved the foothold in the network, the attacker moves toward the objectives (actions on objectives – 7). In order to describe an intrusion event, it is necessary to characterize each of the seven steps and point the related *Indicators Of Compromise* (IOCs) necessary to detect and identify the offensive activities around the network by means of a dedicated security device. This model considers three types of IOCs: the *atomic* ones, indicators which cannot be broken down into smaller parts and are strictly related to a particular intrusion events (e.g. IP addresses, email addresses), the *computed* ones, derived from data involved in the intrusion event (e.g. hash), and the *behavior* ones, logical combination of atomic IOCs.

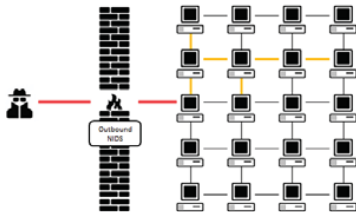


Fig. 1. Example of offensive activities in a LAN

Most of the traditional work carried out in intrusion detection systems focuses on the design of what we refer to as

“outbound NIDS” (Fig. 1), namely systems devised to detect commands between infected hosts inside the network and external Command & Control entities. However, especially in phase 7 (*Actions of objectives*), modern APTs develop stealth techniques to permit an attacker, who has initially taken control of an internal victim and has hence gathered a foothold *inside* the network, to spread the infection and take control of other hosts *without any external interaction*. Rather, the attacker uses the already infected machines inside the LAN as pivot points, send commands (and receives answers), and finally spreads the infection through direct channels established inside the LAN, hence hidden from the Outbound NIDS (orange lines in Fig. 1).

The defense against such lateral movements is complex for many reasons. First, and obvious, there is not anymore a single point in the network such as an outbound NIDS which permits to monitor malicious traffic activities. Second, the attacker employs multiple different techniques in such lateral movement phase. Finally, such techniques are engineered to hide among legitimate traffic and activities, and use standard administration-like operations in this phase. Through the usage of such methods it is difficult to detect the lateral movement, because benign and malicious utilization of these methods looks similar and therefore distinction is complex [10].

A suitable defense against lateral movement thus must face two complementary needs. First, the network monitoring architecture must also spread inside the LAN, via monitoring agents widespread located inside the network (in principle, up to a single vantage point per end device). In this paper we will assume that such a widespread monitoring is in place and we focus on how to effectively model the threats of interest, so as to enable their near real time detection. Since lateral movements are characterized by an administration-like traffic (derived most often by human-driven operations), whose single packets are legitimate events, to prevent false alarms it is necessary to detect how such events occur (and change) in time. As discussed in the next section, we suggest to rely on state machines (executable as signatures to be implemented in StreaMon probes[8]).

III. STATE MACHINE-BASED SIGNATURES

Despite the growing interest in behavioral-based NIDS, designed to automatically identify “deviations” from the normal behavior of a host or network [11], the need for their extensive training in conjunction with their statistical (non-deterministic) operation makes such that signature-based NIDS still play a dominant role in the NIDS market. The key problem in signature-based NIDS is the appropriate definition of signatures for the variety of threats recognized by the system. Threats emerging in lateral movements and APT scenarios are usually very complex, and are hardly summarized into a single matching rule. Rather, they do encompass a multiplicity of serial steps which, if taken alone and detected with atomic IOC by a stateless NIDS, could give many false alarms.

To address such problem, we propose to incorporate the “state” of a threat into a formal model of the threat signature, so that detection of different atomic IOCs can be employed at different “stages” of the threat’s evolution (summarized by

explicit states). It readily follows that an IOC which could be normal or legitimate at a given stage, becomes an indicator of an ongoing attack only when it happens at a different stage, thus significantly reducing false positives. To our surprise, only limited research work has addressed the goal of devising stateful signatures. For instance, [12] promotes such an approach but restricts to layer 3 rules only. Conversely, the StreaMon probe proposed in [8] provides a monitoring probe architecture capable of executing and evolving state-machine-based signatures, but provides only very limited insight on how to model real world threats and whether a state-machine-based signature model is actually suitable for this task.

In this paper, we propose to model behavioral pattern signatures as state diagrams which permit the programmer to **both** formalize the notion of “*attack state*”, as well as define, for each state, the **rules which model the state evolution in the detection process**. In details, a finite-state-machine-based signature is expressed in terms of:

- **entities**: the entity to which as time-varying “state” is associated; it can be a single entity such as a target IP address, or a pair such as [IP attacker, IP victim];
- **states**: for each entities, it formalizes an “attack state” and it associates to each state a possibly different set of features (events or conditions) that shall be monitored. Note that different states may entail the monitoring of different events;
- **state transitions**: similarly, for each state, one or more state transition rules are associated, again expressed as a combination of events and conditions. This permits to formally specify how the state shall evolve.
- **actions**: once a transition state is executed, the finite-state machine can also perform actions, such as changing a variable (e.g. counter) or giving a warning to a security alerts collector.

Once the programmer has identified the set of attack states, the main modeling task consists in associating state transitions to matching rules, expressed as a Boolean predicate on events or conditions (as defined below). In this paper we will adopt the notation and syntax shown in the table below.

<p><b>Transition from <u>State#N</u> to <u>State#M</u>:</b>                  (Event A OR Event B) AND Condition C                  where:                  - <u>Event A</u>: description and indicator of the event A                  - <u>Event B</u>: description and indicator of the event B                  - <u>Cond.C</u>: description and indicator of the condition C                  followed by:                  • <u>Action</u>: description of the action</p>
--

An *event* refers to information that can be detected from the packets, including both information gathered from the packet header as well as information eventually extracted from the packet payload using Deep Packet Inspection (DPI). A *condition* refers to the test of logical operations on collected statistics associated to the entity and/or state (e.g. event counter greater than a threshold, communication direction, etc.). Finally, the signature developer may optionally associate an *action* to each state transition (for instance, send an alert or reset a counter). Note that, with reference to the *Kill Chain* method recalled in Section II.A, events are atomic IOCs,

conditions are computed ones, whereas an entire transition can be interpreted as behavioral IOC.

IV. MODELING THE PSEXEC ATTACK

In order to gather insights on how easy/effective is the proposed model, we applied it to a set of known threats. In this Section, we step-by-step show how the proposed technique can be applied to the detection of the well-known PSEXEC attack [13-15], frequently used in real lateral movement operations.

A. Attack description

PSEXEC is a method firstly conceived for administration purposes (remote execution of process) and then misused for hacking, being one of the most used lateral movement techniques in the APT campaigns [10]. In a LAN, through the SMB (Server Message Block) protocol functionalities, it allows to automatize the process of copying a software on a shared directory of a second host and then remotely execute it as a service. Through the SMB protocol, applications can communicate and share data, by means of pipes and remote calls. In particular, the RPC (Remote Procedure Call) mechanism allows an application to seamlessly invoke remote procedures, as if these procedures were executed locally. The Microsoft implementation, called MSRPC, uses SMB to remotely address the Windows pipes, likes “svcsctl” associated to the Windows process “services.exe”, responsible of the services management inside the operating system. Using these functionalities, once achieved a foothold in a network and grabbed credentials, an attacker can move around the domain using SMB, to be authenticated and move the malware file, and RPC (e.g. SVCCTL), to remotely execute the it as service.

B. Attack pattern modeling

The analysis of this attack was performed using Windows XP and Windows 7 machines, in order to consider different versions of SMB protocol (SMB and SMB2). Concerning the implementation of the PSEXEC attack, we used two well-known different tools, here below listed:

- *PsExec*, belonging to the *Microsoft SysInternal* suite [16], designed and used for administration purpose, but misused for hacking as well.
- the hacking tool *Impacket* [17], collection of python classes for working with network protocols, providing low-level programmatic access to the packets and for some protocols (SMB1-3 and MS-RPC).

In order to identify and then model an effective signature, we analyzed noise-less traffic samples containing the pattern of interest. In our case, we created a virtual environment, representing a basic two-hosts (one WinXP and one Win7) victim local Domain based on Windows Active Directory, and we used an attacker Linux Kali Virtual Machine. After having performed many times the offensive activities on the victims, the traffic extracted from PCAPs was firstly visually analyzed through the *Message Sequence Chart* (MSC) representation, initially developed by International Telecommunication Union for requirement specification of protocols. This representation is useful for extracting patterns because it gives an overview of the message exchange among communicating entities,

considering order of messages and time constraint [24]. For example, Fig. 2 shows the network pattern of a SMB authentication failure, extracted from a PCAP containing a SMB brute force attack (afterwards better explained).



Fig.2. Message Sequence Chart of SMB Authentication Failure

In order to interpret such chart, it is necessary to understand how the SMB protocol works. The SMB packets header contains the command to be executed (expressed as hex code), which is appended with a command-related payload; for example, with reference to the “Session Setup AndX Request” command shown in Fig. 2, at a fix position in the header it is possible to extract the *SMB\_Command* (“0x73” for SMB and “0x0001” for SMB2) or the *ResponseFlag* (“0” for request packet and “1” for the response one), while in the request packet payload it is possible to extract the credentials.

In this example shown in Fig. 2, the graphical analysis of the network pattern allows to easily pinpoint the starting and closing TCP connection sequence (black squares), the SMB dialect negotiation (green square) and the SMB authorization failures (red square) related to the SMB “Session Setup AndX” Request (containing the credentials) and Response (with Status Logon Failure error) commands.

Once modeled, the pattern is implemented as a python script, representing the finite-state machine. Concerning the DPI functionalities, the script firstly used a public library (*pyshark*, the python version of *tshark* [19]) and, in order to be ported in *StreaMon* [8], it uses own implemented libraries, based on Microsoft Windows specification [20] and [21].

C. Finite-state machine pattern signature

The output of the analytical step described in the previous section is a very simple finite-state machine. In principle our approach of course permits more intertwined state transitions, whenever needed. Here below the model of the PSEXEC attack is shown, from the SMB authentication to the remote service call through *SVCCCTL*:

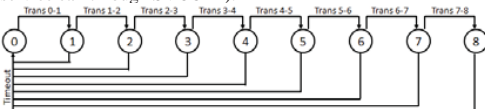


Fig.3. Finite-state machine pattern related to PSEXEC threat

The entities of such finite-state machine are couples [Attacker IP, Victim IP] defined once the first transition happens. Here follows a (commented) description of the state transitions:

**Transition from State#0 (idle) to State#1:**  
 (Ev. 1A OR Ev. 1B) AND Cond. 1C where:  
 -Event 1A: SMB Session Setup AndX (SMB Command="0x73")  
 -Event 1B: SMB2 Session Setup (SMB Command="0x01")  
 -Cond. 1C: Request packet (SMB ResponseFlag="0")  
 • Action: Save the entities (IP Attacker, IP Victims) in the Look-Up Table

The first transition state is related to the SMB authentication (*Session Setup AndX*) request (*SMB\_ResponseFlag*="0" for SMB and SMB2).

**Transition from State#1 to State#2:**  
 (Ev. 2A OR Ev. 2B) AND Cond. 2C AND Cond. 2D AND Cond. 2E where:  
 -Event 2A: SMB Com Tree Connect AndX (SMB Command="0x75")  
 -Event 2B: SMB2 Com Tree Connect (SMB Command="0x03")  
 -Cond. 2C: Request packet (SMB ResponseFlag="0")  
 -Cond. 2D: Pointed resource related to victim (Victim IP in SMB Path)  
 -Cond. 2E: Pointed resource not related to Interprocess Communication ("IPC" string not in SMB Path)

With such SMB command, it is possible to point the resource (pipes) addressed through the “SMB Path” variable. In fact, once authenticated, in order to create a remote call the SMB client (attacker) invokes the special share *Interprocess Communication* (IPC), which provides the information about the RPC named pipes. After that, the SMB client requests to create a SMB session on a shared directory in the victim host. This state transition is then related to the request of SMB session on the \\IP,SharedDir, called *Tree Request* because it allows the client to connect to the shared directory tree.

**Transition from State#2 to State#3:**  
 (Ev. 3A OR Ev. 3B OR Ev.3C) AND Cond.3D AND Cond.3E  
 -Event 3A: SMB Com NT Create AndX (SMB Command="0x2a")  
 -Event 3B: SMB Com Open AndX (SMB Command="0x2d")  
 -Event 3C: SMB2 Create (SMB Command="0x05")  
 -Cond. 3D: Request packet (SMB ResponseFlag="0")  
 -Cond. 3E: Executable requested file (".exe" string in SMB File)  
 • Action: Save SMB File in the Look-Up Table (LUT)

Once created a SMB session on a shared directory, the Attacker ask to copy the malware backdoor through such commands.

**Transition from State#3 to State#4:**  
 Cond.4A  
 -Cond. 4A: Asymmetric traffic from attacker IP to victim IP (transferred bytes greater than a threshold)  
 • Action: Increase the byte transferred counter

In this phase, the Attached transfer the malware file, trough RAW TCP or SMB connections, which can be modeled as an asymmetric communication on a same port of the victim host.

**Transition from State#4 to State#5:**  
 (Ev. 5A OR Ev. 5B OR Ev.5C) AND Cond.5D AND Cond.5E  
 -Event 5A: SMB Com NT Create AndX (SMB Command="0x2a")  
 -Event 5B: SMB Com Open AndX (SMB Command="0x2d")  
 -Event 5C: SMB2 Create (SMB Command="0x05")  
 -Cond. 5D: Request packet (SMB ResponseFlag="0")  
 -Cond. 5E: Pointed resource related to SVCCCTL session ("SVCCCTL" string in SMB Path)

Once transferred the malware file, the SMB client tries to open a *SVCCCTL* session, through the *svccctl* pipe (as in *Transition State 2-3*).

**Transition from State#5 to State#6:**  
 (Ev. 6A OR Ev.6B OR Ev.6C OR Ev.6D) AND Cond.6E AND Event 6F  
 -Event 6A: SMB Com Write AndX (SMB Command="0x2f")  
 -Event 6B: SMB Com Write Row (SMB Command="0x25")  
 -Event 6C: SMB2 Write (SMB Command="0x09")  
 -Event 6D: SMB2 IOCTL (SMB Command="0x0b")  
 -Cond. 6E: Request packet (SMB ResponseFlag="0")  
 -Event 6F: SVCCCTL OpenSCManagerW (SVCCCTL Command="0x0f")

Once created a connection to the *SVCCCTL* pipe, the Attacker asks to establish a connection to the SCM on the specified computer and opens the specified SCM database.

**Transition from State#6 to State#7:**  
 (Ev.7A OR Ev.7B OR Ev.7C OR Ev.7D) AND Cond.7E AND (Event 7F OR Event 7G) AND Cond. 7H  
 - Event 7A: SMB Com Write AndX (SMB Command="0x2f")  
 - Event 7B: SMB Com Write Raw (SMB Command="0x25")  
 - Event 7C: SMB2 Write (SMB Command="0x09")  
 - Event 7D: SMB2 IOCTL (SMB Command="0x0b")  
 - Cond. 7E: Request packet (SMB ResponseFlag="0")  
 - Event 7F: SVCCTL CreateServiceW (SVCCTL Command="0x0e")  
 - Event 7G: SVCCTL OpenServiceW (SVCCTL Command="0x10")  
 - Cond. 7H: Verify if in the name of service created through SVCCTL is contained the name of the executable name defined at State#2 (MW\_FileName -from State#2 - in SVCCTL File.

Once accessed to the SCM, the Attacker asks to create a service associated to the previously transferred backdoor , using the SVCCTL CreateService or OpenService commands. These commands create a Windows service.

**Transition from State#7 to State#8:**  
 (Ev.7A OR Ev.7B OR Ev.7C OR Ev.7D) AND Cond.7E AND Event 7F  
 - Event 7A: SMB Com Write AndX (SMB Command="0x2f")  
 - Event 7B: SMB Com Write Raw (SMB Command="0x25")  
 - Event 7C: SMB2 Write (SMB Command="0x09")  
 - Event 7D: SMB2 IOCTL (SMB Command="0x0b")  
 - Cond. 7E: Request packet (SMB ResponseFlag="0")  
 - Event 7F: SVCCTL StartServiceW (SVCCTL Command="0x13")

Once created the malware service, the Attacker asks to execute it, which returns the shell needed to remotely control the victim.

Once reached the last state, it could be possible to further verify the presence of such threat by remotely checking the status of the new service (whose name is extracted in State#7), with Windows administration utilities. By doing so, the NIDS agent implements the Host Intrusion Detection System (HIDS) functionalities as well.

**D. Signature validation**

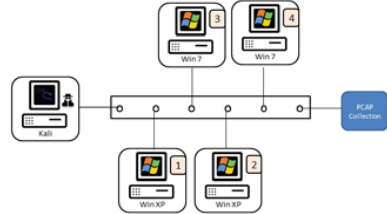
In order to validate the detection technique proposed for PSEXEC attack, we implemented a virtual lab with a Kali attacker box, two WinXP victim hosts and two Win7 victim hosts, as shown in Fig. 4. Although for this application it is necessary to have a widespread NIDS distribution, for this validation step we collect the network traffic through a virtual hub, which also allowed to better assess the capability to manage multiple connections between multiple hosts. We registered a 1-hour traffic track, emulating routine operations typical of an office host, as web browsing, shared directories accesses, operating system and antivirus updates, and we performed different lateral movements (opportunistically scheduled) between all the hosts, as here below shown:

HOSTS		PSEXEC TOOL	COUNT
Attacker	Victim		
K	1,2,3,4	Impacket	3
1	2,3,4	PsExec SysInt.	2
2	1,3,4	PsExec SysInt.	2
3	1,2,4	PsExec SysInt.	2
4	1,2,3	PsExec SysInt.	2

**Table 1. Performed lateral movement for validation**

In order to assess how to face the truncated chain events and false alarms with an effective timeout management, in some cases we voluntarily used incorrect credentials. From Kali box to host (2), we also performed a SMB brute force (trying to access to shared resources). This was continuously break the finite-state machine at State#1 (multiple request of new SMB sessions). For this purpose, we used the

implementation of the well-known hacking platform Metasploit, by means of the module "auxiliary/scanner/smb/smb\_login" [22].



**Fig.4. Validation architecture**

**E. Test results**

The offensive activities, that were expected to be detected, were correctly identified by the proposed solution, without false alarms. In order to understand the advantages related to the use of such approach with respect to the stateless NIDS one, we logged all the warning received using the state transitions individually considered as stateless signatures (apart from transitions 4 and 6, which requires information from the LUT).

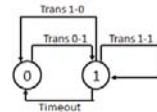
	T 0-1	T 1-2	T 2-3	T 4-5	T 6-7	T 7-8
Pkts	4615	70	54	222	47	68
Pkts/Tot [%]	6.59	0.10	0.08	0.32	0.07	0.10

**Table 2. Avoided false alarms**

Analyzing the results shown in Table 2 (related to a track of ~70k packets), it is possible to verify that:

- if we had analyzed this traffic with a stateless NIDS with those signatures, we would have received many false alarms, mainly associated to administration licit connections;
- the peak present for T0-1 column mainly relates to the SMB brute force attack intentionally performed from Kali (K) machine to WinXP (2) host.

The SMB brute force attack (used also in real cases, as in [23]) could be modeled as finite-state-machine-based signature.



**Fig.5. Finite-state machine pattern related to SMB brute force threat**

Fig. 5 shows finite-state machines, whose entities are couples [Attacker IP, Victim IP] defined once the first transition happens. Here below the descriptions of the state transitions:

**Transition from State#0 (idle) to State#1:**  
 (Ev. 1A OR Ev. 1B) AND Cond. 1C  
 - Event 1A: SMB Session Setup AndX (SMB Command="0x73")  
 - Event 1B: SMB2 Session Setup (SMB Command="0x01")  
 - Cond. 1C: Query packet (SMB ResponseFlag="0")  
 • Action: Save the entities (IP Attacker, IP Victims) in the LUT

This transition is the same as the first one of PSEXEC attack, because related to the SMB session creation request.

**Transition from State#1 to State#1:**  
 (Ev. 1A OR Ev. 1B) AND Cond. 1C  
 - Event 1A: SMB Session Setup AndX (SMB Command="0x73")

- *Event 1B*: SMB2 Session Setup (SMB Command="0x01")  
 - *Cond. 1C*: Answer packet (SMB ResponseFlag="1")  
 - *Cond. 1C*: Failure Logon (SMB FailureLogonFlag="1")  
 • *Action*: Increase the Failure Logon counter

Differently from the previous transition state, the *SMB\_ResponseFlag="1"* refers to the opposite direction of the packet and the *SMB\_FailureLogonFlag="1"* refers to the use of incorrect credentials. Once this condition is satisfied, the *FailureCounter* is increased.

*Transition from State#1 to State#0:*

*Cond. 1D*

- *Cond. 1D*: Failure Logon counter greater than the threshold

The SMB brute force event happens when the *FailureCounter* is greater than the number of accepted logon failures.

## V. RELATED WORKS

IDS can be host-based (Host IDS - HIDS) or network-based (NIDS): the first type is a system installed on single hosts, mainly dedicated to the monitor of files integrity, application logs and system calls, while the NIDS attempts to discover unauthorized access to a computer network by capturing the network traffic [24]. Almost all the detection techniques used for PSEXEC attacks are based on the Windows event logs analysis, through HIDSs and centralized SIEM applications. In [4] and [5], for example, in order to detect pass-the-hash attack (first step of the PSEXEC technique) it is suggested to implement an host based detection (using the Windows and antivirus logs in order to monitor the user authentication events - comparing the results against an user and/or IP approved list); in addition, a network based detection approach is proposed, based on the user habits (once created a baseline of normal system behavior, the detection techniques looks for anomalies on the number of connection in a short period of time on a specific port). A similar wider approach is proposed in the Microsoft guide dedicated to the mitigation of pass-the-hash attack [25], where they propose a complete list of anomalous user behaviors, which can be interpreted as stolen credential use, and the Windows log events related to such events. This kind of host based analysis can be performed in decentralized or centralized way, using SIEM solution whose high performance allows to implement complex statistical algorithms, as in [26]. For the detection of such threat, HIDS approach are hence preferred to the NIDS ones, probably because of the use of licit protocols and administration commands, although a network analysis can be performed at run time and integrated with Software Defined Network devices for mitigation. One example of NIDS detection for this threat is proposed in [27], whose idea is anyway based on the conversion of the Windows log indicator. In particular, indicators of lateral movement proposed in (Windows event type and relative information) are converted in SNORT format [6] using regular expressions and text search. A mixed approach, similar to one proposed in this article, is provided by [3]: the detection algorithm is based on the SMB authentication process chain in Windows architecture and performs DPI in order to extract credentials.

## VI. CONCLUSIONS AND FUTURE WORK

This work proposes a possible detection approach for the detection of PSEXEC and SMB brute force lateral movements inside a local network. The technique is based on the modeling

of such threats as finite-state machines, whose state transitions are linked to SMB protocol events (protocol negotiation, session creation, authentication, remote service execution, ...). Such behavioral patterns, implemented as IOC proof of concept, was then successfully validated in a virtual environment. Future work includes porting these signatures as StreaMon [8] modules, as well as the other cyber threats used by real APTs.

## VII. ACKNOWLEDGEMENT

This research was partially supported by the EU Commission within the Horizon 2020 program, SCISSOR project grant no 644425.

## REFERENCES

- [1] "APT Detection – Closing the Gaping Hole", 2014
- [2] Dan Reis, "It's Only the Beginning for Endpoint Security", *FireEye 05*
- [3] A. Oberler et al., "Preventing Pass-the-Hash and Similar Impersonation Attacks in Enterprise Infrastructures," 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, 2016, pp. 800-807.
- [4] "Pass-the-hash attacks: Tools and Mitigation", SANS, 2010
- [5] Jadeja et al., "Implementation and Mitigation of Various Tools for Pass the Hash Attack." *Procedia Computer Science* 79 (2016): 755-764.
- [6] Roesch, Martin. "Snort: Lightweight Intrusion Detection for Networks." *LISA*. Vol. 99. No. 1. 1999.
- [7] Marchetti, Mirco, et al. "Analysis of high volumes of network traffic for Advanced Persistent Threat detection." *Computer Networks* (2016).
- [8] G. Bianchi et al., "StreaMon: A software-defined monitoring platform", *Teletraffic Congress (ITC)*, 2014 26th International, pp.1 -6
- [9] Hutchins et al., "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains" *Leading Issues in Information Warfare Security Research*, vol.1, p.80, 2011.
- [10] M. Usath, D. Jaeger, Feng Cheng and C. Meinel, "Advanced persistent threats: Behind the scenes," 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, 2016, pp. 181-186.
- [11] Mitchell, David. "Intrusion Detection from Simple to Cloud" (2015).
- [12] Sekar, R., et al. "Specification-based anomaly detection: a new approach for detecting network intrusions." *Proceedings of the 9th ACM conference on Computer and communications security*. ACM, 2002.
- [13] "The MsnMM campaign", Kaspersky, 2015
- [14] "Arbanak APT: the great bank robbery", Kasperky, 2015
- [15] "The Dukes: 7 years of Russian cyberespionage", F-Secure, 2015
- [16] Microsoft SysInternals Suite - [technet.microsoft.com/en-us/sysinternals/](http://technet.microsoft.com/en-us/sysinternals/)
- [17] CoreSecurity Impacket - <http://www.coresecurity.com>
- [18] Borza, Viktor, "Generation of MSC Diagrams from Network Traffic" *Bakalárska práca, Masarykova univerzita, Fakulta informatiky* (2013).
- [19] Combs, Gerald. "TShark-The Wireshark Network Analyser"
- [20] "Server Message Block (SMB) Protocol Versions 2 and 3", Microsoft Corporation, release: October 16, 2015
- [21] "Service Control Manager Remote Protocol", Microsoft, 2015
- [22] Metasploit - Scanner SMB Auxiliary Modules "<https://www.offensive-security.com/metasploit-unleashed/scanner-smb-auxiliary-modules/>"
- [23] "Operation Titty Tiger – The Eye of the Tiger", AirBus Defence and Space, 2014
- [24] Raghav, Iit, Shashi Chhikara, and Nitasha Hasteer. "Intrusion Detection and Prevention in Cloud Environment: A Systematic Review." *International Journal of Computer Applications* 68.24 (2013).
- [25] "Mitigating Pass-the-Hash and Other Credential Theft", Microsoft, 2014
- [26] G. Rush et al., "DCAFE: A Distributed Cyber Security Automation Framework for Experiments," *Computer Software and Applications Conference Workshops*, 2014 IEEE 38th International, Vasteras, 2014.
- [27] "Detecting "Pass-the-hash" attacks with Sagan in real time." [https://quadrantsec.com/about/blog/detecting\\_pass\\_the\\_hash\\_attacks\\_wi\\_th\\_sagan\\_in\\_real\\_time/](https://quadrantsec.com/about/blog/detecting_pass_the_hash_attacks_wi_th_sagan_in_real_time/)

# Experimental Low Complexity CubeSat-based Network for Alert Messages Broadcasting

Alessio Fanfani, Simone Morosi, Luca Ronga, Sara Jayousi, Enrico Del Re

Department of Information Engineering of the University of Florence - CNIT

Via S. Marta 3, Florence, 50139, Italy

Telephone: +39 055 4796485

email: alessio.fanfani@unifi.it, simone.morosi@unifi.it, luca.ronga@cnit.it, sara.jayousi@cnit.it, enrico.delre@unifi.it

**Abstract**—This paper presents an end-to-end proof of concept of the low complexity CubeSat-based network architecture for alert dissemination based on the adoption of the MAMES (Multiple Alert Message Encapsulation over Satellite) as alert messages transport protocol and a satellite Portable SDR (Software Defined Radio) receiver. The overall system architecture is described from the generation to the reception of MAMES messages together with the experimental activities carried out to prove the feasibility of the proposed alert network and the applicability of MAMES to LEO satellites.

## I. INTRODUCTION

Recent technologies have now opened the doors for new satellite application and services. Thanks to innovative manufacturing process, e.g. additive manufacturing, and the miniaturization of electronics components has allowed the origin of a new class of satellite, called "Small-Sat", that is growing exponentially in last years. The term "Small-Sat" was initially used for mini and micro satellites with mass lower than 500 kg; however, today often refers to even smaller satellites, namely Nano satellites and Pico satellites or even smaller size [1].

The growing trend of Small-Sat is confirmed by global experts in satellite market consulting and analysis: Euroconsult says that a total of 551 small satellites has been launched over the past 5 years and 1380 are expected to be launched up to 2020 [2]. A quantitative description of the exponential growing trend of the number of small satellites launched per year is reported in [1].

The most common form factor for the small satellite is the CubeSat standard that consist of a cube with an edge length of 10 cm and a mass of up to 1.33 kg. This standard unit has been defined to be a 1U CubeSat and can be combined in structures of these cubes for obtaining bigger CubeSat platform as 2U and 3U. The standardization of CubeSat satellite has allowed a global viral diffusion of them: several launch opportunities with traditional and new generation of launchers are rising all over the world, hundreds of companies and universities are developing satellite sub-system based on COTS electronic and the satellite cost and manufacturing time have been reduced respectively up to few dozen of thousand dollars and a couple of weeks [3]. The cost reduction both for the manufacturing and for the launch recurrent cost as well the brief manufacturing time will benefit the diffusion of satellite constellation. In particular, Small Satellite will generate new

opportunities for the communication services and the Earth observation application that are based on huge number of satellite in Low Earth Orbit (LEO).

The increasing population of satellite leads also to a consequent growing of space debris that currently includes more than 300 million of fragments and about 5,000 of thousands defunct satellites. In order to guarantee a sustainable access to space, it is necessary to stop the systematic increment of non-operative satellites in orbit [4]. This seems to be feasible thanks to preventive debris removal solution, as the Decommissioning Device developed by D-Orbit, whereby a dedicated and independent propulsion system is installed on the satellite prior to launch allowing a safe and quick decommissioning of the spacecraft before it will become a debris [5].

Another enabling factor for the diffusion of next satellite networks is the evolution of Software Defined Radios (SDR) [6], which are software-based radios composed by a wide-band RF front-end plus a direct modulator/demodulator that can work in several sampling rates and central frequencies, controlled by a Field Programmable Gate Array (FPGA) [7]. Data is integrated and sent using Universal Serial Bus (USB) or Ethernet controller to the host or computer, which can be handled by an easy to use common API, GNURadio [8]. Thanks to SDR, all functionalities of an expensive satellite decoder can be implemented in a single board controlled by computer.

The use of the previous mentioned technologies allows the definition of low complexity systems which can be adopted in a large variety of application contexts. Focusing on emergency management scenarios, a growing number of standardization activities and research projects addresses the topic of Emergency Services provided by satellite systems [9] [10] [11] and the exploitation of different communication channels (e.g. Global Navigation Satellite System [12]). The provision of an effective alerting service relies on a timely and reliable distribution of alert messages to the intended audience of the incident affected area. Already existing alerting systems/technologies and an efficient transport of different alert message formats are key elements in the provisioning of an alerting service. To this aim an extensible multiple alert message encapsulation protocol for transporting differently formatted alert messages over satellite links has been defined by ETSI: MAMES (Multiple Alert Message Encapsulation

over Satellite) [13] [14].

This paper proposes a low complexity satellite alert network based on the adoption of: i) a CubeSat for alert messages broadcasting; ii) MAMES protocol for alert messages transport; and iii) SDR technology for messages reception. The experimental activities are carried out in the framework of the D-SAT Mission (II) and aim at demonstrating the feasibility of the proposed low complexity alert network and the MAMES applicability to LEO satellites. The paper is organized as follows: Section II presents the operative context and objectives of D-SAT mission. An overview of the overall architecture of the proposed alerting system and a detailed description of its main components, namely the MAMES Protocol, the Space (CubeSat) and Ground (Satellite Mission Control Center and portable SDR user terminal) Segments, is reported in Section III. Section IV is devoted to the description of the results and current status of the experimental activities. Finally, conclusions are drawn in Section V.

II. OPERATIVE CONTEXT: DSAT MISSION

The proposed satellite alert network will be experimented in the framework of the D-SAT Mission. D-SAT spacecraft has been designed and manufactured by D-Orbit (Italy) and it will be launched by the Polar Satellite Launch Vehicle (PSLV) in LEO Sun-Synchronous Orbit at 500 km altitude in the first quarter of 2017. The Mission includes two main experimental activities: the SatAlert Experiment and the Decommissioning Device Experiment.

1) *SatAlert Experiment: Objectives and Details:* The SatAlert experiment aims to verify the transmission setup of MAMES messages from the MAMES Alert Provider to the MAMES Alert Receiver. The ground infrastructure responsible for the MAMES Message generation and the transmission over the space segment, the on-board software that parses and broadcasts the MAMES messages and the portable SDR UT for the reception of the alert will be validated.

During the visibility period, the Satellite Earth station sends to the satellite a MAMES message (maximum size of 1.5 kB) that will be processed by the satellite and stored in the On-Board Computer memory for broadcasting upon reception of a proper trigger command sent by the Earth station (Fig.1).

According to the international agreements on non-interference regulated by no. 4.4 of the ITU Radio Regulations, MAMES broadcast transmission will be activated only during the available visibility windows in order not to cause harmful interference to other satellite and terrestrial services. The yellow circle in Fig.1 shows the satellite footprint on ground that covers a wide area on ground comprising several nations. For each user on ground, 3 or 4 visibility windows/day (depending on the specific day) and a visibility period of about 480s will be provided.

2) *Decommissioning Device Experiment: Objectives and Details:* D-SAT Mission will demonstrate in orbit the capabilities of D-Orbit Decommissioning device through the satellite decommissioning and controlled atmospheric re-entry. Thus, D-SAT will be the first CubeSat ever to be actively de-orbited



Fig. 1: D-SAT Ground Footprint

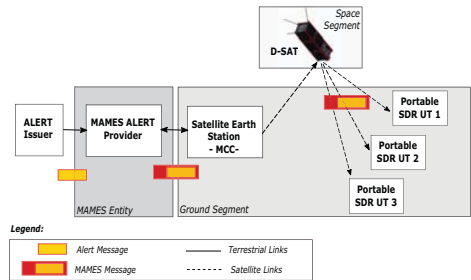


Fig. 2: Overall System Architecture

in a quick, safe, reliable and controlled way. After two months of nominal operations and experimental validation activities of the MAMES Protocol, the D-Orbit Decommissioning Device will be activated and D-SAT will be removed from its operational orbit and disposing it into the atmosphere where it will burn over an uninhabited region of the Earth.

III. SYSTEM OVERVIEW: LOW COMPLEXITY SATELLITE ALERT NETWORK

The overall system architecture developed for the demonstration of a low complexity satellite alert network for MAMES messages broadcasting is depicted in Fig.2. Its main components are:

- *MAMES Alert Provider*, the network entity responsible for encapsulating the alert messages coming from an alert authority (Alert Issuer) into MAMES message as defined in [13].
- *Space Component*, represented by a 3-Unit CubeSat (D-SAT), which is in charge of processing and broadcasting the uploaded MAMES Messages.
- *Ground Component*, which consists of the Satellite Mission Control Center (whose main function is the transmission of the MAMES Messages and the management of the satellite subsystems) and the portable SDR user terminal (used for MAMES messages reception).

As shown in Fig.2, an Alert Protocol Message is sent by the Alert Issuer to the MAMES Alert Provider which encapsulates the received message into a MAMES Message and makes it available (Web Services) for the satellite communication network. The Satellite Mission Control Center (MCC) periodically queries the MAMES Alert Provider for updated MAMES Message and uploads it to the D-SAT. After MAMES message on-board processing (validity and priority checks), D-SAT broadcasts the received message back to Earth. The MAMES Message is then received by an SDR receiver.

**A. MAMES Protocol**

This section presents the MAMES protocol features relevant for the experimental activity, including the MAMES message structure and the network operations.

The MAMES Protocol [13] [14] is defined taking into account the heterogeneity of Alerting Systems and the need of providing a flexible encapsulation scheme for different alert messages (e.g. Common Alert Protocol, unstructured text, image, paging protocols, etc). It defines additional (optional) functions for service extension, enabling the adoption towards a large variety of situations (including network resources limited context). Moreover, it aims to be integrated with the main telecommunication satellite architectures (Galileo Services, DVB-Suite, any IP-based satellite access, etc.) and with already existing terrestrial networks [15].

One of the main objective of the proposed system is to demonstrate MAMES applicability to LEO satellites (as D-SAT is), showing the functioning of a low complexity alert network, from the generation to the reception of MAMES Messages. The experimental activity focuses on the Direct MAMES Alerting operation mode [13]: the Portable SDR UT (User Terminal) directly receives MAMES Messages, acting as the satellite terminal and MAMES Receiver (Fig.2).

**MAMES Messages Design:** Fig.3 reports the extensible structure of the MAMES Frame. It is composed of:

- a set of headers;
- a payload, comprising a concatenation of Alert Protocol Messages (zero, one or multiple Alert Protocol Messages).

About the MAMES Headers, MH (Mandatory Header) and EHs (Extension Headers) pertain to the entire MAMES Message and while the former is of fixed length and contains mandatory information, the latter are optional features (e.g. integrity, encryption, etc.) that enhance the MAMES message. On the other hand, each AMH (Alert Message Header) pertains only to a corresponding Alert Message (AM) contained in the MAMES Payload and specifies the type, language and length of the AM it refers to.

Five types of MAMES Messages are defined: i) ALERT, which enables the encapsulation of a single or multiple Alert Protocol Messages; ii) Ultra-short, which represents the shortest MAMES Message and is defined for the transmission of messages over narrowband satellite channels; iii)UPDATE, which contains updated information for a previously transmitted MAMES Message; iv) CANCEL, which declares a

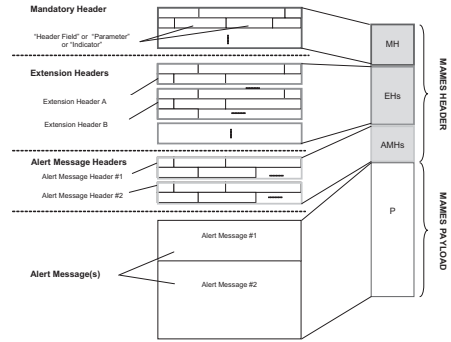


Fig. 3: MAMES message structure [14]

previously transmitted MAMES Message as obsolete.; v)ACK, which provides an acknowledgement at MAMES level for a correctly received MAMES Message.

**Experiment Relevant MAMES Features:** Among the defined MAMES Header fields, it is worth highlighting the role of the priority and time validity information, which are involved in the D-SAT MAMES Message processing, as detailed in the following. The MAMES Frame priority with respect to other MAMES Frames at transport level is managed by the MAMES Transport Priority field of the MH; while the MAMES Validity Start and End fields indicate the time instant when the MAMES Frame shall become valid and invalid, respectively. Upon reception of a MAMES Message from the Satellite Earth Station, both the validity and the priority checks (with the previous stored message, if any) are performed on board of the D-SAT. While upon reception of a trigger command for the transmission, the on-board validity verification of the stored message is carried out before the broadcast of the stored message.

**B. Space Segment (Platform, COM and SW)**

The platform used for the Alert Network proof-of-concept is a 3-Unit CubeSat. D-SAT is designed according to 3U+ CubeSat standard with a mass lower than 4.5 kg and a dimension of about 30 x 10 x 10 cm [16]. The satellite since the first conceptual idealization has been designed to maximize the reliability and consequently the mission success. Therefore, a unique feature of D-SAT is that the main functions of the system are single point of failure free [17].

**Platform:** D-SAT spacecraft is constituted by two CubeSat-class systems, called Unit 1 and Unit 2, as shown in Fig.4. Each unit contains the following subsystems:

- an On-Board Computer (OBC) including Attitude and Orbit Control System (AOCS) sensors (magnetometers, sun sensors, gyroscope) and the drivers for AOCS magnetorques;
- a communication subsystem that consists of an UHF radio module and a turnstile antenna;



- an Electric Power System (EPS) including batteries and solar arrays;
- one commanding board for the control of the decommissioning manoeuvre;
- a GPS receiver to accurately track the re-entry trajectory after the decommissioning maneuver;
- a spin wheel to impart angular momentum about the z-axis of the satellite, for passive stability during fire;
- an Electro Explosive Subsystem (EES) comprehensive of the EED, the electric "Firing Circuit" and mechanical "Safe and Arm Device" part;
- a solid propellant motor delivering about 750 Ns total impulse to the satellite necessary to directly re-entry from LEO orbit to ground in less than 1 orbit period.

In nominal conditions, Unit 1 is the Master Unit acting as hosting satellite and takes control of the spacecraft, while the Unit 2 acts as "Slave", playing the role of an independent Decommissioning Device and it is set to idle mode. Before the decommissioning phase, the control is transitioned to the Slave Unit, which provides a manoeuvre-suitable attitude, a stabilizing spin motion and finally the deorbiting manoeuvre. Thanks to D-SAT architecture, the decommissioning function is possible even if one of the two units is malfunctioning or dead.

**Communication Subsystem:** It is composed by a turnstile omnidirectional antenna and a UHF transceiver module able to transmit up to 1 Watt. The transceiver uses a GMSK modulation over FM signal at a baud rate of 4800 bps: this guarantees a sufficient bandwidth for both telemetry and the service data link. The data are coded with a Reed-Solomon (223,255) coding and randomized according to CCSDS standard [18] to reduce the probability of error and therefore increasing the reliability of the communication link. The UHF communication subsystem has three main functions: i) to transmit a beacon signal (a morse signal FM modulated) as tracking signal; ii) to transmit and receive Telemetry and Telemetry; iii) to transmit and receive MAMES messages.

The link frequency value is 437.505 MHz and it is coordinated by IARU (International Amateur Radio Union) and assigned by ITU. As Transport level, NanoCom U482C uses CubeSat Space Protocol, a small network-layer delivery protocol based on a 32-bit header.

**On-Board Software:** The communication subsystem is supported by the satellite on-board computer, which runs a real-time operating system with a high degree of reliability. Among the different functions of the operating system there are the SatAlert Experiment task (SAE) and the Interface SatAlert Experiment task (ISAE). The ISAE manages the software-hardware interaction between the transceiver and the on-board computer guaranteeing the independence of the experiment from the rest of the mission, while SAE implements the logical functions of the MAMES communication protocol. SAE is invoked by the on-board operating system only in the case of an upload of a new MAMES message or a transmission trigger command from the main ground station. When invoked, SAE receives through the ISAE interface information about

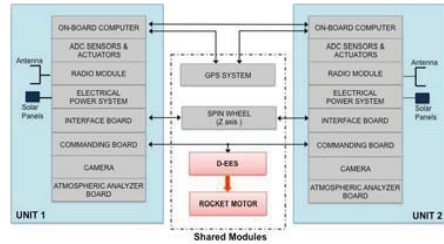


Fig. 4: D-SAT Block Diagram

the uploaded MAMES message, the on-board time and the last MAMES message stored in the transmission queue. Thus, SAE is responsible for MAMES message validation:

- Time validity check of MAMES message previously stored in the transmission queue, in case of transmission trigger command.
- Time validity and priority checks, in case of an upload of a new MAMES message; SAE parses the received message to compare priority and validity of the new uploaded message with the previously stored message.

Based on the validation results, SAE decides which messages have to be transmitted or deleted from the transmission queue, providing this information to the ISAE interface.

The core of On-Board Computer is a flight proven High-performance 32-bit ARM7 CPU, integrated combined with 2MB of static RAM, 4MB of flash memory for data storage and 4 MB of flash memory for code storage. The OBC has an on board timer synchronized with UTC time by means of a GPS receiver within the satellite. The on-board software is also in charge of handling the communication interfaces with other subsystems through data buses (I2C bus, CAN bus and UART).

### C. Ground Segment

**Mission Control Center:** The D-SAT satellite will be operated by a MCC based in the D-Orbit facility in Fino Mornasco, close to Como lake. The main features of the MCC to provide a UHF communication link with satellite are:

- the capability to decode the Morse Beacon Continuous Waive signal and to estimate the frequency of signal transmitted by the satellite affected by Doppler effect and temperature shift;
- the possibility to decode and process telemetry and to transmit telecommand to the satellite through the digital GMSK over FM signal;
- the capability of satellite tracking and frequency prediction to compensate the Doppler effect [19].

Almost all the ground station hardware is COTS amateur radio hardware and was selected to meet the requirements imposed by the uplink and downlink budgets (Table I).

The block diagram of the MCC (Fig.5) includes an equipment that is mounted on the roof-top of the D-Orbit facility

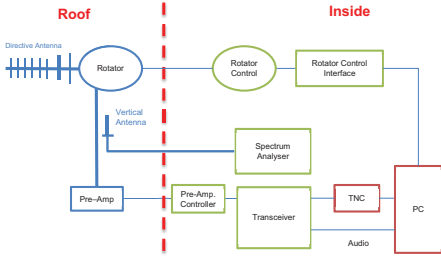


Fig. 5: D-SAT MCC Block Diagram

and hardware inside the D-Orbit control room.

The antenna section consist of a vertical 70 cm antenna with linear polarization and a high-gain Yagi with 19 elements that requires accurate pointing. For instance, the 3dB half power beam width for the directive is about 21 deg: hence, the ground station includes an azimuth/altitude rotator system which provides a pointing accuracy better than 5 deg. Closer to the Yagi antenna is mounted a 20 dB gain low-noise amplifier (LNA) with a noise figure equal to 0.6 dB for receiving weak signals. The vertical antenna provides a wider sky coverage and it mainly used to track and acquire D-SAT signal during the Early Phase of the mission when its position is unknown.

The control room contains the remainder of the equipment for the satellite ground station. The transceiver is a Kenwood TS-2000 featuring a high sensitive receiver and high transmit power output. Thanks to a Serial Data interface connected to a computer running Gpredict satellite tracking program, the transceiver frequency can be tuned to compensate for Doppler shift. [20]. Gpredict is a real-time satellite tracking and orbit prediction application. It uses the satellite Kepler elements released by NORAD to track D-SAT and computes its position, the time of next passes, the azimuth and elevation angle and the Doppler Shift.

The hardware include also a Yaesu G-5500 azimuth and elevator rotators that feed directly into the Rotator Control Port interface with Gpredict. Under this configuration, satellite tracking becomes automated across the sky.

The base baseband modulation-demodulation section, that includes the interface between the computer and the transceiver is delegated to the Terminal Node Controller (TNC). The TNC provides a full-duplex baseband modulation and demodulation using the MSK scheme with baud rates at 4800 baud, a channel coding and decoding function using consisting of a forward error-correction (FEC) with an additional scrambling compliant with CCSDS standard and frame format and routing data function according to the CubeSat Space Protocol (CSP) that is based on a simple header structure [21]

*Portable User Terminal:* Due to the limited availability of resources and infrastructure characterizing an emergency scenario, a satellite portable user terminal based on an low cost SDR solution that enables a simple and flexible configuration

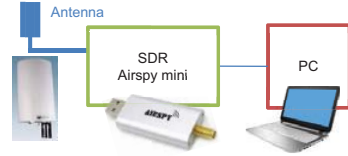


Fig. 6: SDR UT Block Diagram

has been designed for the reception of MAMES messages. The SDR Portable UT (6) includes three main elements: a Circular Polarized Quadrifilar Helix Antenna, an Airspy Mini SDR radio and a personal computer configured with GNU Radio software. The choice of the Quadrifilar Helix Antenna instead of a directive one is justified by the operative scenario and the impossibility of using a rotator mechanism to point the satellite. The Helix Antenna ensures a semi-spherical radiation pattern and, although the maximum antenna gain is limited up to 5 dB the  $E_b/N_0$  value (computed with standard link budget [19]), it is still enough for correctly decoding the signal as shown in Table I. The used SDR-platform is the AirSpy Mini SDR, supported by a GNU Radio implementation on a computer. The main performance of the AirSpy Mini SDR are: a 3.5 dB Noise Figure in UHF frequency band, a 35dBm IIP3 RF front end, a 12bit ADC with a Dynamic Range up to 80dB and a 0.5 ppm high precision, low phase noise clock.

The host computer is a Linux system running GNU Radio, an open source radio toolbox with digital signal processing functions and hardware drivers. It is the framework used to implement the ground station software. A GNU Radio application usually consists of several modules, called blocks, each performing a specific task. They are connected in a chain, forming a flowgraph, where each block consumes and produces data and the resulting program is a runnable application, referred to as the top block [8]. GNU radio performs all remaining operations of the receiver: modulation and demodulation, error detection and correction, as well as clock synchronization and Doppler compensation [22].

#### IV. STATUS AND RESULTS

The D-SAT Qualification Model (QM) has successfully passed the qualification tests campaign following the Launch Service Provider requirements, including mechanical (Sine and Random Vibration, Shock and Acceleration Test) and thermal-vacuum test. The D-SAT Flight Model (FM) integration has been completed, the acceptance test campaign will be performed in the next month and the D-SAT FM satellite will be delivered to the Launch Service Provider (LSP) within the end of the year. In January, the satellite will be integrated within the P-POD deployer [16], and shipped to the PSLV's launch site in India. The expected launch date is in March 2017. Fig.7 shows both D-SAT models in D-Orbit Clean Room facility.

During functional test activities an extended test campaign of Sat Alert operating functions (MAMES Message generation and on-board processing) has been performed with good

Link Budget Parameters	MCC	SDR UT	Unit
	Value	Value	
Satellite Range	909.5	909.5	km
Elevation Angle	30	30	deg
Frequency	437.5	437.5	MHz
Transmission Rate	4800	4800	bps
<b>Spacecraft</b>			
Transmitted Power	1	1	W
Antenna Gain	0.4	0.4	dBi
Line Losses	-1.7	-1.7	dB
Pointing Losses	-3	-3	dB
<b>Downlink</b>			
PathLoss	-144.4	-144.4	dB
Polarization Loss	-3	0	dB
Atmospheric Losses	-1.8	-1.8	dB
<b>Ground Station</b>			
Antenna Gain	17.6	5	dBi
HPBW	21	180	deg
Figure of Merit, (G/T)	-10.4	-10.3	dB/K
<b>Link Budget Results</b>			
Received Isotropic Power	-113.5	-113.5	dBm
$E_b/N_0$	28.3	18	dB
Required $E_b/N_0$ @ BER = $10^{-5}$	13.5	13.5	dB
Margin $E_b/N_0$	14.8	4.5	dB

TABLE I: Downlink Budget Table

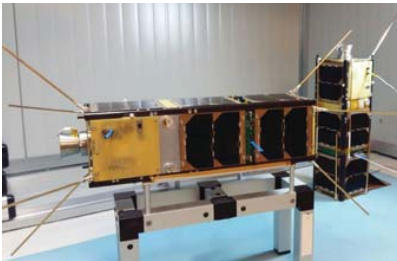


Fig. 7: D-Sat FM and QM - courtesy of D-Orbit

results. The end-to-end communication between MCC and Portable SDR user terminal was experimented in D-Orbit facility with all the real equipment that will be used during D-SAT mission and emulating the path-loss attenuation using power attenuators in receiving chain.

V. CONCLUSIONS

An experimental low complexity alert satellite network based on the adoption of a CubeSat, MAMES protocol and SDR implementation has been presented. The experimental activities have proven the feasibility of the proposed alert network and the applicability of MAMES to LEO satellites.

VI. ACKNOWLEDGEMENT

This study is the result of a cooperation between CNIT (University of Florence) and D-Orbit. The D-SAT project was initially financed by D-Orbit investments and from December 2015 has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 711193.

REFERENCES

- [1] M. Buscher, T. Funke, A. Patzphal, K. Brie, "Satellites Might go viral - Data Proves It", Final Proceedings at Small Satellites, System & Services Symposium (4S), 30 May -3 June 2016, Valletta, Malta
- [2] R. Villain, "Facts & figures on possible future directions of the small-sat market", Final Proceedings at Small Satellites, System & Services Symposium (4S), 30 May -3 June 2016, Valletta, Malta
- [3] N.H. Crisp, K. Smith, P. Hollingsworth, Launch and deployment of distributed small satellite systems, Acta Astronautica, vol 114, pp. 65-78, 2015
- [4] European Code of Conduct for Space Debris Mitigation - June 2004 - Issue 1.0.
- [5] S. Antonetti, L. Ferrario, E. Toson and L. Rossetini, "Implementation of Debris Prevention and Reduction Measures by an Independent Propulsive Decommissioning Device", 66th International Astronautical Congress, Jerusalem, 12 - 16 October 2015.
- [6] M. Panizza, C. Sacchi, J. Varela-Miguez, S. Morosi, L. Vettori, S. Digenti, E. Falletti, "Feasibility study of a SDR-based reconfigurable terminal for emergency applications", Proc. of the 20011 IEEE Aerospace Conference, Mar. 2011, Big Sky, Montana, USA.
- [7] T. Ulvversy, "Software defined radio on digital communications: A new teaching tool", Wireless and Microwave Technology Conference (WAMICON), 2012 IEEE 13th Annual , vol 12 , no 978-1-46730128-2, p.18
- [8] P. Dominguez, Ian, Software Defined Radio: USRP y GNU Radio. Ivn Pinar Dominguez1 ed , Vol. 6.2011.
- [9] E. Del Re, S. Morosi, S. Jayousi, C. Sacchi, "SALICE Satellite-Assisted Localization and Communication systems for Emergency services", Proc. of the 1st International Conference on Wireless Communications, Vehicular Technology, Information Theory, Aerospace and Electronic Systems (WVITAE), Aalborg, Denmark, 2009.
- [10] E. Del Re, S. Jayousi, S. Morosi, L.S. Ronga, M. De Sanctis, E. Cianca, M. Ruggieri, E. Falletti, A. Iera, G. Arantici, C. Sacchi, "SALICE project: Satellite-Assisted Localization and Communication Systems for Emergency Services", IEEE Aerospace and Electronics Systems Magazine, Vol. 99, n. 9, Sept. 2013, pp. 4-15.
- [11] T. Pecorella, L.S. Ronga, F. Chiti, S. Jayousi, L. Franck, "Emergency satellite communications: research and standardization activities", Communications Magazine, IEEE, vol.53, no.5, pp.170,177, May 2015.
- [12] D. Iwazumi, T. Ishida, S. Iino, N. Kohtake, P. Buist, "GNSS-based emergency message service: Lessons learned and future prospects", Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC), 2014 7th, vol., no., pp.276- 283, 8-10 Sept. 2014.
- [13] ETSI, "Satellite Earth Stations and Systems (SES); Satellite Emergency Communications; Multiple Alert Message Encapsulation over Satellite (MAMES)", ETSI, TS 103 337, May 2015.
- [14] ETSI, "Satellite Earth Stations and Systems (SES); Satellite Emergency Communications (SatEC); Multiple Alert Message Encapsulation over Satellite (MAMES) deployment guidelines ", ETSI, TR 103 338, May 2015.
- [15] L. S. Ronga, S. Jayousi, R. Pucci, S. Morosi, M. Berlioli, J. Rammer, A. Fanfani, S. Antonetti, "Multiple Alert Message Encapsulation Protocol: Standardization and Experimental Activities", Proceedings of the ISCRAM 2015 Conference - Kristiansand, May 24-27 - Palen, Bscher, Comes & Hughes, eds.
- [16] "CubeSat Design Specification" Rev. 13, The CubeSat Program, Cal Poly SLO
- [17] E. Toson, A. Fanfani, S. Morosi, L. S. Ronga, P. Martegani, L. Rossetini, "D-SAT Mission: An In-Orbit Demonstration of a solid propellant satellite decommissioning device", The 4S Symposium 2016
- [18] CCSDS Recommended Standard 131.0-B-2 TM Synchronization and Channel Coding
- [19] G. Maral and M. Bousquet, "Satellite Communications Systems: Systems, Techniques and Technology", Pearson Education India.
- [20] C. Cooper, R. Fevig, J. Patel, "The CubeSat Ground Station at the University of Arizona"
- [21] Technical Datasheet: GS-DS-TNC-4.0 NanoCom TNC1 Datasheet Base-band MSK modem for satellite ground stations
- [22] A. Lofaldli, R. Birkeland, "Implementation of a Software Defined Radio Prototype Ground Station for CubeSats", Final Proceedings at Small Satellites, System & Services Symposium (4S), 30 May -3 June 2016, Valletta, Malta

# A coherence study on EEG and EMG signals

Giulia Cisotto, Umberto Michieli, Leonardo Badia  
 Dept. of Information Engineering, University of Padova  
 via Gradenigo 6B, 35131 Padova, Italy  
 email: {giulia.cisotto, leonardo.badia}@dei.unipd.it

**Abstract**—The aim of this study is to investigate bursts-related EEG signals in a focal hand dystonia patient. Despite of considering time domain and frequency domain techniques as mutually exclusive analysis, in this contribution we have taken advantage from both of them: particularly, in the frequency domain, coherence was used to identify the most likely frequency bands of interaction between brain and muscles; then, in the time domain, cross-correlation was exploited to verify the physiological reliability of such a relationship in terms of signal transmission delay from the centre to the periphery. Our preliminary results suggest - in line with recent literature - that activity in the high  $\beta$  band (around 30 Hz) could represent an electroencephalographic correlate for the pathological electromyographic bursts affecting the focal hand dystonia condition. Even though a future study on a larger sample is needed to statistically support these preliminary findings, this contribution allows to think of new kinds of rehabilitation from focal hand dystonia that could target the actual electroencephalographic correlate of the pathology, i.e. phenotypically expressed by bursts, with the consequence of a relevant functional improvement.

## I. INTRODUCTION

There is a wide literature robustly reporting on how a voluntary motor output is prepared and driven by the central nervous system, the sensorimotor circuit, in particular. Brain signals recorded at several depths of the brain, i.e. (from outer to inner layer) the scalp, the dura, the cortex and the basal ganglia, have shown robust patterns of activation and deactivation in specific regions, frequency bands and time periods.

Specifically, power decrease at the contralateral hand-related side (to the movement) in the so-called  $\mu$  and  $\beta$  bands, around 10 Hz and 20 Hz, respectively, occurs as soon as 1s before movement onset. This phenomenon is known as *event-related desynchronization* (ERD) [1].

In the time domain, time-locking each electroencephalographic (EEG) signal on its corresponding electromyographic (EMG) activation onset, and averaging among several responses, a typical waveform could be observed: indeed, the so-called *readiness potential* starts as soon as 1.5 to 1 s before movement onset with a slow decrease of signal amplitude; other known components follow, each of them with a specific clinical meaning. This complex behavior, overall, is labelled as *movement-related cortical potential* (MRCP) [2].

In case of neuro-motor pathologies, ERD and MRCP could become carriers of important information related to abnormal behaviours of the patient. Particularly, in case of motor disorders, where movements are often involuntarily produced, a

central, i.e. conscious, control has been suggested [3] [4] [5], but not consistently proved and accepted, yet.

The *Jerk-locked back averaging* (JLBA) technique has been effectively employed on MRCP to identify the central origin of a specific kind of involuntary movements observed in myoclonus, Tourette's syndrome and other psychogenic motor disorders [6]. Thanks to such technique, a sharp biphasic waveform could be consistently seen on the averaged EEG signal, especially at the central and contralateral areas. Moreover, this EEG *potential* anticipated the EMG onset by 15 to 20 ms on average [7] [8].

Later on, other techniques have been utilized to quantify the influence of brain activity on the motor output. Moreover, coherence between EEG and EMG has been computed in many studies for different kinds of patients: widely-known as *cortico-muscular coherence* (CMC) [9], it is usually evaluated during sustained contractions at a predetermined level, e.g. 15 to 20 percent of the maximal voluntary contraction, in order to ensure stable motor units engagement.

However, CMC was also employed in few recent works on Parkinson's disease resting tremor [10][11]: in those studies, no movement was accomplished by the patients but their postural tremor was recorded by EMG together with synchronous EEG or magnetoencephalographic (MEG) activity.

A clear peak of activity at the frequency of tremor and its second harmonic, around 5 Hz and 10 Hz respectively, could be seen in the EMG power spectrum; moreover, CMC showed a statistically significant peak of coherence between the EEG signal recorded from the contralateral hand-related scalp area and the fingers extensor muscle.

Therefore, in case of rhythmic pathological behaviour, a relationship between central and peripheral activity could be significantly quantified at rest, too.

The aim of this study is to apply a similar concept to the investigation of bursts-related EEG signals in one focal hand dystonia (FHD) patient.

FHD is a movement disorder that causes people who are affected by it to experience an abnormal and involuntary co-contraction of the agonist and antagonist muscles of the hand and the forearm. It has been shown to originate in the central nervous system and to cause abnormal patterns of brain activation, especially in the  $\beta$  band (as suggested by recent literature) [12].

Bursts are abrupt and giant involuntary muscular contraction events that typically affect EMG of this kind of patients,

especially at rest when they largely exceed the very low background activity.

In this contribution, CMC as well as cross-correlation have been computed between EEG and EMG to assess, both in the frequency and in the time domain, the effect of a central driver onto the motor output. Our preliminary results show the influence of EEG on pathological EMG oscillatory activity. Coherence was employed to identify the most involved frequencies, while cross-correlation was used to support the physiological meaning of such EEG-EMG relationship.

In the rest of the paper, section II will present the methods of the study, section III the most interesting preliminary results, while the final section IV will discuss them in comparison with existing literature on the topic along with an overview of some limitations to be overcome in the future; a perspective view for clinical applications in motor rehabilitation will be provided, also.

## II. MATERIALS AND METHODS

One FHD patient and one healthy subject (HS) were involved in this pilot study. The EEG was recorded from one monopolar EEG channel placed on C3, the standard location of the International 10-20 System over the left hemisphere corresponding to the brain region related to the functioning of the (dominant) right-hand. The EMG was recorded from one bipolar channel placed on the *abductor pollicis brevis*, the intrinsic hand muscle responsible for the abduction of the thumb. Both signals were sampled with a sampling frequency of 1 kHz and quantized at 16 bit. In the experiment, the participants were sitting quietly on a comfortable chair in front of a screen placed 1 meter apart from them, on a table. They were simply required to rest with opened eyes for about 3 minutes with their limbs laying on the table in front of them.

At a first glance, it was possible to assess a clear difference between the two EMG signals: in the FHD patient, the amplitude of the signal assumed values up to  $\pm 200 \mu\text{V}$ , while its power spectral density (PSD) took significant values in the frequency band [5, 200] Hz. However, in case of HS, the amplitude of the EMG signal did not exceed  $\pm 20 \mu\text{V}$ , with a significant PSD extended from 5 to 50 Hz.

In the offline analysis, signals were preprocessed to limit their frequency range in the frequency band of interest. Specifically, the EEG was filtered through an elliptic filter of order 24 with a passband of [4, 45] Hz. The EMG was processed by a high-pass elliptic filter of order 11 with cut-off frequency at 5 Hz. A series of notch filters of order 14 were used with cut-off frequencies at 50 Hz and subsequent harmonics up to 350 Hz were put to reduce the effect of the mains.

Then, CMC as well as cross-correlation have been computed between the EEG (otherwise labelled as signal  $x[m]$ ) and the EMG (otherwise labelled as signal  $y[m]$ ) signals, in order to assess the quantitative relationship between them, both in the frequency domain and in the time domain.

### A. Frequency domain analysis: cortico-muscular coherence

The coherence of two discrete-time signals  $x[m]$  and  $y[m]$ , regarded as stochastic processes, is given by:

$$\text{Coh}_{xy}(f) \triangleq \frac{\mathcal{P}_{xy}(f)}{\sqrt{[\mathcal{P}_x(f)]} \cdot \sqrt{[\mathcal{P}_y(f)]}}, \quad (1)$$

where  $\mathcal{P}_x(f)$  is the PSD of  $x[m]$  and  $\mathcal{P}_{xy}(f) = \frac{1}{n} \sum_{i=1}^n X_i(f)Y_i^*(f)$  is the cross-power spectral density (CPSD) between  $x[m]$  and  $y[m]$ .

In order to provide a statistically significant result, a confidence level  $CL$  of 95 %, i.e. a critical level of  $\alpha = 0.05$ , was obtained from the following formula [13]:

$$CL = 1 - (1 - \alpha)^{\frac{1}{N-1}}, \quad (2)$$

where  $N$  is the number of signal segments used to estimate the coherence value.

The PSD, the CPSD and the coherence values were estimated via the Fast Fourier Transform (FFT)-based Welch's method: specifically, the signal length was set to  $L = 1024$  samples (1.024 s) and the number of FFT points to 1024 samples. Border effects were mitigated by a Hann sliding windowing with overlap of 50 % [14].

### B. Time domain analysis: cross-correlation function

Generally speaking, given two discrete-time signals  $x[m]$  and  $y[m]$ , their cross-correlation function is defined as:

$$r_{xy}[n] \triangleq \sum_{m=-\infty}^{+\infty} x^*[m]y[n+m]. \quad (3)$$

Cross-correlation is particularly useful to evaluate the similarity between two signals as a function of the time shift  $n$  (expressed in number of samples) of the second signal behind the first one.

In the present analysis, the absolute value of the correlation between the EEG and the EMG signals computed at its maximum and normalized by the square root of the product of the signals energies  $E_x$  and  $E_y$  was evaluated. Therefore, the quantity:

$$r_{max} = \frac{\max(r_{xy}[n])}{\sqrt{E_x E_y}}. \quad (4)$$

was considered as a measure of similarity between the two signals.

Moreover, the lag  $n$  which the maximum was found at was taken into account as a measure of the transmission delay from the brain to the muscle, i.e. the time taken for a motor command to travel from its origin in the central nervous system to the target effector at the periphery.

Particularly, 71 pairs of EEG and EMG signals were extracted from the whole EEG and EMG recordings of the FHD patient. They were selected empirically as examples of bursty EMG activity (with their corresponding EEG). The duration of these signals was variable ( $0.70 \pm 0.66$  s): all of them were included in the analysis to keep into account the variability of the burst events.

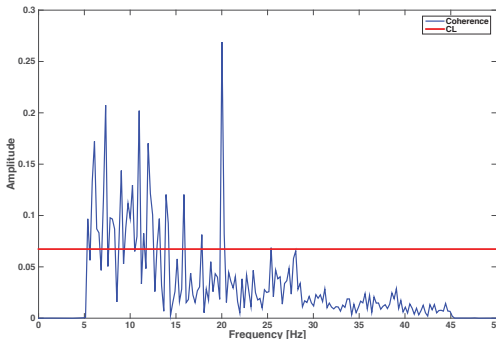


Fig. 1. Absolute value of CMC for the healthy participant ( $CL = 0.067$  with  $N = 44$  and  $\alpha = 0.05$ ).

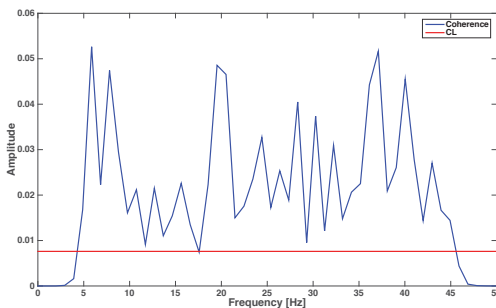


Fig. 2. Absolute value of CMC for the pathological subject ( $CL = 0.008$  with  $N = 393$  and  $\alpha = 0.05$ ).

To support the physiological meaning of the EEG-EMG coherent components, the cross-correlation function was computed between the narrow-band EEG signals filtered in the high  $\beta$  band, i.e. between 26 and 31 Hz, and the EMG signal limited to 250 Hz by a band-pass filter with frequency band [5, 250] Hz.

### III. RESULTS

#### A. EEG-EMG coherence

In this section the results are shown in regards to the CMC measure for both the HS and the FHD patient. In the case of the healthy participant, the CMC spectrum could be seen in Fig. 1. It has to be noted that peaks above the confidence level can be observed in the frequency range between 5 and 20 Hz, only, with a particularly strong coherence at 20 Hz.

On the other hand, the CMC spectrum of the FHD patient is reported in Fig. 2. It can be observed that a larger frequency band contribute to the coherence between EEG and EMG signals. It is also important to highlight the presence of peaks in the upper side of the spectrum, i.e. [20, 45] Hz. This is

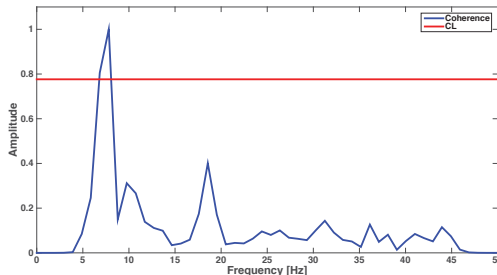


Fig. 3. CMC, in absolute value, between chunks of healthy-like EEG and EMG for the pathological subject ( $CL = 0.776$  with  $N = 3$  and  $\alpha = 0.05$ ).

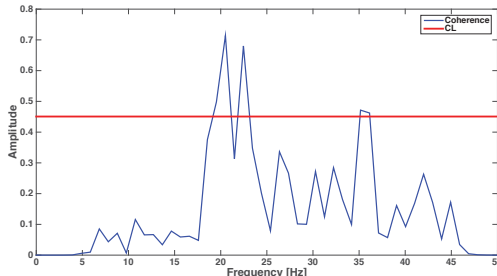


Fig. 4. CMC, in absolute value, between chunks of bursts-affected EEG and EMG for the pathological subject ( $CL = 0.451$  with  $N = 6$  and  $\alpha = 0.05$ ).

probably due to the larger bandwidth of the pathological EMG of the patient, as mentioned in the previous section.

In order to confirm our hypothesis, we selected a portion of the whole recorded data where bursts mostly affected the EMG signal and evaluated the CMC in this particular case. As a further support, we selected another portion of EMG signal where healthy-like activity could be observed and computed CMC as well. Two typical examples of both situations are reported next. Fig. 3 shows the coherence result when comparing two chunks of the EEG and EMG signals for the healthy-like case. Here, two main peaks can be seen at the frequencies of 8 Hz and 18 Hz, but no significant coherence values at frequencies higher than 30 Hz.

On the contrary, Fig. 4 reports the coherence spectrum in case of bursts-affected chunks. Significantly, the figure shows that coherence values at low frequencies are heavily reduced, whereas some peaks around 20 and 35 Hz appeared, hence the hypothesis that higher frequencies components are related to bursty EMG activity in the FHD patient could actually be supported.

#### B. EEG-EMG cross-correlation

As mentioned above, cross-correlation of EEG and EMG was computed to investigate the physiological reliability on

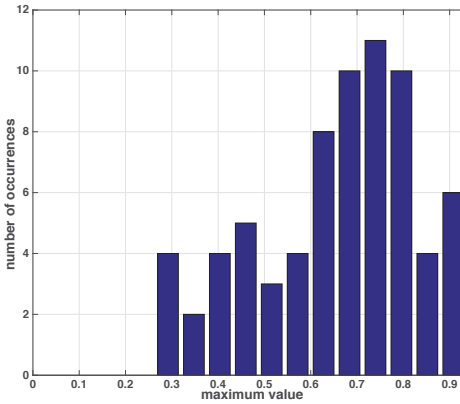


Fig. 5. Empirical distribution of the maximum value of the cross-correlation function of the 71 pairs of EEG and EMG signals of the FHD patient.

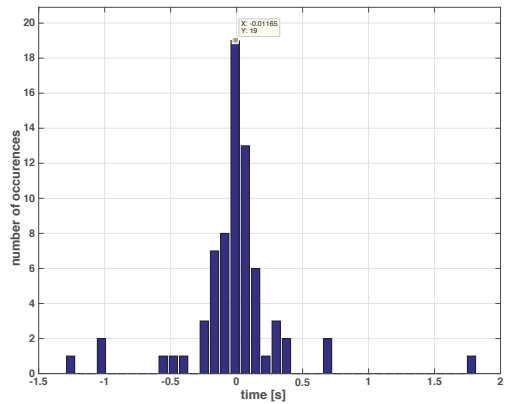


Fig. 6. Empirical distribution of the lags where the maximum of the cross-correlation was found.

the relationship between the high  $\beta$  band EEG component with the EMG.

Fig. 5 reports the empirical distribution of the maximum values of the cross-correlation function found from the 71 pairs of EEG and EMG signals. Mean value was found of 0.683, with variance of 0.0293. This result certainly shows a strong connection between the narrow-band EEG and the EMG, as indicated by the high average value.

Finally, Fig. 6 displays the empirical distribution of the lag where the maximum value of the cross-correlation function of the 71 pairs of EEG and EMG signals was found. Mean value occurred at  $-11.65$  ms. As neural impulses propagate at a speed of about 100 m/s, the transmission of signals from the brain to the hand muscles could be estimated of about 10 ms, which is in line with the results we achieved.

The standard deviation is considerably high (it was found to be about 100 ms) but this could be explained because of the limited size of the data sample. Indeed, we expect that the tendency observed in this study could be further confirmed (with a reduced standard deviation), with an increased sample size.

#### IV. DISCUSSION

FHD and other movement disorders show jerks during movement or even during rest. Identifying possible central drivers of such abnormal muscular activity is a very relevant issue and could represent a key aspect for improving diagnosis and rehabilitation. JLBA is the only technique that could directly reveal a causal relationship between an average brain activity and an average muscular response in many pathological cases.

However, JLBA is not always applicable and requires many chunks of EEG and EMG to be robust. Moreover, identification of jerks and averaging among jerk-locked EMG signals raise

some technical questions in its application to FHD, as reported by other literature [15][16]. Specifically, the following critical points were highlighted:

- high-frequency and irregularity of the jerks might prevent their correct identification (EMG activation should be absent for 100ms at least before the EMG burst);
- similarly, the absence of giant somatosensory evoked components in many subjects could prevent success of the JLBA procedure;
- frequency domain-based analysis have been successfully proposed to identify EEG correlates of myoclonic jerks, even in case where time domain JLBA has failed.

Since similarity could be noted between EMG bursts seen in FHD and jerks present in myoclonus, many considerations explained above hold true for analysis of EMG bursts in case of FHD.

Therefore, in this study we proposed the use of CMC in case of FHD, too. Besides, we do not believe time and frequency-domain analysis to be mutually exclusive but, on the contrary, we claim the opportunity to use them to complement each other. Therefore, in this contribution, coherence analysis was used to identify the most likely frequency bands of central-periphery communications; then, cross-correlation function was employed to verify the physiological reliability of such relationship (as suggested by [7]).

In line with well-known literature about CMC [17] [13], we found a peak-component at 20 Hz ( $\beta$  band) together with lower peaks at lower frequencies ( $\theta$  and  $\alpha$  bands) in the HS.

On the contrary, the FHD patient showed different CMC patterns in different behavioral scenarios: specifically, when a healthy-like recording period is selected, CMC spectrum displayed two major peaks, one around 5 Hz and the other one at 20 Hz. In this case, we could advance the hypothesis - supported by literature [10] [11] - that the slower component

is related to the postural tremor affecting this subject, while the second one seems to reflect the healthy-like  $\beta$  band synchronous activity between EEG and EMG that has been seen in the control subject, as well. Worthy of notice is the fact that the tremor-related component significantly prevails onto the other one and it is the only one to exceed the confident level of 95 %. When an EMG period heavily characterized by the presence of bursts is selected, we could observe two major frequency components in the CMC spectrum: the first one was around 20 Hz ( $\beta$  band), it exceeded the confidence level of 95 % and it could resemble the healthy-like  $\beta$  band component of the normal condition. The second one, interestingly, was significant at the 95 % as the first component, and could be suggested to represent the correlate of the bursty activity observed in the pathological EMG of the patient.

An average delay of 10 ms of EMG behind EEG has been observed through the cross-correlation analysis and it is in line with physiology and other literature [13].

Overall, our preliminary results claimed that activity in the high  $\beta$  band [25,35] Hz could represent an EEG correlate for the EMG bursts affecting the FHD condition.

Even though the small size of the data sample could be a limitation to the study, our conclusions were primarily driven by our data analysis, but their reliability was strongly supported by existing literature about similar studies accomplished with larger samples of data and patients, even though in slightly different diseases, e.g. myoclonus and Parkinson's disease.

Even though a future study on a larger sample is needed to statistically support these preliminary findings, this contribution allows to think of new kinds of rehabilitation interventions [18] [19] for focal hand dystonia patients that could target the actual EEG correlate of the pathology, i.e. phenotypically expressed by bursts, with the consequence of a relevant functional improvement.

## V. CONCLUSIONS

The aim of this study is to investigate bursts-related EEG signals in a focal hand dystonia patient. Despite of considering time domain and frequency domain techniques as mutually exclusive analysis, in this contribution we have taken advantage from both of them: particularly, in the frequency domain, cortico-muscular coherence was used to identify the most likely frequency bands of interaction between brain and muscles; then, in the time domain, cross-correlation was exploited to verify the physiological reliability of such a relationship in terms of signal transmission delay from the centre to the periphery. The most interesting result suggested that the high  $\beta$  band activity in the EEG could be responsible for the bursty activity observed in the EMG. Even though a future study on a larger sample is needed to statistically support these preliminary findings, this contribution allows to think of new kinds of rehabilitation interventions for focal hand dystonia patients that could target the actual EEG correlate of the pathology with consequence improvement of the motor functions.

## REFERENCES

- [1] G. Pfurtscheller and F. H. Lopes Da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842-1857, 1999.
- [2] H. Shibasaki, G. Barrett, E. Halliday, and A. Halliday, "Components of the movement-related cortical potential and their scalp topography," *Electroencephalography and Clinical Neurophysiology*, vol. 49, no. 3, pp. 213-226, 1980.
- [3] P. T. Lin and M. Hallett, "The pathophysiology of focal hand dystonia," *Journal of Hand Therapy*, vol. 22, no. 2, pp. 109-114, 2009.
- [4] M. H. Ruiz, P. Senghaas, M. Grossbach, H.-C. Jabusch, M. Bangert, F. Hummel, C. Gerloff, and E. Altenmüller, "Defective inhibition and inter-regional phase synchronization in pianists with musician's dystonia: an EEG study," *Human brain mapping*, vol. 30, pp. 2689-700, Aug 2009.
- [5] H. Shibasaki and M. Hallett, "Electrophysiological studies of myoclonus," *Muscle and Nerve*, vol. 31, no. 2, pp. 157-174, 2005.
- [6] H. Shibasaki and Y. Kuroiwa, "Electroencephalographic correlates of myoclonus," *Electroencephalography and Clinical Neurophysiology*, vol. 39, no. 5, pp. 455-463, 1975.
- [7] H. Shibasaki, "Cortical activities associated with voluntary movements and involuntary movements," *Clinical Neurophysiology*, vol. 123, no. 2, pp. 229-243, 2012.
- [8] G. Avanzini, H. Shibasaki, G. Rubboli, L. Canafoglia, F. Panzica, S. Franceschetti, and M. Hallett, "Neurophysiology of myoclonus and progressive myoclonus epilepsies," *Epileptic Disorders*, vol. 18, no. September, pp. 11-27, 2016.
- [9] T. Mima, S. Ohara, and T. Nagamine, "Corticomuscular coherence," *International Congress Series*, vol. 1226, pp. 109-119, 2002.
- [10] L. Timmermann, J. Gross, M. Dirks, J. Volkmann, H. J. Freund, and A. Schnitzler, "The cerebral oscillatory network of parkinsonian resting tremor," *Brain*, vol. 126, no. 1, pp. 199-212, 2003.
- [11] B. Hellwig, S. Häußler, B. Scheller, M. Lauk, B. Guschlbauer, J. Timmer, and C. Lücking, "Tremor-correlated cortical activity in essential tremor," *The Lancet*, vol. 357, no. 9255, pp. 519-523, 2001.
- [12] F. Tecchio, J. M. Melgari, F. Zappasodi, C. Porcaro, D. Milazzo, E. Cassetta, and P. M. Rossini, "Sensorimotor integration in focal task-specific hand dystonia: A magnetoencephalographic assessment," *Neuroscience*, vol. 154, no. 2, pp. 563-571, 2008.
- [13] T. Mima and M. Hallett, "Electroencephalographic analysis of corticomuscular coherence: reference effect, volume conduction and generator mechanism," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1892-1899, 1999.
- [14] S. K. Mitra, *Digital signal processing: a computer-based approach*. McGraw-Hill, 2007.
- [15] P. Brown, S. F. Farmer, D. M. Halliday, J. Marsden, and J. R. Rosenberg, "Coherent cortical and muscle discharge in cortical myoclonus," *Brain*, vol. 122, no. 3, pp. 461-472, 1999.
- [16] Y. Xu, M. V.M., C. Z., and M. K.R., "Cortico-Muscular Coherence with Time Lag with Application to Delay Estimation," *IEEE Transactions on Biomedical Engineering*, May 2016.
- [17] T. Mima and M. Hallett, "Corticomuscular coherence: a review," *Journal of Clinical Neurophysiology*, vol. 16, no. 6, pp. 501-511, 1999.
- [18] K. von Carlowitz-Ghori, Z. Bayraktaroglu, G. Waterstraat, G. Curio, and V. V. Nikulin, "Voluntary control of corticomuscular coherence through neurofeedback: A proof-of-principle study in healthy subjects," *Neuroscience*, vol. 290, pp. 243-254, 2015.
- [19] Y. Hashimoto, T. Ota, M. Mukaino, M. Liu, and J. Ushiba, "Functional recovery from chronic writer's cramp by brain-computer interface rehabilitation: a case report," *BMC neuroscience*, vol. 15, p. 103, jan 2014.



# EGNSS High Accuracy System Improving Photovoltaic Plant Maintenance using RPAS integrated with Low-cost RTK Receiver

Marco Nisi, Fabio Menichetti

*Sistematica S.p.a, Via Bramante, 43, 05100 Terni TR, Italy*

Bilal Muhammad, Ramjee Prasad

*CTIF Section, Department of Electronic Systems, Aalborg University, Denmark*

Ernestina Cianca

*CTIF Section, Department of Electronics, University of Rome Tor Vergata, Italy*

Alberto Mennella, Graziano Gagliarde

*TopView srl, Via Santa Croce, 9, San Nicola La Strada, 81020 (CE), Italy*

Davide Marenchino

*Entec S.p.a, Via Montegrappa, 2, 12038 Savigliano (CN), Italy*

**Abstract**—Global Navigation Satellite System (GNSS) Real Time Kinematic (RTK) is the key enabling technology for a number of applications demanding very high positioning accuracy as their operational requirement. This include, but not limited to, mapping, surveying, robot guidance, and precision agriculture to name a few. Typically, GNSS RTK employs high-end dual-frequency receivers and antennas to deliver precise positioning that, in some way, restricts the use of GNSS RTK to a subset of user market due to very high cost. The emerging mass-market user applications, however, require centimeter positioning accuracy considering a cost-effective solution. This calls for low-cost GNSS RTK technology to create new possibilities for mass-market user applications to make use of GNSS high accuracy positioning in a variety of ways. One of the applications that make use of low-cost RTK is EGNSS high accuracy system for improving photovoltaic (PV) plants maintenance. The proposed solution aims at automating the maintenance of PV plant with enhanced reliability in a time and cost effective manner, which otherwise requires intermediate human intervention. This paper presents system architecture, design, and core algorithm that plays a pivotal role in enabling the automatic report generation of PV plant status.

**Index Terms**—Photovoltaic; RPAS; GNSS; RTK

## I. INTRODUCTION

PV plants experience aging over a time span, which results in lower production mainly due to inefficiencies caused by faults and or defects in unknown modules. PV plant owners are responsive about enhancing their energy production and recognize the importance of maintenance in the life cycle of a PV plant according to the policies of incentives activated in the past years inside the European Union member states in favor of clean energy production. However, the inspection of faulty modules over a PV plant is still achieved by using hand-held thermal camera, kept by a person walking over the whole PV plant, on the ground or on the rooftop. This method has itself significant drawbacks, for instance, the need for ground

plants, walking over large spaces of wild grass, snow, and rough terrain. Another approach is realized by the installation of expensive systems, including sensors communicating with a center station, allowing the detection of a series of modules over a PV array, with no information about the unique faulty module inside the string. Nevertheless, even in this case human intervention is needed in anomaly identification. The maintenance becomes even more cumbersome for those PV plants that are installed in rural areas where communication and surveillance infrastructure is hard to deploy; consequently a dedicated crew or service is required.

To minimize the operational complexities in maintaining a PV plant, the proposed solution presented in this paper aims at automating the overall process of maintenance by employing a Remotely Piloted Aircraft System (RPAS) equipped with a thermal camera and low-cost GNSS RTK receiver, which performs a mission flying over a photovoltaic field and collecting optical and thermal images. The collected data processed by means of a computer vision algorithm and referenced by means of on-board L1-only RTK GNSS receiver would allow the maintainer to identify and locate the defective cell to be replaced without intermediate human inspections. The remainder of this paper is organized as follows. Section II describes the proposed system architecture concept, system design, core algorithm, and subsystem error sources. Section III focuses on GNSS centimeter positioning accuracy requirement and potential solution to meet the positioning accuracy requirement. Section IV presents preliminary results and analysis that have been carried out so far, followed by conclusion in Section V.

## II. PROPOSED SOLUTION

The proposed solution aims at automating the maintenance of PV plant with enhanced reliability in a time and cost effective manner from an operational perspective. This purpose

shall be achieved by enhancing the overall safety during the on-field inspection process using a custom RPAS payload design for PV plants maintenance to avoid the employment of people checking modules on rooftop plants (running risks for their safety) or in ground plants (sometimes crossing rough terrain and tall grass).

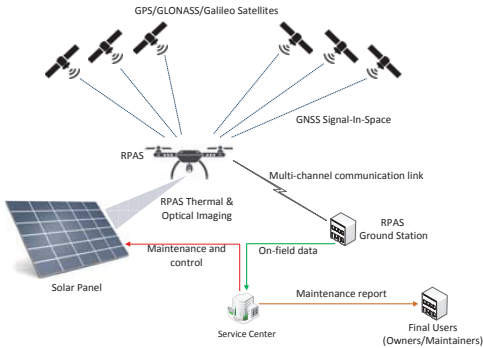


Fig. 1. Proposed architecture concept.

### A. System Architecture Concept

A depiction of proposed system architecture concept is presented in Fig.1. As shown, the proposed solution employs RPAS equipped with a thermal camera, low-cost GNSS RTK receiver and other related components. The RPAS shall perform a mission flying over a photovoltaic field to collect optical and thermal images of solar panels in a PV plant. The data collected data by RPAS shall be processed by means of a computer vision algorithm, and the captured images shall be tagged by means of centimeter-level position fix provided by L1 RTK GNSS receiver. The accurate positioning provided by GNSS enables the automation of the entire process allowing to correctly geo-reference the defective panel inspected by the on-board thermal camera. Finally, this information shall be provided to the remote service center in charge of thermal anomalies identification and management.

### B. Subsystem Design Overview

The proposed system is composed of three subsystems:

- 1) Service center
- 2) RPAS
- 3) RPAS Ground Segment (RGS)

1) *Service Center*: Service center is devoted to manage all aspects related to production, delivery, archiving and cataloguing of data acquired and processed. It performs:

- Product Archiving and Cataloguing of mission data such as Optical Images, Thermal Images, Videos and post processed products. It includes retrieval of previously archived and cataloged data/products and catalogue browsing.
- Mission Planning to arrange organization and cooperation between the various entities involved throughout the mission process.
- Post Processing Management of acquired/catalogued products. These activities can be operated in automatic mode or manually by Thermographic Expert Operator (TEO).
- Product Validation before delivery is a function mainly used by the TEO to see the products, processing and making analysis, and write annotations, etc.
- Product Delivery and Distribution to the final user of validated products, including report formatting, archiving and distribution.
- Monitoring and Control including all the activities involved in the management and presentation of information coming from the PV plant such as monitoring and control of plant elements status, alarm supervision and production parameters supervision.

2) *RPAS*: A block diagram of RPAS subsystem is provided in Fig. 2. As shown, RPAS is composed of COTS (Commercially Of the Shelf) or self-built drone platform, Communication and Control (C&C) unit, and the RPAS Payload. The RPAS Payload is consist of the following units:

- On-Board Computer (OBC) responsible for interfacing and handling the payload sensors (Barometer, GNSS Receiver, IMU, Thermal and Optical Camera), storing thermal and visual video/image, geo-Tagging, and enriching each frame with ancillary information such as GNSS horizontal positioning fix, attitude and related metadata.
- GNSS Receiver equipped with active patch antenna, which acts as a rover, is a dual-constellation low-cost L1-only RTK receiver supporting high position update rate (1-8 Hz) .
- 2 or 3 axis mechanical gimbal with its dedicated electronic control unit, which is designed for housing thermal sensor and optionally visual sensors with a custom design.
- Thermal and visual sensors implemented by commercial sensors (COTS) available in the market, such as, FLIR VUE PRO or FLIR TAU2.

3) *RPAS Ground Segment*: The RGS subsystem is composed of the following:

- Central Processing Unit (CPU) with dedicated Graphical Processing Unit (GPU) to perform demanding graphical operations. As a first choice, Jetson Tx1 (256 GPU Cuda cores) development board based on Nvidia Tegra processor has been selected. The software modules within CPU are:
  - + Computer vision algorithm performs various operations (offline) on the data collected by RPAS, which include

detection of thermal anomalies and identification of solar panels with unique ID based on its geographic placement using precise GNSS position fix and related metadata. The computer vision algorithm can be implemented over a distributed architecture by porting some of the recognition capabilities on-board RPAS.

- + Pilot Video Feedback acts as a real time video for the RPAS pilot, which enables the pilot to have a real time feedback of panels irradiation and an estimation of the area covered directly during on-field operations.
- + Service Center gateway is responsible for the generation of all the application and transport layers towards service center, according to the defined protocol (e.g. FTP / XML).
- GNSS receiver module at RGS is a low-cost L1-only RTK receiver that performs as a RTK base station providing RTK correction in RTCM (v.3.x) format to the RPAS through a UHF radio link or internet modem.
- Radio Modem is used for receiving the real time ancillary data (AUX channel - NAVCOM data, attitude, etc.) that will be used by the pilot for various operational purposes.

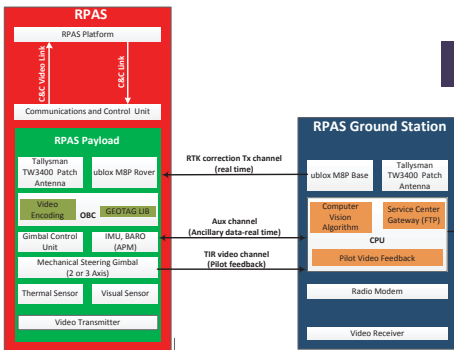


Fig. 2. RPAS and RPAS ground subsystem overview.

C. Core Algorithm

The core algorithm merges centimeter-level RTK GNSS receiver positioning capabilities with computer vision and geodesics algorithms largely used in the literature [1] [2] [3] [4] to identify and locate the faulty solar panel modules. The algorithm performs the analysis of acquired PV panels images to detect possible thermal anomalies and subsequently uses direct geo-referencing of images acquired through infrared (IR) camera to identify the geographical location of defective modules. Finally, a database (DB) is created by assigning each panel a unique identifier based on its precise geographical placement. The key steps performed during the process are summarized as below:

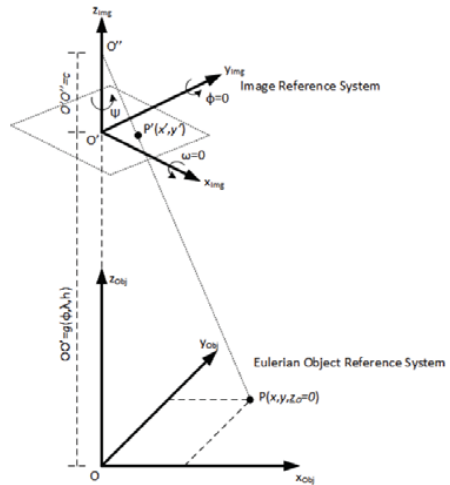


Fig. 3. Geometric illustration of Eulerian object reference system and image reference system.

1. Image Acquisition

The image acquisition shall be performed by implementing *radial acquisition photogrammetric technique*, as described in [5] [6]. The image is acquired by IR camera when the optical axis of camera is perpendicular to the plane formed by  $X_{obj}$  and  $Y_{obj}$ . The plane formed by  $X_{obj}$  and  $Y_{obj}$  is referred to as *Eulerian object reference system*, which a right oriented East, North, Up (ENU) originated in an arbitrary known point  $O(x_0, y_0, z_0)$  as illustrated in Fig. 3. In order to keep the optical axis of the camera normal to the *object space*, on-board gimbal adjustment shall be performed using the information obtained from on-board sensors.

2. Direct Geo-referencing

In order to geo-reference every point  $P_{(x,y,z_0=0)}$  acquired with the IR camera in the *object space* direct geo-referencing image algorithms shall be implemented [7]. Considering that the object is placed in the planer surface (2-dimensional) and the optical axis of the camera is perpendicular to the *object space*, the geo-referencing of point  $P_{(x,y,z_0=0)}$  is the same as knowing  $(x, y)$  scalar parameters of the image. For this purpose, a reference system, which is referred to as the *image reference system* is defined as shown in Fig. 3. The image reference system is a right oriented reference system originated in  $O'$  while having  $z_{img}$  axis parallel to  $z_{obj}$  and  $x_{img}$  and  $y_{img}$  forms a plane parallel to  $x_{obj}$  and  $y_{obj}$  axis. Furthermore  $x_{img} \cdot x_{obj} = \cos(\Psi)$  and  $P_{(x,y,z_0=0)}$  becomes  $P'_{(x',y')}$  *image space*. The aforementioned process of geo-referencing is performed by a transformation function  $\mathcal{F}(\cdot)$

given in Eq. 1

$$P_{(x,y,z_0=0)}^k = \mathcal{F}(P_{(x',y')}^k, c, \Psi, h, \phi_c, \lambda_c) \quad (1)$$

Where:

- $P_{(x,y,z_0=0)}^k$  is the point in *object space*.
- $P_{(x',y')}^k$  is point (in pixel) in the *image space* provided by the camera sensor.
- $c$  is the focal distance that depends upon the camera sensor.
- $h$  is the RPAS altitude obtained from on-board barometric sensor.
- $\phi_c$  is the longitude and  $\lambda_c$  is the latitude of on-board moving camera sensor.

It has to be remarked that  $\phi_c$  and  $\lambda_c$  are obtained through a transformation function from on-board GNSS Antenna Phase Center (APC) latitude  $\phi_{GNSS}$  and longitude  $\lambda_{GNSS}$ , which are accurate to centimeter level mainly due to RTK based positioning. This indicates the significance of centimeter level positioning in enabling the automation of PV plant maintenance

Eventually, the point  $P_{(x,y,z_0=0)}^k$  in the *object space* shall be transformed from object reference system (ENU) to World Geodetic System (WGS84) representation of geographical coordinate as:

$$P_{(\phi,\lambda)}^k = \mathcal{G}(P_{(x,y,z_0=0)}^k) \quad (2)$$

**3. Panel recognition and faulty objects detection** The process of panels recognition and its related faulty objects (e.g. hotspot, hot strips and other kind of thermal anomalies) is implemented with image processing [8]. For our application, given a video sequence containing one or more faulty moving objects, the desired result is the set of the trajectories of these objects and an estimation of their position.

#### 4. DB Cataloguing

The PV Panels can be uniquely catalogued by the generation of an alphanumeric primary key exploiting the position information. Each panel, surrounded by its bounding box, is tracked by the algorithm  $N$  times and the position of its center in the *object space* is estimated after averaging over  $N$  position estimates as shown in Eq.

$$P_{(\phi_m,\lambda_m)}^k = \sum_{i=1}^N \frac{P_{(\phi,\lambda)_i}^k}{N} \quad (3)$$

$P_{(\phi_m,\lambda_m)}^k$  represents the center of the  $k^{\text{th}}$  panel, which shall be used to generate a unique key to exploit the position information in order to access the DB.

#### D. Subsystem Error Sources

The proposed system involves various units within each subsystem such that each of the units introduces a certain

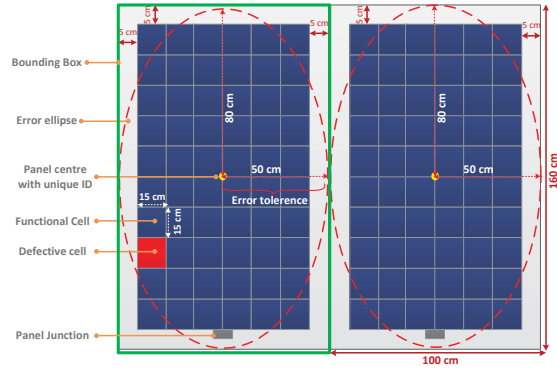


Fig. 4. Standard solar panel dimensions, error ellipse, and maximum error tolerance.

amount of error that contributes to the total subsystem error budget. Currently the following error sources contributing are identified:

- 1) Payload altimeter resolution: A barometric altimeter is used to assess the RPAS height  $h$  with a resolution of about 14 cm. Such an error impacts  $O'$  i.e. image reference system determination.
- 2) RPAS velocity: The RPAS Ground Speed of typically less than 3 m/s is taken into account with respect to the GNSS receiver position update
- 3) Gimbal: The orientation of gimbal must be normal to object reference system, however, wind gusts may introduce for few milliseconds misalignment to be evaluated by tests. Typical resolutions for COTS gimbal is about  $0.05^\circ$
- 4) PV module Height: The height of PV panels is assessed in case of lack of information about PV panel placement in planar surface. This error source causes  $P_{(x,y,z_0=0)}$  not belonging to the *object space*.
- 5) PV module inclination: Panels have inclination which is typically a function of the latitude for PV plants on the ground. Such inclination reduces the cross section of the Panels as seen from RPAS.
- 6) Sensor lens distortion: Distortion introduced by sensors optics, to be assessed and mitigated during tests. Sense lens distortion account one of the most significant source of error budget.
- 7) GNSS accuracy: Error in position solution obtain from on-board RTK GNSS rover receiver contributes to the error budget. Further tests need to be conducted to quantify GNSS accuracy impact.
- 8) Sensor resolution: The error introduced by the resolution of commercial thermal cameras of available resolution up to 640x480 pixels.

- 9) Algorithm quantization: The errors introduced by the algorithm in the image reference system taking into account the panels dimensions expressed in pixels.

A thorough assessment of the aforementioned error sources shall be done during the test and validation phase of the study in order to quantify the contribution of the error sources to the subsystem error budget.

### III. GNSS ACCURACY REQUIREMENT FOR RPAS POSITIONING

Considering the standard dimension of the solar panel the maximum error tolerance is  $\min(50 \text{ cm}, 80 \text{ cm})$  to uniquely identify each solar panel regardless the inclination of the solar panel as shown in Fig. 4 that . This implies that the subsystem error budget always be less than the maximum error tolerance of 50 cm. To meet subsystem error requirement, the accuracy of L1 RTK GNSS receiver is of paramount importance in order to restrict the GNSS positioning error contribution as minimum as possible keeping in view the low-cost, low-power and small form factor deployment constraints. Though, at this point, the exact positioning requirement is not known and shall be estimated during a dedicated validation and testing campaign. However, the expected GNSS positioning accuracy should be not more than 20 cm (2DRMS) to provide enough error margins for other error sources contributing to the subsystem error budget. It has to be remarked that the positioning accuracy requirement is subject to horizontal positioning only since the computer vision algorithms operate in two-dimensional space, therefore, no vertical positioning information is required.

Currently, a number of L1-only RTK OEM receivers are available in market, which offer centimeter positioning accuracy employing RTK messages in RTCM format. As listed in Table I, ublox M8P [9] offers the least expensive L1-only RTK solution with an added feature of integrated UHF radios for RTK RTCM message transmission/reception. The RTK RTCM messages can either be generated employing one of the OEM module as base station or it can be obtained from a public Continuously Operating Reference Station (CORS) such as EUREF permanent GNSS network [10] using Network Transport of RTCM via Internet Protocol (NTRIP) [11]. Typically, CORS is a desirable choice for using RTK in order to avoid base station setup complexity, however, the availability of CORS nearby PV plant is not always possible. For this reason, a local base station approach, which is supported by all the existing L1-only RTK OEM receiver, presents a handy choice at the cost of base station installation setup for on-field operations in a PV plant.

#### IV. UBLOX M8P HORIZONTAL POSITIONING ACCURACY

ublox M8P base and rover equipped with active GPS/GLONASS Tallysman antenna, as shown in Fig. 5, were installed on the rooftop of Aalborg University buildings at a spatial separation of 160 meter between base and rover as shown in Fig. 6. The rover position data in NMEA format was logged at rate of 1 Hz for a period of three days. The

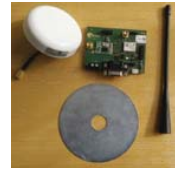


Fig. 5. ublox M8P, Tallysman TW3400 antenna, ground plane, and 433 MHz UHF antenna for RTCM transmission/reception



Fig. 6. ublox M8P base (left) and rover (right) mounted on rooftop of Aalborg University buildings. The spatial separation between the rooftop of base and rover is 160 m.

logged file is then processed by RTKPLOT utility [12] to obtain rover ground track and related positioning statistics. Fig. 7 shows plot of rover ground track considering position solution only with RTK fix. It can be observed that the rover provides 2DRMS positioning accuracy of 1.08 cm, while the RTK fix was available 99% of the time. Though the rover receiver was static, however, such positioning performance is significant for L1-only RTK receiver considering the fact that no surveyed position was provided to the base receiver. Nevertheless, we expect that the degradation in positioning accuracy is inevitable under dynamic condition of moving RPAS. We expect that positioning error should not exceed 20 cm in order to provide reasonable margin for other error sources. Unlike base, rover receiver benefits from improved signal reception as the GNSS signal experiences relatively less multipath and obstruction due to RPAS flight at a height.

### V. CONCLUSION

The target market of the proposed solution comprises of both maintainers and owners of large PV plants as final users, which are encouraged to use customized RPAS payload design together with RPAS ground segment to enhance PV plant energy production in a cost-effective way by automating PV plant maintenance.

As a first step towards a market-ready solution aligned with PV maintenance user needs, this paper presented system architecture and subsystem design at a glance as well as expected error sources contributing to the total error budget. Furthermore, the paper discussed the algorithm, which performs the key tasks of direct geo-referencing, image acquisition, panel recognition, faulty module detection, and solar panel cataloging in DB. Finally, low-cost L1-only RTK GNSS solutions are briefly discussed along with expected error

TABLE I  
MARKET READY L1-ONLY GNSS RTK OEM EVALUATION BOARDS

Product Name	GNSS	Features	Est. cost (€)
ublox M8P	GPS/GLONASS	Base-Rover/CORS, UHF & GNSS Antenna	340
Piksi	GPS	Base-Rover, UHF & GNSS Antenna	1000
Navspark S2525F8-BD-RTK	GPS/BeiDou	Base-Rover/CORS	450
NV08C-EVK-RTK	GPS/GLONASS	Base-Rover/CORS	1600
Reach RTK	GPS/GLONASS	Base-Rover/CORS	500

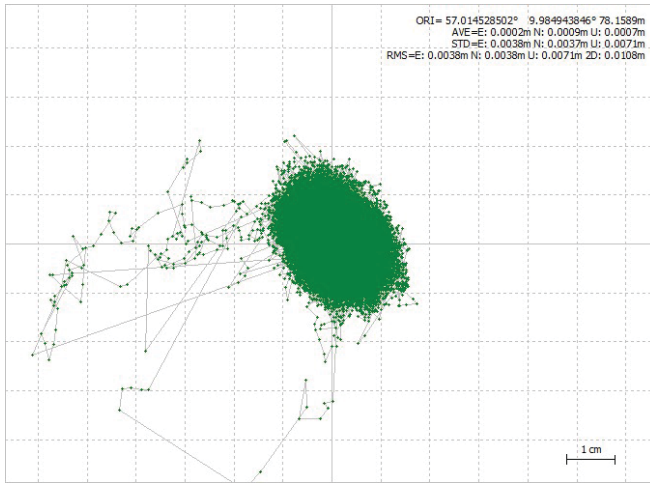


Fig. 7. ublox M8P rover (static) horizontal positioning accuracy, 2DRMS = 1.08 cm, and RTK Fix availability of 99%.

contribution to the overall subsystem error budget. The paper presented static horizontal positioning accuracy of ublox M8P, which indicated that M8P is a potential candidate RTK GNSS receiver to be integrated on-board RPAS. In the upcoming phase of the project, a thorough test and validation campaign shall be performed in order to realize the proposed concept and core algorithm taking in to account on-field PV plant conditions.

ACKNOWLEDGMENT

The work presented in this paper is a part of the ongoing activities performed in the framework of EASY-PV project sanctioned under H2020-Galileo-2015-1 call for Small and Medium Enterprise (SME) based EGNSS application having grant agreement number 687409. The project started off on Feb. 01, 2016 and shall end on Jan. 31, 2018. The project is coordinated by Sistematica S.p.A with contributions from other partners, namely, Aalborg University, TopView srl, DeepBlue, Entec, and Alpha consultants.

REFERENCES

[1] L. OpenCV, "Computer vision with the opencv library," GaryBradski & Adrian Kaebler-O'Reilly, 2008.

[2] G. Cai, B. M. Chen, and T. H. Lee, *Unmanned Rotorcraft Systems*. Springer Science & Business Media, 2011.

[3] T. Vincenty, "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations," *Survey review*, vol. 23, no. 176, pp. 88-93, 1975.

[4] C. F. Karney, "Algorithms for Geodesics," *Journal of Geodesy*, vol. 87, no. 1, pp. 43-55, 2013.

[5] K. Kraus, *Photogrammetry*. Ferdinand Dummlers Verlag, 1993, vol. 1.

[6] —, *Photogrammetry*. Ferdinand Dummlers Verlag, 1997, vol. 2.

[7] M. Cramer, "Performance of gps/inertial solutions in photogrammetry," 2001.

[8] H. P. VC, "Method and means for recognizing complex patterns," Dec. 18 1962, US Patent 3,069,654.

[9] C. Mongrdien, J.-P. Doyen, M. Strom, and D. Ammann, "Centimeter-Level Positioning for UAVs and Other Mass-Market Applications," in *Proceedings of the 29th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ )*, Portland, Oregon, 2016.

[10] C. Bruyninx, "The euref permanent network: a multi-disciplinary network serving surveyors as well as scientists," *GeoInformatics*, vol. 7, no. 5, pp. 32-35, 2004.

[11] G. Weber, D. Dettmering, and H. Gebhard, "Networked transport of rtm via internet protocol (ntrip)," in *A Window on the Future of Geodesy*. Springer, 2005, pp. 60-64.

[12] T. Takasu, N. Kubo, and A. Yasuda, "Development, Evaluation and Application of RTKLIB: A program library for RTK-GPS," in *GPS/GNSS symposium*, 2007, pp. 213-218.

# Challenges in self-coordinating and adaptive CPS with human in the loop

Agata Manolova<sup>1</sup>, Vladimir Poulkov<sup>2</sup>, Krasimir Tonchev<sup>3</sup>, Ognian Boumbarov<sup>4</sup>

Faculty of Telecommunications

Technical University of Sofia

Sofia, Bulgaria

Email: { amanolova<sup>1</sup>, vkp<sup>2</sup>, k\_tonchev<sup>3</sup>, olb<sup>4</sup> }@tu-sofia.bg

**Abstract**— New development in information and communication technologies created the way for ubiquitous Cyber-Physical Systems (CPS) implanted with self-coordination and self-adaptive capabilities. CPSs are mainly related with the control and monitoring of physical environments and phenomena through sensing and actuation systems consisting of distributed computing and communicating devices. CPS has seen a great expansion with enormous societal and economic impact facilitating various services such as medical systems, assisted living, traffic control and safety, advanced automotive systems, process control, energy conservation, distributed robotics, weapons systems, manufacturing, distributed sensing command and control, critical infrastructure, smart structures, bio-systems, and communications systems. In order for these systems to better serve the human needs, the CPSs will need to acknowledge the influence of the user or the operator, through Human-in-the-Loop controls that take into consideration human intents, psychological states, emotions and actions. As these completely incompatible worlds have to be integrated, innovative solutions are required. The aim of this paper is to provide an overview of the current challenges in designing a self-coordinating and adaptive CPS with human in the loop. In this paper we focus on the multi-disciplinary challenges taking into account the perspectives of the physical processes related to the human behavior, computation and integration in CPS.

**Keywords**—Cyber physical systems; Human in the loop; Self-coordination; Self-adaptation

## 1. INTRODUCTION

Computing and communication capabilities will soon be embedded in all types of everyday objects in the physical environment that surround us. Applications with enormous societal and economic impact will be created by harnessing the capabilities of these web-connected devices for good or bad. Systems that “bridge the cyber-world of computing and communications with the physical world are referred to as Cyber-Physical Systems (CPS)” [1]. CPSs are mainly related with the control and monitoring of physical environments and phenomena through sensing and actuation systems consisting of distributed computing and communicating devices [2]. CPS are systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core. Due to these facts in the last few years they have attracted significant interest and are being considered to be a frontier

scientific research field, which will be of major interest for the years to come as it was mentioned in a report of the European Commission [3]. According to this report a key aspect that is currently missing is the consideration of the socio space, cyber space and physical space interacting at the same time.

The core philosophies of CPS and the Internet of Things (IoT) are very similar in the field of intensive information processing, comprehensive intelligent services and efficient communications [4]. But CPS are interconnected systems of collaborating heterogeneous units and components that are envisioned to provide integration of communication, computation and also physical processes. The potential of CPS is boosted by several recent trends in wireless communications; low cost, low-power, high-capacity, small form-factor computing devices and increased-capability sensors, continuing improvements in energy distribution, alternative energy sources and energy harvesting [5]. CPS can bring together the discrete and powerful logic of computing and communications, to monitor and control the continuous dynamics of physical and engineered systems. As already described in [6] CPS must cope with the complexity in the physical environment, together with the lack of perfect synchrony across time and space.

CPSs operate in dynamic contexts and thus have to handle uncertainty that results from phenomena such as interference and noise, abnormal behavior, rare events, evolving structure, etc. This uncertainty takes another dimension when humans are included in the loop of the CPS. In addition CPSs have to cope with noisy and heterogeneous data [7], to offer robust performance over often unreliable wireless and open communication networks, to operate safely, securely, and efficiently and all this in real-time as they interact with the physical world.

All these different properties and characteristics of CPS enable new opportunities, pose new research challenges and call for the creation of innovative scientific foundations and engineering principles for CPS. New approaches based on statistical analysis and mathematics must replace inefficient and testing-intensive techniques. The coupling between the cyber and physical contexts will be driven by new demands and applications. Novel interactions among communications, computing and control must be analyzed and understood, and based on this new methods and algorithms that explicitly address the interaction between the physical and cyber

This work was supported in part by the Grant Agreement No: 610658, eWALL: eWall for Active Long Living” of the EU Seventh Framework Programme. The authors wish to thank the invaluable help received from all the consortium members.

subsystems must be developed. They must be based on the integration of the theories of communication systems and computing, sensing and control of physical systems and the interaction between humans and CPS. These systems are mostly interdisciplinary, requiring expertise and skills in communication, and data processing, smart devices and services, security and privacy, mathematical abstractions (algorithms, processes, etc.).

CPS will transform how humans interact with and control the physical and cyber worlds. The application domains of CPSs involve [8], but are not limited to medical systems, assisted living, traffic control and safety, advanced automotive systems, process control, energy conservation, distributed robotics, weapons systems, manufacturing, distributed sensing command and control, critical infrastructure, smart structures, bio-systems, and communications systems. Practical examples of CPS nowadays include different types of medical devices, aerospace systems, intelligent transportation vehicles, defense systems, robotic systems, factory automation, building and environmental control and smart homes and smart cities. Some of the many societal benefits that CPS will deliver will be in the domain of zero-energy buildings and cities, extreme-yield agriculture, near-zero automotive fatalities, perpetual life assistants, location-independent access to medical care, situation-aware physical critical infrastructure, blackout-free electricity, and safe evacuation from hazardous areas.

The main contribution of this paper is to provide an overview of the current challenges in designing a self-coordinating and adaptive CPS with human in the loop. We will focus on the multi-disciplinary challenges taking into account the perspectives of the physical processes related to the human behavior, computation and integration in CPS. The outcome of the survey will provide knowledge about how some of these challenges are currently handled in different example CPS, what problems have been tackled, which methods have been used to solve them, and how solutions have been evaluated. These insights will help identify areas for further investigation in this particular scientific area and outline concrete challenges for future research concerning the self-adaptation and coordination in CPS with human in the loop (CPS with HiL).

The rest of the paper is organized as follows: the next section describes in detail the main challenges for CPS with HiL. Section III introduces some opportunities and solutions to the presented challenges. The final section draws the conclusion and suggests the scope of future work.

II. CHALLENGES

A conceptual model of CPS with HiL is illustrated on Fig. 1. CPS are typically designed as networks of interacting elements with physical input and output instead of being simply a combination of standalone devices. The inherent heterogeneity and integration of different components pose new challenges to traditional data analysis, communication, control, and software theories [9]. This often makes system design analysis and modeling inefficient with current technologies.

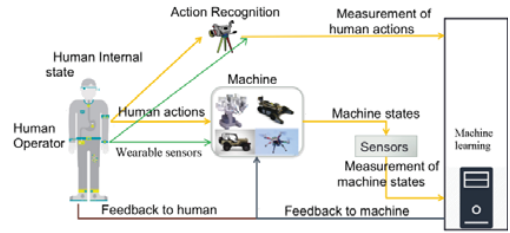


Fig. 1. Conceptual model of a CPS with Human in the Loop

A major difference between CPS and a typical control system or an embedded system is the use of communications, which adds re-configurability and scalability, as well as complexity and potential instability. Furthermore, CPS has intelligent sensors and actuators, as well as substantially stricter performance and energy constraints, which are critical for its efficiency and lifetime. On the other hand, cyber capabilities are embedded in every physical process and component, networking is employed at multiple scales, complexity lies at multiple temporal and spatial scales, and high heterogeneity is seen across devices and protocols [10]. Timing and spatial precision, uninterrupted connectivity, predictability and repeatability are extremely critical for CPS [11]. As very well explained in [12] the scientific community needs to build new theoretical foundations, scientific models, abstractions and to rethink and reinvent user oriented interface and connectivity functions.

A. Self-coordination and adaptation

In many CPS research initiatives focus is on self-monitoring, self-diagnosis and adaptation to maintain both operability and safety, while also taking into account humans-in-the loop for system operation and decision making [13], [14], [15], [16]. Typical goals of such self-diagnosis approaches are the detection and isolation of faults, identifying and analyzing effects of degradation, providing fault adaptive control, and optimizing energy consumption [17]. So far, the majority of projects and papers for analysis and diagnosis rely on manually-created diagnosis models of the system's physics and operations [18], [19]. However, the last years have clearly shown that such models are rarely available for complex CPSs and when they do exist, they are often incomplete and sometimes inaccurate, and it is hard to maintain the effectiveness of these models during a system's life-cycle. A promising alternative is the use of data-driven approaches, where monitoring and diagnosis knowledge can be learned by observing and analyzing system behavior. Such approaches have only recently become possible as CPSs now collect and communicate large amounts of data [20]. This large amount of data can be exploited for the purpose of detecting and analyzing performance. The vision is developing CPSs that can observe their own behavior, recognize unusual situations during operations, inform experts and/or control systems, which can then update operations procedures, and also inform operators, who use this information to modify operations or



plan for repair and maintenance. The CPS need to adapt and remain operable even in restrictive and hostile conditions, such as network unavailability, hardware failures, resource scarcity, etc. The biggest problem is that an exhaustive catalog of configurations and situations at design time is not a viable solution in the domain of CPS, to model all unanticipated situations in the physical environment. This is followed by the fact that self-adaptive CPS need to base their adaptation actions not only on the current situation, but also to learn from previous situations and improve its performance [21]. Muccini et al. [22] did a very extensive review on self-adaptation in CPS and their main conclusion was that the primary concerns of adaptation in CPS are performance, flexibility, and reliability. The authors' findings show that adaptation in CPS is a cross-layer concern of the technology stack and the main future challenges are how to map concerns to layers and adaptation mechanisms, how to coordinate adaptation mechanisms within and across layers, and how to ensure system-wide consistency of adaptation. In [23] the author discusses in detail the aspect of self-coordination in CPS. Ramming argues that to bridge the gap between the embedded systems with their deterministic behavior and the globally available data and services with their highly probabilistic behavior a certain degree of self-organization together with cognitive capabilities of CPS is crucial.

### B. Human in the Loop

Coverage and connectivity should be redefined in the framework of CPS with human in the loop (HiL). Such systems will usually consist of both wired and wireless sensor and actuation networks with different capacities and reliability, but the human presence in the loop will add another dimension of uncertainty. To model such heterogeneous system, an innovative scientific and technological turn is necessary. To compromise all those critical aspects of CPS with HiL, sophisticated signal and data processing techniques, estimation and analysis of human influence, coupled with novel network, connectivity and security approaches, should be researched and designed to provide highest performance and efficiency levels for such CPS.

In CPS the human context will become increasingly more important as most future technologies will be human-aware. Future CPSs will most likely bolster a much stronger tie between humans and control loops. A CPS with Human-in-the-loop is a system that must take human response into consideration and human presence and behavior becomes a key part of the system instead. In this case it is essential to develop and integrate reliable and accurate human behavior modeling techniques that attempt to learn and predict human behavior. Capturing human behavior by extending system identification or other modeling techniques is extremely difficult due to complex physiological, psychological and behavioral aspects of human beings, some examples are presented in [24], [25]. Also, the level of modeling depends on application requirements. Although requirements are different for different applications, a significant portion of human-in-the-loop applications have to address some common challenges, e.g., user specific metrics, thresholds and parameters, estimation of the change of human behavior over

time, and required sensing technology to sense the appropriate aspects of human behavior. Human behavior must be modelled for large number of applications before general principles and theories emerge to address these issues. Clustering, data mining, specialized models based on human physiology and behaviors may all be techniques to be enhanced before applied in the development of CPS with HiL. Robust CPS systems will likely require predictive models to avoid problems before they occur; consequently advances to stochastic model predictive control are also required. It is also unlikely that any models developed initially to design the controllers will remain accurate as the system and human behaviors evolve over time. Hence, adaptive control with humans-in-the-loop will be necessary.

Currently, the techniques that model certain aspect of human behavior are either very general or very specific. For example, Smart Thermostat [26] uses a Hidden Markov Model (HMM) to model occupancy and sleep patterns of the residents in a home to save energy, which captures human behavior from a very high level. Another example, describes a new paradigm called Body Coupled communication (BCC) for Wireless Body Area Network (WBAN) that leverages the human body as a communication channel [27]. In this application, sensors are implanted in a human body that are capable of monitoring a wide range of physiological and emotional states and human serves as a communication channel to transmit the sampled data to a centralized monitoring entity. Another challenge is related to determining how to incorporate human behavior models into the formal methodology of description of CPS. Even if we have a model of human behavior, it is not clear where to place the model for "each" specific application. The newest challenge seems to be how to incorporate the human behavior as part of the system itself.

### C. Predictability and scheduling

Since computing and networking systems in CPS interact with the physical world, predictability (or timeliness) of these systems is an important property that should be provided. Real-time scheduling theory is the area that studies this issue in computing and networking systems. Some examples of real-time scheduling theory are feasible region calculus [28] or resource sharing, real-time queueing theory [29], etc. In CPS, temporal predictability of communication systems is also an important issue for overall system-wide stability, performance, and safety. Achieving such real-time properties is much harder in wireless networking systems than in wired situations due to several issues such as interference between nodes, dynamically changing network topology, power, etc. Practical realization of wireless sensor networks (WSN), which are considered the enablers of CPS architecture, is characterized by periods of severe impairments due to propagation phenomena, noise, and interference, dictated by the characteristics of the communication channel [30] and they cannot be treated with conditional methods, such as increased transmission power [31]. These limitations have a direct impact on the performance and reliability of the connectivity of the CPS, which are typically expressed in terms of packet losses, and excessive delays. In such cases standard

scheduling or routing approaches such as packet scheduling cannot be utilized. The problem of real-time packet scheduling is that schedulability condition is derived by considering the unreliability of wireless communication links [32], [33]. Some of our research group is focused in this research area [34], [35]. Wireless sensor networks involve several of the components related to CPS, such as sensing, computing, and communication. There are many important research issues of importance to WSN, such as connectivity, energy-efficient networking protocols, routing and platforms which can be refined, reshaped and considered for use in CPS [36], [37], [38], [39]. In addition the great advances in radio technologies with the design of software radios and cognitive radios, which offer better flexibility, could be applied for improving connectivity and ensuring better quality of service in CPS. For example, it can be conceived that some frequency bands can be allocated for critical cyber-physical traffic using robust modulation schemes.

#### D. Security

New security challenges are introduced by CPS, especially with human in the loop, challenges that cannot be fully handled via traditional cryptographic means. In many wireless CPS the devices have limited capabilities and are not controlled by a central control center; thus, the implementation of computationally expensive cryptographic techniques can be challenging. One can find many examples where "smart" home devices are used as weapons in website attacks [40]. In this article one of the contributors explains that: "In a relatively short time we've taken a system built to resist destruction by nuclear weapons and made it vulnerable to toasters."

Motivated by these considerations, substantial recent research has been investigating the use of the physical layer as a means to develop low-complexity and effective wireless security mechanisms. Physical layer security has become an emerging hot topic in wireless systems. At the heart of this enthusiasm is the belief that the physical layer represents a previously untapped resource for enhancing wireless security [41]. In particular, rather than rely solely upon generic higher-layer cryptographic mechanisms, as has been the norm, there is a belief that it is possible to design lower-layer services that support security objectives such as authentication and confidentiality. There are several good surveys and collections that explore the fundamentals of physical layer security [42], [43], [44].

### III. SOME OPPORTUNITIES

#### A. Cyber-physical transportation systems (CPTS)

CPTS are rapidly advancing due to progress in real-time computing, control and artificial intelligence [45]. All contemporary cars are sold with a good amount of smart components. Such components include engine control sensors, comfort and safety features. In order to meet environmental standards, ICT technologies are required in every aspect of the vehicle design. There is also huge progress in autonomous driving [46] where the human centric element is a primary concern. A human-centric vehicular system uses smart sensors

to gather information, make fast decisions and adapt its behavior, leaving the human free to make higher level decisions such as preference for the fastest route or fuel consumption. In recent years, various threats, vulnerabilities, and attacks have been discovered from different models of smart cars [47] and many solutions are proposed.

#### B. Medical CPS

Today, many components of the health system are combined into networked closed-loop systems with humans in the loop to improve medical workflows and patient safety. Emerging technologies aimed at enabling remote care of patients will provide care practitioners with information on how activities of daily living affect healthcare and allow practitioners to make more informed decisions about interventions, rehabilitation and in general the well-being of the patient. One such example can be considered: eWALL home caring environment [48]. The architecture is described in [49]. eWALL is a platform providing dynamic environment for elderly patients with MCI for social interaction and continuous medical surveillance. The system offers personalized and adaptive services such as daily activity monitoring, suitable exercises, reminders and others.

In [50] an extensive review is done concerning medical CPS such as Implantable Medical Devices, Physiological Control Systems and Model-based Clinical Trials. The conclusion of the authors is that the healthcare system will transform into one large and complex and safety-critical cyber-physical system with many advantages as well as challenges.

Very interesting example of such system for improvement of emotional well-being is published in [51]. This work presents the implementation of a Human-in-the-loop emotionally-aware Cyber-Physical System that attempts to positively impact its user's mood by creating emotionally-aware application to benefit people's daily lives.

#### C. Cyber Physical Energy Systems (CPES)

Energy production, distribution, and optimization are all CPS problems. For example, the smart grid combines multiple electric power production plants with a multiplicity of loads using dynamic load balancing and dynamic pricing with demand-response strategies. Smart buildings integrate sensors into control systems for lighting, heating, ventilation, and air conditioning, safety (fire monitoring and evacuation) and home entertainment. The main CPES research areas are identified as: modelling energy systems, energy efficiency, energy resource management, and energy control. In [52] and [53] are mentioned some examples of such CPES acknowledging the issues of security and the influence of the human user in order to provide reliable energy supply and maximize user comfort while minimizing energy usage.

### IV. CONCLUSION

The research challenges related to CPS with HiL come from the fact that CPS connect completely incompatible worlds: computing and communications with the physical world and integrating the human behavior in the loop. For bridging the above mentioned gap towards really functional

CPS with HiL a certain degree of self-organization together with cognitive capabilities are crucial. These engineered systems are built from, and depend upon, the seamless integration of computational algorithms and physical components. Advances in CPS will enable capability, adaptability, scalability, resiliency, safety, security, and usability that will far exceed the simple embedded systems of today.

The overall impact of the CPS on the human society is currently largely unknown, since social sciences can hardly cope with the speed of introduction of new technologies. But one important question arises about CPS with HiL: Whose Responsibility: Human vs. CPS? How much do users merge with CPS? Are they partially integrated in the existing system or do they fully depend on the advanced conveniences offered by the CPS? For example, when an emergency happens on a freeway, would systems release the control of automobiles to drivers or not? If not, would drivers allow their lives to be in the sole control of machines, argue the authors in [54] and [55]. And in the case of some damage, who or what is liable for the damage. The lack of answers to these legal liability issues could actually prevent useful CPS technology from being implemented.

#### REFERENCES

- [1] Ragunathan Rajkumar, Insup Lee, Lui Sha, John Stankovic, Cyber-physical systems: The next computing revolution, Design Automation Conference (DAC), 2010 47th ACM/IEEE, 13-18 June 2010.
- [2] Lee, E.A.: Computing foundations and practice for cyber-physical systems: a preliminary report. Technical Report UCB/EECS-2007-72, EECS Department, University of California, Berkeley (2007).
- [3] Cyber-Physical Systems: Uplifting Europe's Innovation Capacity, Report from the Workshop on Cyber-Physical Systems: Uplifting Europe's Innovation Capacity, 29th and 30th October 2013, Brussels, Belgium.
- [4] J. Wan, M. Chen, F. Xia, L. Di, and K. Zhou, "From machine-to-machine communications towards cyber-physical systems," Computer Science and Information Systems, vol. 10, pp. 1105–1128, 2013.
- [5] Nafaâ Jabeur, Nabil Sahli, Sherali Zeadally, Enabling Cyber Physical Systems with Wireless Sensor Networking Technologies, Multiagent System Paradigm, and Natural Ecosystems, Journal of Mobile Information Systems, Volume 2015 (2015), Article ID 908315.
- [6] Edward A. Lee, Albert M. K. Cheng, The Past, Present and Future of Cyber-Physical Systems: A Focus on Models, Sensors (Basel). 2015 Mar; 15(3): 4837–4869.
- [7] Akshay Rajhans; Ajinkya Bhavne; Ivan Ruchkin; Bruce H. Krogh; David Garlan; André Platzer; Bradley Schmerl, Supporting Heterogeneity in Cyber-Physical Systems Architectures, IEEE Transactions on Automatic Control ( Volume: 59, Issue: 12, Dec. 2014 ), pp. 3178 – 3193
- [8] Eva Geisberger, Manfred Broy, Living in a networked world: Integrated research agenda Cyber-Physical Systems (agendaCPS), *Herbert Utz Verlag*, 2015.
- [9] G. Tzagakarakis, G. Tsagakatakis, D. Alonso, C. Asensio, E. Celada, A. Panousopoulou, P. Tsakalides, and B. Beferull Lozano, Signal and Data Processing Techniques for Industrial Cyber-Physical Systems, *Cyber-Physical Systems From Theory to Practice*, CRC Press 2015, p. 181–226
- [10] X. Cao, P. Cheng, J. Chen, and Y. Sun, "An online optimization approach for control and communication codesign in networked cyber-physical systems," Industrial Informatics, IEEE Transactions on, vol. 9, no. 1, pp. 439–450, Feb 2013
- [11] K.-D. Kim and P. Kumar, "Cyber physical systems: A perspective at the centennial," Proceedings of the IEEE, vol. 100, no.Special Centennial Issue, pp. 1287–1308, May 2012.
- [12] J.R. Poovendran, Cyber-physical systems: close encounters between two parallel worlds [point of view], Proc. IEEE 98 (8) (2010) 1363–1366.
- [13] E.A. Lee. Cyber physical systems: Design challenges. In Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on, pages 363–369, 2008.
- [14] Ragunathan (Raj) Rajkumar, Insup Lee, Lui Sha, and John Stankovic. Cyber-physical systems: The next computing revolution. In Proceedings of the 47th Design Automation Conference, DAC '10, pages 731–736, New York, NY, USA, 2010. ACM.
- [15] Peter C. Evans and Marco Annunziata. Industrial internet: Pushing the boundaries of minds and machines. Technical report, GE, 2012.
- [16] Promotorengruppe Kommunikation. Im fokus: Das industrieprojekt industrie 4.0, handlungsempfehlungen zur umsetzung. Forschungsunion Wirtschaft-Wissenschaft, March 2013.
- [17] L. Christiansen, A. Fay, B. Ogenoorth, and J. Neidig. Improved diagnosis by combining structural and process knowledge. In Emerging Technologies Factory Automation (ETFA), 2011 IEEE 16th Conference on, Sept 2011.
- [18] Johan de Kleer, Bill Janssen, Daniel G. Bobrow, Tolga Kurtoglu Bhaskar Saha, Nicholas R. Moore, and Saravan Sutharshana. Fault augmented modelica models. The 24th International Workshop on Principles of Diagnosis, pages 71–78, 2013.
- [19] D. Klar, M. Huhn, and J. Gruhsr. Symptom propagation and transformation analysis: A pragmatic model for system-level diagnosis of large automation systems. In Emerging Technologies Factory Automation (ETFA), 2011 IEEE 16th Conference on, pages 1–9, Sept 2011.
- [20] GE. The rise of big data - leveraging large time series data sets to drive innovation, competitiveness and growth - capitalizing on the big data opportunity. Technical report, General Electric Intelligent Platforms, 2012.
- [21] Ilias Gerostathopoulos, Tomas Bures, Petr Hnetynka, Jaroslav Keznikl, Michal Kit, Frantisek Plasil, Noel Plouzeau, Self-adaptation in software-intensive cyber-physical systems: From system goals to architecture configuration, The Journal of Systems & Software (2016), doi: 10.1016/j.jss.2016.02.028.
- [22] Henry Muccini, Mohammad Sharaf, Danny Weys, Self-adaptation for cyber-physical systems: a systematic literature review, SEAMS '16 Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems 2016, pp. 75–81.
- [23] Franz J. Rammig, Self-Coordination as Fundamental Concept for Cyber, 2nd SBC Workshop on Autonomic Distributed Systems - WoSIDA 2012, Ouro Preto, Brazil, S. 45–48, Apr. 2012.
- [24] Krasimir Tonchev, Stanislav Panev, Agata Manolova, Nikolay Neshov, Ognian Bumarov, Vladimir Poulkov." Gaze Tracking, Facial Orientation Determination, Face and Emotion Recognition in 3D Space for Neurorehabilitation" in Neuro-Rehabilitation with Brain Interface. L.G. Lighthart, R. Prasad, S. Pupolin Ed. Aalborg, Denmark: River Publishers, 2015, ch. 4, pp. 51–87, ISBN: 9788793237438.
- [25] Krasimir Tonchev, Strahil Sokolov, Yuliyana Velchev, Georgy Balabanov, Vladimir Poulkov. Recognition of Human Daily activities. 2015 Proceedings of the IEEE International Conference on Communications: Workshop on ICT-enabled services and technologies for eHealth and Ambient Assisted Living. June 2015, London, United Kingdom. 2015 IEEE, pp. 284–289.
- [26] Lu, J., Sookoor, T. et al. The smart thermostat: using occupancy sensors to save energy in homes. In *SenSys* (2010).
- [27] Schirner, G., et al. The future of human-in-the-loop cyber-physical systems. *Computer* 46, 1 (2013), 36–45.
- [28] T. Abdelzaher, V. Sharma, and C. Lu, "A utilization bound for aperiodic tasks and priority driven scheduling," IEEE Transactions on Computers, vol. 53, no. 3, 2004.
- [29] L. Sha, T. Abdelzaher, et al. "Real time scheduling theory: A historical perspective," Real-Time Systems, vol. 28, no. 2, pp. 101–155, 2004.
- [30] F.-J. Wu, Y.-F. Kao, and Y.-C. Tseng, "From wireless sensor networks towards cyber physical systems," Pervasive and Mobile Computing, vol. 7, no. 4, pp. 397 – 413, 2011.

- [31] J. Ferrer Coll, "Rf channel characterization in industrial, hospital and home environments," pp. xiii, 65, 2012, qC 20120119.
- [32] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS in wireless," in Proceedings of the IEEE INFOCOM, 2009, pp. 486–494.
- [33] I.-Hong Hou, P. R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in Proceedings of the IEEE INFOCOM, 2010, pp. 1–9.
- [34] Plamen Semov, Vladimir Poulkov, Albena Mihovska, Ramjee Prasad. Self-Resource Allocation and Scheduling Challenges for Heterogeneous Networks Deployment. Springer Journal, Wireless Personal Communications. Vol. 87; Issue: 3; pp. 759-777 April 2016. ISSN: 0929-6212. DOI: 10.1007/s11277-015-2640-7.
- [35] Plamen Semov, Hussein Al-Shatri, Krasimir Tonchev, Vladimir Poulkov, Anja Klein. Implementation of Machine Learning for Autonomic Capabilities in Self-Organizing Heterogeneous Networks. Springer Journal, Wireless Personal Communications. 2016. DOI: 10.1007/s11277-016-3843-2
- [36] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in Proceedings of the 2nd international conference on Embedded networked sensor systems, 2004, pp. 95–107.
- [37] C. Lu, B. Blum, T. Abdelzaher, J. Stankovic, and T. He, "RAP: a real-time communication architecture for largescale wireless sensor networks," in Proceedings of the IEEE Real-Time and Embedded Technology and Applications Symposium, 2002, pp. 55–66.
- [38] J. Stankovic, T. Abdelzaher, C. Lu, L. Sha, and J. Hou, "Realtime communication and coordination in embedded sensor networks," Proceedings of the IEEE, vol. 91, no. 7, pp. 1002–1022, 2003.
- [39] X. Cao, P. Cheng, J. Chen, and Y. Sun, "An online optimization approach for control and communication codesign in networked cyber-physical systems," Industrial Informatics, IEEE Transactions on, vol. 9, no. 1, pp. 439–450, Feb 2013.
- [40] 'Smart' home devices used as weapons in website attack, <http://www.bbc.com/news/technology-37738823>
- [41] R. Liu and W. Trappe, *Securing Wireless Communications at the Physical Layer*, Springer, 2010.
- [42] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, Cambridge, 2011.
- [43] Y. Liang, H. V. Poor, and S. Shamai, *Information Theoretic Security*, NOW, 2009.
- [44] Wade Trappe. The Challenges Facing Physical Layer Security. IEEE Communications Magazine, June 2015.
- [45] Bradley JM, Atkins EM, Optimization and Control of Cyber-Physical Vehicle Systems, Sensors (Basel). 2015 Sep 11; 15 (9): 23020-49. doi: 10.3390/s150923020.
- [46] Okuda, R., Kajiwara, Y., Terashima, K.: A survey of technical trend of ADAS and autonomous driving. In: Proceedings of Technical Program—2014 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), pp. 1–4 (2014).
- [47] Abdulmalik Humayed, Bo Luo. Cyber-physical security for smart cars: taxonomy of vulnerabilities, threats, and attacks, ICCPS '15 Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems 2015, pp. 252-253.
- [48] eWALL - Home Caring Environment: [www.ewallproject.eu](http://www.ewallproject.eu)
- [49] Georgi Balabanov, Krasimir Tonchev, Pavlina Koleva, Agata Manolova, Vladimir Poulkov, "Cloud Based Service Bricks Architecture for Ambient Assisted Living System", in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST. Vol. 159, 2015, pp. 69-75. DOI: 10.1007/978-3-319-27072-2\_9. ISBN: 978-331927071-5. Springer Verlag.
- [50] Rahul Mangharam; Houssam Abbas; Madhur Behl; Kuk Jang; Miroslav Pajic; Zhihao Jiang, Three challenges in cyber-physical systems, Communication Systems and Networks (COMSNETS), 2016 8th International Conference on, DOI: 10.1109/COMSNETS.2016.7440015.
- [51] Pedro Carmona; David Nunes; Duarte Raposo; David Silva; Jorge Sa Silva; Carlos Herrera, Happy hour - improving mood with an emotionally aware application, Innovations for Community Services (I4CS), 2015 15th International Conference on, DOI: 10.1109/I4CS.2015.7294480.
- [52] K. Paridari, A. E.-D. Mady, S. La Porta, R. Chabukswar, J. Blanco, A. Teixeira, H. Sandberg, and M. Boubekeur, "Cyber-physical-security framework for building energy management system," in 7th International Conference of Cyber-Physical Systems (ICCPs), 2016.
- [53] Grigore Stamatescu, Iulia Stamatescu, Nicoleta Arghira, Vasile Calofir, Ioana Fagarasan, Building Cyber-Physical Energy Systems, May 2016, arXiv:1605.06903.
- [54] Weisong Shi, Shinan Wang, Building Self-Adaptive Cyber Physical Systems Using Unreliable Components, National workshop for research on high-confidence transportation cyber-physical systems: automotive, aviation & rail, November 18-20, 2008.
- [55] Peter Marwedel and Michael Engel, Cyber-Physical Systems: Opportunities, Challenges and (Some) Solutions, Management of Cyber Physical Objects in the Future Internet of Things Part of the series Internet of Things pp 1-30, Springer 2016.

# Advanced Multi-Constellation EGNSS Augmentation and Monitoring Network (AUDITOR)

E. López<sup>1</sup>, J. Dominguez<sup>1</sup>, B. Quijano<sup>1</sup>, C. Fernández<sup>2</sup>, J. Arribas<sup>2</sup>,  
 M. Hernández-Pajares<sup>3</sup>, A. García-Rigo<sup>3</sup>, M. Schmidt<sup>4</sup>, A. Goss<sup>4</sup>, E. Spaltro<sup>5</sup>, J. Grosso<sup>5</sup>,  
 M. Symeonidou<sup>6</sup>, S. Fountas<sup>6</sup>, Z. Tsiropoulos<sup>6</sup>, F.K. van Everf<sup>7</sup>, P. Blok<sup>7</sup>, D. Roma<sup>8</sup>

**Abstract**— This paper presents the H2020 project AUDITOR. The goal of the project is the implementation of novel precise-positioning differential techniques based on augmentation data in custom GNSS receivers to improve the performance of current augmentation services and reducing costs. These techniques have proven to offer better accuracy with faster convergence times than differential solutions commercially available. Moreover, more sophisticated atmospheric models will be implemented to provide better corrections of ionospheric errors to enable further increased accuracy at the user side. All these advances will be integrated in a software demonstrator that will use public data from GNSS networks and will be able to generate and apply these corrections in near real time.

A custom dual-frequency receiver module will be implemented, following an innovative approach by porting a GNSS software-defined receiver to an embedded system that will integrate hardware accelerators to enable real-time operation in a low power system. The form factor and capabilities of the resulting receiver will be comparable to existing professional market receivers, while retaining all the advantages of software receivers: modularity, scalability, upgradability and flexibility. Besides, providing multi-frequency multi-constellation support, this advanced receiver will allow very low level access to key internals even at sample level, enabling the integration of other complementary techniques. The fact that the software layer will be the evolution of an existing and successful open-source project, GNSS-SDR, will allow GNSS developers and researchers to customize the code of the receiver and tailor it to their own applications or test their algorithms using this flexible receiver module, from reflectometry to ultra-tight GNSS/INS coupling.

Based on the above aspects, low-cost effective precision agriculture (PA) services to farmers will be enabled, especially to those with small and medium-sized businesses in areas of Europe where EGNOS availability may be poor under certain conditions.

**Keywords**— GNSS; augmentation; SRD; receivers; PPP; RTK

## I. INTRODUCTION

This paper presents an overview of the H-2020 project entitled AUDITOR, which started on January 1<sup>st</sup> 2016, with a total duration of 24 months.

AUDITOR is a project of the H2020-Galileo-2015-1 call, topic GALILEO-2-2015, coordinated by ACORDE<sup>1</sup>, with participation from CTTC<sup>2</sup>, UPC<sup>3</sup>, DGFI-TUM<sup>4</sup>, Alpha<sup>5</sup>, DRAXIS<sup>6</sup> and WR<sup>7</sup>.

<sup>1</sup> ACORDE Technologies S.A. (ACO), Santander, Spain

<sup>2</sup> Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels, Spain

<sup>3</sup> Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

<sup>4</sup> Deutsches Geodätisches Forschungsinstitut TUM (DGFI-TUM), München, Germany

<sup>5</sup> Alpha Consultants S.R.L.(ALPHA), Milano, Italy

<sup>6</sup> Draxis Environmental S.A. (DRAXIS), Thessaloniki, Greece

<sup>7</sup> Wageningen Research (WR), Wageningen, The Netherlands

<sup>8</sup> Dept. of Engineering: Electronics, University of Barcelona, Barcelona, Spain

## II. AUDITOR CONCEPT

### A. AUDITOR Objectives

The purpose of AUDITOR is first, to develop an improved GNSS ground-based augmentation system, and, second, to deliver services in precision agriculture based on the new augmentation system.

The GNSS augmentation system will implement novel precise positioning techniques with modern and proven algorithms in highly configurable, cost-effective receivers.

These new receivers will enable cost-effective precision agriculture services to farmers, especially those with small and medium-sized businesses in areas of Europe where EGNOS availability may be poor under certain conditions.

AUDITOR will deliver:

- GNSS network software that enables enhanced features, including advanced precise positioning with high accuracy and fast fix mainly based on Wide Area Real Time Kinematic (WARTK).
- Flexible dual-frequency GNSS receiver platform, hardware/software hybrid, with SDR-based low-power embedded receiver, fully reconfigurable, which will serve as a test-bed for advanced signal processing algorithms, and with low-level access to data in order to enable new functionalities. It will be Galileo/EGNOS enabled, with reconfigurable front-ends for alternative double band operation (L1/E1 and selectable L2 or E5), and a commercial off-the-shelf (COTS) approach for hardware components enabling short time-to-market through a faster implementation process.
- Implementation of GNSS augmentation methods based on open software, including observable generation base on GNSS-SDR [1] and data exchange within the user receiver based on RTKlib [2].
- Web-based platform to collect, store and process georeferenced data for precision agriculture and mobile applications. Services will be provided through an intuitive mobile and web interface, adapted to the needs of farmers. The platform will be highly automated, and simple to implement in different landscapes/agricultural practices.

- Services in precision agriculture based on the above, including data collection and recommendations to farmers regarding site-specific application of water, fertilizer and pesticides.

**B. AUDITOR Approach**

The AUDITOR approach implies innovation in several stages, with a potential multiplier effect. The two most remarkable are:

- Use of GNSS software receivers, in order to overcome the lack of flexibility of existing GNSS receiver in terms of configurability, flexibility and capacity to be upgraded.
- Use of techniques to allow for precise GNSS positioning (based on WARTK) in order to enhance the services of precise GNSS positioning (shorter cold start time, less dense ground network, further increase of accuracy for both single and multi-frequency users).

Based on these advances, AUDITOR will bring specific services to the precision agriculture market.

Fig. 1 illustrates this concept, with the confluence of the scientific and technical advances in precise GNSS with the final aim to offer services for precision agriculture.



Fig. 1. Layout of the main AUDITOR approach

**III. SCIENTIFIC AND TECHNICAL KEY ELEMENTS**

**A. Architecture definition**

AUDITOR high-level architecture is presented in Fig. 2.

The main elements of the proposed architecture are:

- **Dual-band RF GNSS front-end (FE):** configurable RF dual-band multi-constellation receiver which acquires the GNSS signals. The FE embeds all analogue and digital hardware required to convert the RF signal to digital samples. It provides a data output interface with I/Q samples, clock and a management interface (UART/SPI) to configure its main parameters.

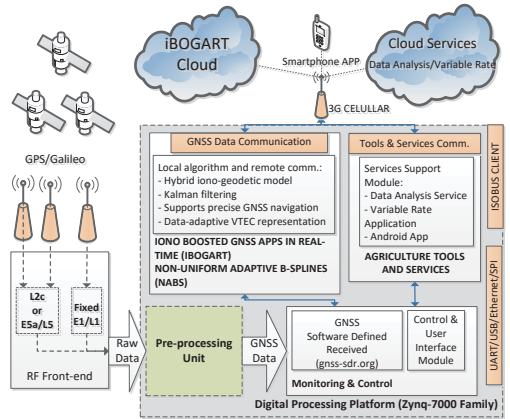


Fig. 2. Architecture of AUDITOR

- **Digital Processing Platform:** hybrid embedded platform with FPGA/ARM hardware

- **Pre-processing unit:** high performance real-time processes that perform the signal acquisition and conversion of the GNSS signals, implementing:
  - GNSS signal sampling and data buffering,
  - low level data pre-processing adapting the data stream throughput,
  - configuration interface for the front-end parameters,
  - GNSS data interface used by the ARM processing system.
- **Monitoring & Control:**
  - Linux OS with third-party libraries and drivers,
  - based on GNSS-SDR library (gnss-sdr.org)
  - Wired external interfaces (SPI / UART / USB / Ethernet)
- **iBOGART model & algorithms:** the ionospheric Boosted GNSS Applications in Real-Time (iBOGART) implements the Central Processing Facility (CPF) and the associated communications, which main functions are:
  - retrieve the real-time NTRIP data stream [3] RT-IGS data providing a serial pipe of data streams,
  - perform pre-processing of the measurements and run the hybrid ionospheric-geodetic model based on TOMION SW [4],
  - implement a Kalman filter estimation based on the previous model,
  - support the precise GNSS navigation by merging the GNSS receiver data and the precise corrections distributed to users
- **Agriculture Tools & Services:** provide the basis to implement higher level services for agriculture end-users:

- precise navigation of autonomous agriculture machinery,
- data analysis service of the positioning data,
- variable rate application based on the data analysis services.
- services will be provided via web/android applications that will be implemented as cloud services

*B. GNSS Receiver Module*

*1. RF front-end and processing platform*

There are two main HW elements in the GNSS receiver the RF Front-End module and the Digital Processing Platform.

The Front-End module (FE) is composed of a RF core that embeds a dual channel GNSS receiver and additional digital control logic to Monitor & Control the main receiver parameters. An overview of the FE module is shown in Fig. 3.

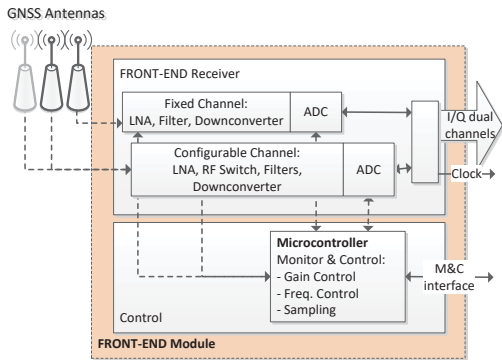


Fig. 3. HW GNSS Receiver

The proposed FE includes a dual-band receiver divided into a fixed channel (Galileo E1 / GPS L1) and a configurable one (GPS L2 or Galileo E5a / GPS E5).

The fixed channel implements an E1/L1 optimized down-converter and hardware sampler to offer a continuous stream of GNSS I/Q samples, and the configurable channel provides extra adaptability to the GNSS receiver allowing selecting between the L2 band and the more convenient dual constellation E5a/L5. This feature is implemented using RF switches and dynamic reconfiguration of the internal mixer frequencies.

Additionally to the previous RF elements, a support microcontroller will be integrated to perform the associated tasks for the FE module.

The Digital Processing Platform supports multiple software subsystems i.e. the GNSS software receiver, monitoring & control tasks, custom network software modules and the external interfaces to the agriculture tools and services.

GNSS baseband processing requires a high computational load. Even when executed in modern, powerful computers, the

number of satellite channels that the receiver is able to process in real-time is limited, especially for BOC-based signals. The solution proposed in AUDITOR is based on a System-on-Chip architecture that combines power-efficient, all-purpose ARM® processors and powerful FPGAs in a single chip based on the Zynq®-7000 family [5]. This allows executing the GNSS software receiver and other software modules in the ARM® processors, while offloading the most demanding computations to the FPGA.

*II. Software-defined GNSS receiver*

In the proposed architecture, the flow of digitized and quantized signal samples delivered by the analog-to-digital converter is entirely processed in real-time by a software application running in the processing platform. Hence, the whole GNSS baseband processing chain, from the output of a RF front-end up to computation of GNSS observables and the position solution, is fully defined by software. This approach provides and unparalleled degree of flexibility in terms of which and how the GNSS observables are obtained and delivered to the network.

The implementation of such software-defined receiver is based on GNSS-SDR [1], a free and open source software application which modular design allows full receiver customization. By the beginning of the project, the GPS L1 C/A and Galileo E1 OS processing chains were already included in GNSS-SDR, while Galileo E5b, GPS L2C and L5 are being developed in the AUDITOR framework. Its open source license ensures the community inspection and its continuous improvement, while its modular design allows for a further AUDITOR-specific customization, in which the most computationally demanding parts of the baseband processing are implemented in VHDL and executed on the FPGA device embedded in the Zynq processor, while all the other processing is being run on general-purpose ARM devices, allowing the software application to be run in real-time in a small-sized, low-cost user equipment. This innovative approach provides advantages in many dimensions: it dramatically cuts development, deployment and operational costs; the integration of design, development, testing and deployment in the same workflow ensures rapid prototyping, testing and user feedback; it ensures dissemination (GNSS-SDR is now included as a software package in major GNU/Linux distributions, such as Debian and Ubuntu), maintainability and technological endurance beyond the project framework; and provides a solid base of highly-skilled users and use cases that continually improve the source code in many different senses.

*C. Advanced Algorithms for GNSS*

The contributions of DGGI-TUM are focusing on the development of high accuracy GNSS algorithms to provide enhanced ionospheric corrections. This includes two major issues, namely the

1. improvement of existing mapping functions to convert the slant total electron content (STEC) measurable by GNSS into the vertical total electron content (VTEC), and the

- development of a data and signal adaptive approach to produce high resolution VTEC maps.

In two-dimensional global approaches the ionosphere is often identified as a spherical layer of an infinitesimal thickness in which all the electrons are concentrated. The height of this so-called single layer model (SLM) – the ionospheric effective height (IEH) – corresponds approximately to the altitude of the maximum electron density and it is usually set to values between 350 and 450 kilometers [6]. The improvement of the mapping function shall be achieved by estimating more realistic values for the so-called effective ionospheric height (EIH) by using radio occultation data, e.g. from the Formosat-3/COSMIC mission.

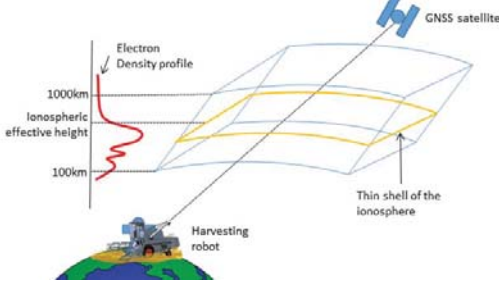


Fig. 4. Geometry of the SLM; electron density profile

Usually VTEC is modeled by series expansions in terms of spherical harmonic base functions [7]. However, it is a well-known fact that the computation of the corresponding series coefficients requires a homogeneous data distribution. An adaptive approach for global VTEC modelling based on B-spline functions has already been developed at DGFI-TUM to consider the inhomogeneous distribution of the observation sites in a previous project. Within AUDITOR an improved version of this adaptive approach will be developed and incorporated into the system architecture to provide higher quality products. This approach will allow for the sequential processing of measurements. In other words, the data of different ground GNSS networks together with observations of the flexible low-cost receivers shall be sequentially processed by means of a Kalman filter to produce high rate VTEC maps.

#### D. Network and user SW for precision agriculture

At the network side, the CPF will offer a passive service, computing and distributing WARTK corrections in real time through the Internet to any potential AUDITOR user. The CPF output messages will contain many different messages, including the ionospheric model, the tropospheric estimation, the satellite clocks, and the rover navigation status, among others.

In order to derive such terms, it is necessary to run a Kalman filter (taking into account a process noise for each term to be modelled) to solve the navigation equations and derive the post-fit residuals as well as the partial TEC values for each ionospheric voxel in a dual-shell tomographic modelling (i.e. TOMION model; [4]). In addition, this ionospheric information

is compacted in a slant ionospheric model per satellite, and stored in WARTK messages (labelled DSM). This is typically slightly better in Wide Area scenarios than broadcasting 3D ionospheric grid values, with an implicit mitigation of the mapping function errors, extending the RTK distances up to several hundreds of kilometres, among other advantages.

In addition, double differences are applied between pairs of satellites and pairs of receivers and double differenced postfit residuals are derived, fixing or constraining WARTK-required ambiguities (Bc, Bi, Bwc), checking wrong ambiguity fixes and looking for small undetected cycle-slips

At the user side, the application of such WARTK messages is considered of great value to reach decimeter or few centimeter positioning accuracy even though the baseline wrt a reference permanent receiver is hundreds of kilometres away. In this context, the corrections enable the WARTK user to build its own STEC values through interpolation, and get a prompt precise guess of the ambiguities of both carrier phases.

Note that the WARTK user algorithm is based on the fast ionospheric-free ambiguity constraint thanks to the availability of precise ionospheric corrections from the iBOGART Cloud (the WARTK CPF, running the hybrid ionospheric-geodetic model) and the smoothed double-differenced Melbourne-Wübbenna combination.

In AUDITOR, RTKlib as the baseline PVT (Position, Velocity, Time) solver at the user side, is being considered. This is considered of key importance to facilitate integration of WARTK algorithms inside the COTS receiver at a later stage.

Last but not least, the Travelling Ionospheric Disturbances (TIDs) are a source of errors for single-frequency users, and mostly useful information for multi-frequency ones, that will be mitigated in the case of AUDITOR users with respect to the reference receiver. In particular, the most frequent ionospheric wave signatures are the Medium Scale TIDs (MSTIDs; with typical periods ranging from several minutes to less than one hour, and velocities from 50 to 300 m/s; [8] since they introduce a differential error that is non-linearly dependent on the baseline distance, affecting directly to precise GNSS positioning.

#### E. Implementation for precision agriculture

Precision agriculture is a method of farming based on measuring and responding to temporal and/or spatial variability [9] in soils and crops (Fig. 5). Spatial variability of soils may be expressed in maps of soil properties such as lutum content, organic matter content, and soil depth. Spatial variability of crops is often expressed in yield maps. In many cases, when soil and/or crop maps are available, the optimal amounts of fertilizer, water and pesticides that are needed can be determined precisely. Thus, the cost of production is reduced and, at the same time, the negative impact on the environment is lessened.



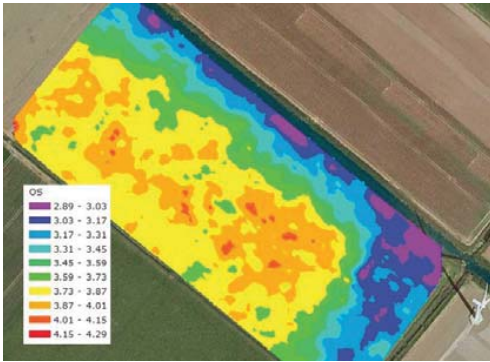


Fig. 5. Map showing soil organic matter content (“OS”) for a field in The Netherlands. Soil variability is high, with organic matter content ranging from barely 3% to more than 4%

The making of soil and crop maps, as well as the spatially varying application of fertilizer which the maps enable, is utterly dependent on the availability of a high-quality GNSS signal. However, in several areas of Eastern and South Europe GPS accuracy is much less than the standard provided by existing augmentation services. This fact is a significant burden for site specific crop management and precision agriculture in general.

High accuracy and precision in navigation and positioning in agriculture is also needed when autonomous mobile robotic units (Fig. 6) are deployed [10]. Increasingly, agricultural robots are developed to navigate in the demanding environment, to identify weeds, pests and diseases, to identify the crops’ needs for nutrients or water, and to apply fertilizers and pesticides.

AUDITOR focuses on three aspects of precision agriculture. First a service will be developed to store, with centimeter accuracy, data collected by sensors mounted on agricultural vehicles, combined with the positioning data of the new receiver. A web portal will ingest the georeferenced measurements. Farmers will be able to apply spatial analysis on the data for gaining optimized and accurate information and maps about their crops.



Fig. 6. Agricultural robots developed by Wageningen Research.

Second, a service targeted to variable rate application for high value crops (orchards, vineyards, vegetables) and arable crops will be developed. By using it, the farmer will be able to get optimized data for irrigation, fertilization or pesticide needs

thus achieving optimal plant growth conditions, saving valuable resources and reducing production costs. For achieving this, farmers will apply decision rules on Data Analysis Service stored data and then will download a task map to their equipment (e.g. pesticide sprayer).

Finally, the new GNSS receiver will be used to improve the accuracy of positioning of agricultural machinery and agricultural robots, both from one pass over the field to the next, and from one year to the next. An Android application will be developed that will allow users through its Graphical User Interface (GUI) to monitor, control and configure the navigation system.

#### IV. MARKET OPPORTUNITY

AUDITOR’s focus is mainly on **PA in Europe**, given its initial aim to develop a cost effective PA service for farmers with small and medium-sized businesses in the continent. However, given the nature of the solution, AUDITOR’s augmentation service is highly scalable and could cope with other geographies with vast agricultural areas (e.g. LATAM and Africa) in the mid-term, also addressing big businesses of these regions, in which investments in machinery are usually lower than in developed regions. In general, the major drivers for this market are the ever increasing requirement for augmented yield and profitability and energy and cost savings, whilst the high initial investments still remains as the primal barrier for farmers to introduce precision agriculture products in their assets. The consortium believes that by focusing on existing applications of PA in Europe, AUDITOR can compete on a market valued at roughly 180 €m in 2023, composed by c. 300 k GNSS devices by that year.

In the near future, other applications for PA with similar requirements of decimetre-level real-time positioning may strongly benefit from the results of AUDITOR, namely Remotely Piloted Aircraft Systems (**RPAS**) and **agriculture robots**. A very optimistic outcome from the research of the US Association for Unmanned Vehicle Systems International (AUVSI), puts PA at approximately 80% of the potential commercial market for UAVs in the USA, which translates, from the estimations, on a value of 21.8 €b within 2014-2024.

Along with PA, other professional activities requiring high precision positioning, such as **surveying, civil engineering, construction and mining** have always been early adopters of new location technologies. Given the potential portability of the product and service stemming from AUDITOR, these could be of interest in the long term.

#### V. CONCLUSIONS

AUDITOR developments will lead to a solution with several improvements as compared to existing systems:

- The use of WARTK/MSTIDs techniques permits longer baselines between the permanent receivers in the GNSS network, reducing dramatically the required infrastructure to provide prompter and more accurate corrections to enable precise positioning.

- The newest improvements and findings on MSTIDs modeling will be taken into account, resulting into less accuracy performance limitations.
- Reduction in convergence times.
- Hardware engine and embedded software implementation to allow low-power real time operation while retaining all the flexibility of SDR implementations, targeting specifically open, low cost and multi-frequency Galileo and EGNOS receivers

All these advances will be offered to the Precision Agriculture market in different key services, such as agricultural machinery navigation, data analysis service or variable rate applications.

#### ACKNOWLEDGMENT

This paper has been generated within the AUDITOR project, which is being funded by the European Union's H2020 Programme, under grant agreement number 687367.

#### REFERENCES

- [1] GNSS-SDR: Global Navigation Satellite System Software Defined Receiver. <http://gnss-sdr.org/project>
- [2] RTKlib: An Open Source Program Package for GNSS Positioning. <http://www.rtklib.com/>
- [3] Weber, G., Dettmering, D., & Gebhard, H. (2005). Networked transport of RTCM via internet protocol (NTRIP). In *A Window on the Future of Geodesy* (pp. 60-64). Springer Berlin Heidelberg.
- [4] Hernández-Pajares, M., J.M. Juan, J. Sanz, and O.L. Colombo (2000). Application of ionospheric tomography to real-time GPS carrier-phase ambiguities resolution, at scales of 400-1000 km and with high geomagnetic activity, *Geophysical Research Letters*, 27:13, 2009-2012.
- [5] Zynq 7000: <http://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html>
- [6] Hernandez-Pajares M., Juan J.M., Sanz J., Garcia-Fernandez M.: Towards a more realistic ionospheric mapping function. XXVIII URSI General Assembly, Delhi, 2005
- [7] Schaer, S.: Mapping and predicting the Earth's ionosphere using the Global Positioning System. PhD thesis, Bern Univ., Switzerland, 1999
- [8] Hernández-Pajares, M., Juan, J. M., Sanz J., Aragón-Ángel A. (2012) Propagation of medium scale traveling ionospheric disturbances at different latitudes and solar cycle conditions, *Radio Sci.*, 47, RS0K05, doi:10.1029/2011RS004951.
- [9] Kempenaar C, van Evert Fk, Been T, Kocks Cg & Westerdijk Ce (2016) Towards data-intensive, more sustainable farming: advances in predicting crop growth and use of variable rate technology in arable crops in the Netherlands. In: 13th International Conference on Precision Agriculture (ICPA), St. Louis, MO, USA.
- [10] Hiremath Sa, van Der Heijden Gwam, an Evert Fk, Stein A & Ter Braak Cjf (2014) Laser range finder model for autonomous navigation of a robot in a maize field using a particle filter. *Computers And Electronics In Agriculture* 100, 41-50

# Performance Analysis of Filter Bank Based Multicarrier Waveform for Future Wireless Systems

Deepa T and T. Rama Rao

Department of Telecommunication Engineering,  
SRM University, Chennai, Tamil Nadu, India  
deepa.t@ktr.srmuniv.ac.in, ramaraoatr@gmail.com

**Abstract**— Multicarrier Modulation (MCM) systems partition the transmission band into narrower sub-bands, which enables multiplexing the symbols in the frequency domain. Among the existing MCM systems the orthogonal frequency division multiplexing (OFDM) is the most prominent since it has been considered in wireless systems with several advantages. In this work, a filter bank based multicarrier (FBMC) waveform with advantages over OFDM according to a set of future wireless system requirements is introduced. Simulation results show that the FBMC technique offers high spectrum resolution, little spectral regrowth and can provide independent sub-channels, while maintaining or enhancing the high data rate capability compared with the conventional OFDM.

**Keywords**— Multicarrier Modulation, Filter bank based multicarrier, IoT, OFDM, Future Wireless System.

## I. INTRODUCTION

DATA traffic demand in access networks has been exponentially increased as a result of the expansion of various applications due to mobile devices. The requirements on future access networks for 5G wireless communication after LTE-A are currently being discussed [1-2]. 5G is the next generation of wireless communication technology that is expected to be rolled out around the year 2020. While the standard for 5G has yet to be defined, it is expected that 5G will open up opportunities for the Internet of Things (IoT) and will also deliver far more flexibility than current generation technologies. Improvements are expected to be seen in the form of better coverage, spectral efficiency, data rates, signaling efficiency, lower power consumption [3]. The research of new PHY's in 5G NOW aims to offer scalable and spectrally malleable modulations that are attractive for 5G requirements, especially addressing non-orthogonal and asynchronous principles within the trinity of the corner scenarios MTC, CoMP/HetNet, carrier aggregation, and Tactile Internet. Moreover, in order to support various applications and legacy services, a future access network should be flexible, scalable, reliable and spectrum efficient.

The IoT will certainly play a key role, as it extends the internet today by attaching a multitude of sensors/actors (e.g. power meters) to it wirelessly via 5G [4]. It is yet too early to fully survey all possible use cases of the IoT, however, many of them will rely on devices differing heavily from the ones attaching to 4G, but business models have not

started off yet. And, a system incorporating IoT devices may not be able to apply as tight synchronization requirements as 4G does due to energy and cost reasons. The IoT could change the way we see the Internet as a human-to-human interface toward a more general machine-to-machine platform. Ensuring completely orthogonal wireless access becomes costly or even impossible, hence robust modulation schemes with a confined spectral response beyond orthogonal frequency division multiplexing (OFDM) may be a solution. Therefore, there are currently three foremost waveforms for the 5G wireless mobile communication systems which has attracted more and more researchers, and now filter bank based multicarrier (FBMC) has been used in many fields and the newer Universal Filtered Multi-Carrier (UFMC) [5 - 6].

Multicarrier modulation (MCM) exhibit a higher robustness against multipath propagation. Among the existing MCM systems the OFDM is the most prominent since it has been considered in several systems such as, IEEE 802.16a and IEEE 802.11a/g [7]. It is important to remark that OFDM shapes the signals transmitted on each sub-band with a rectangular window. However, major drawbacks of OFDM systems are the high spectral side-lobe due to the use of rectangular window and transmission of redundancy in the form of cyclic prefix (CP), which reduces the spectral efficiency and is sensitive to the frequency offset [8]. Aiming at overcoming all these issues, an FBMC techniques have been heavily promoted [9]. FBMC applies a filtering functionality to each of the subcarriers in contrast to OFDM while dividing the spectrum into multiple orthogonal sub bands. So, with FBMC the side-lobes are much weaker and thus the inter-carrier interference issue described above is by far less crucial than with OFDM. Therefore, a system applying FBMC is much more suited to a potential 5G system to that respect.

In this paper, we discussed and compared FBMC waveform with unquiet advantages over OFDM according to a set of 5G requirements. FBMC transmission technique has been recently attracted to offer a competitive alternative to the long established OFDM as the modulation scheme of choice in future wireless communication and networks. Section II gives an evaluation of filter bank multicarrier techniques. In Section III, the filter bank multicarrier waveform for IoT applications is discussed. The advantages of FBMC waveform over OFDM is analyzed in section IV. Conclusion and future work is presented in Section V.

II. EVOLUTION OF FILTER BANK MULTICARRIER TECHNIQUES

DIGITAL filter banks are now widely used in source coding, for audio, image, and video signals, as well as for communication systems, for instance, in trans-multiplexers and multicarrier modulations. In this regard, several filter bank multicarrier structures that are proposed in the literature. Among filter bank based multicarrier structures are discrete wavelet multitone (DWMT), filtered multitone (FMT), cosine modulated multitone (CMT), modified discrete Fourier transform (MDFT), exponentially modulated filter bank (EMFB), and FBMC/OQAM [10-11]. The last is the most popular filter bank scheme that results in high bandwidth efficiency.

Evolution of FBMC techniques, the first multicarrier methods that were introduced, were based on filter bank [12]. FBMC modulation can be considered as an evolved OFDM. The first proposal came in the 1960 [13], the conditions required for signaling a parallel set of Pulse amplitude modulated (PAM) symbol sequences through a bank of overlapping vestigial side band (VSB) modulated filters are presented. This work was extended with transmission of Quadrature amplitude modulated (QAM) symbols [14]. In 1980s, an efficient poly-phase implementation for OFDM based offset QAM is proposed [15]. This method is also referred to as staggered modulated multitone (SMT). A particular type of FBMC, the so-called FBMC/OQAM (or OFDM/OQAM) system, consisting of pulse shaped OFDM carrying offset QAM (OQAM) symbols, has received increasing attention due to, among other features, its potential for maximum spectral efficiency [16]. In the 1990s, the application to data transmission for very high speed digital subcarrier line (VDSL) technology led to more work on two classes of FBMC communication systems, namely FMT and DWMT modulation.

FMT modulation is an interesting solution for VDSL transmission, being intermediate to other proposed single carrier and multicarrier methods, as well as providing some unusual advantages in terms of spectrum management, unbundling, and duplexing. FMT generates digitally a set of tightly packed single-carrier-like multiband signals with minimum analog filtering requirements and offers filtering-based implementation. This modulation technique can be applied to a number of difficult transmission environments, not limited to VDSL, but possibly including home local area networks (LANs), corporate LANs, cable modems and possibly even wireless transmission. In FMT, the adjacent subcarriers do not overlap as they are separated by guard bands. Hence, FMT is less bandwidth efficient than the FBMC methods [17].

In [18], an adaptively modulated optical filter bank multicarrier with offset quadrature amplitude modulation-based side lobes suppression, and verified both the improvements in signal performance and spectral efficiency compared with OFDM in a multi-sender- supported optical network after a 50-km transmission. In the literature [19], [20] comparison of FBMC and OFDM is compared in time-

frequency efficiency and concluded that FBMC outperformed OFDM by about 10% in case of very short packets while performing similar for long sequences.

III. FILTER BANK MULTICARRIER WAVEFORM FOR IOT APPLICATIONS

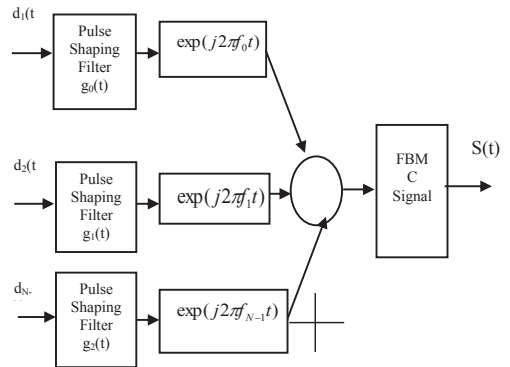


Fig.1 Block diagram of FBMC/OQAM Modulator

In this scheme, a filter is applied per sub carrier and can be modulated as shown in Fig.1. For FBMC, each of the individual sub-channel is filtered on its own. It uses the very narrow band filter with long time length. This gives us very good control of each filter bank. It gives a good control over emission of each of the subcarriers. The important thing is that each of the subcarrier goes through a filter called 'Pulse Shaping Filter [21]. Therefore the subcarrier signals are shaped with very good spectral efficiency and very good data rate but because no CP is added. FBMC modulation, which is also implemented based on IFFT/FFT pairs, and a prototype filter designed with the Nyquist pulse shaping principle is adopted, which can reduce greatly the spectral leakage problem of OFDM [22]. This results in negligible inter symbol interference (ISI) and inter carrier interference (ICI).

$$S(t) = \sum_{n=0}^{N-1} \sum_{k=-\infty}^{\infty} d_{n,k}(t)g(t - kT) \exp[j2\pi\Delta fn(t - kT)] \tag{1}$$

Where,  $d_{n,k}(t)$  are modulated symbols associated with nth subcarrier of the kth multicarrier symbol  $g(t)$  is the prototype filter of the symbol period duration. N is the total number of subcarriers;  $\Delta f$  is subcarrier spacing.

IV. RESULTS & DISCUSSION

The critical steps for FBMC is to implement filters for each sub channels and align the multiple filters into a filter

bank. We design a basic form of a filter called prototype filter. Frequency response of prototype filter is shown in Fig.2. Frequency domain prototype filter of FBMC signed is expressed as;

$$P(f) = H(f) = \sum_{k=-(K-1)}^{K-1} H_K \frac{\sin(\pi(f - \frac{k}{MK})MK)}{MK \sin(\pi(f - \frac{k}{MK}))} \quad (2)$$

Where M is the number of sub-channels and K is the overlapping factor. Then the impulse response of the FBMC's prototype filter is given as;

$$IFFT[P_{FBMC}(f)] = p(t) = 1 + 2 \sum_{k=1}^{K-1} H_K \cos\left(2\pi \frac{kt}{KT}\right) \quad (3)$$

where, T is the time duration of FBMC symbol. Fig. 3 illustrates the prototype filters of OFDM and FBMC in frequency domain when the overlapping factor K = 4. In the conventional OFDM, prototype filter is a rectangular pulse of height 1 and width T. In frequency domain the large side lobes result in high out of band emission of signal power among subcarriers. In contrast, the duration of prototype filter in FBMC systems is 'KT'. Therefore, FBMC symbols overlap in time domain but leaks negligible signal power beyond different subcarriers.

As FBMC applies a filtering functionality on a per subcarrier basis, the performance of FBMC systems in burst transmission is impacted by the length of the impulse response of the prototype filter. Assuming that N multicarrier symbols have to be transmitted in a burst. With FBMC, the emitted burst begins with the first signal sample associated with the first symbol and it finishes with the last signal sample associated with the last symbol. Due to the overlapping factor K, the length of the burst is increased by K-1 symbol periods, compared to a scheme with no overlapping, such as OFDM.

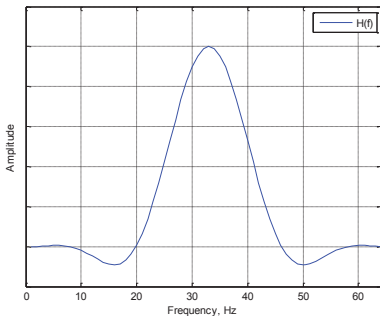


Fig.2 Frequency Response of Prototype filter

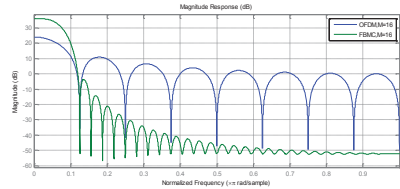


Fig.3. Spectra of the prototype filter for OFDM and FBMC

An FBMC burst is shown in Fig.4, for K=4. The initial and final transition phases are both equal to (K-1)/2 symbol periods. It is clear from the figure that, the signal in some parts of the transitions is very small. However, if the spectra of adjacent users must be fully preserved, it is necessary to keep the entire transitions. Conversely, if a user is separated from the neighboring users by spectral gaps, then the transitions can be shortened, with the corresponding increase in data rate.

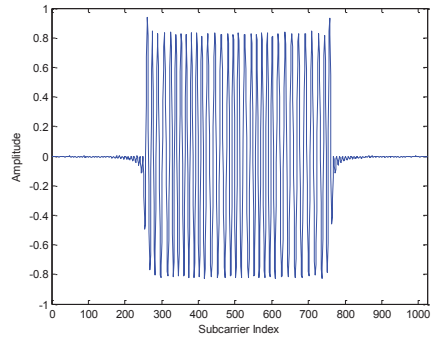


Fig.4 Structure of the transmitted burst with FBMC

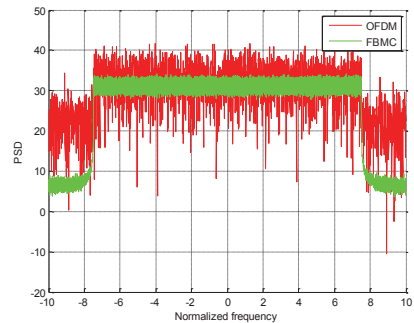


Fig.5 Spectra of OFDM and FBMC sub-channels

Essentially, filter banks offer more degrees of freedom which can be exploited to mitigate certain issues associated with OFDM. For instance, OFDM suffers from poor spectral

selectivity since the frequency response of adjacent sub-channels overlap significantly with each other. The poor spectral selectivity may pose problems in the presence of narrowband noise because sub-channels adjacent to the narrowband noise will provide a rather poor attenuation which can severely affect the performance of OFDM. In this case, filter banks can be designed to provide much better spectral selectivity. Fig.5 shows the power spectrum for conventional OFDM Waveform and FBMC waveform. As shown in Fig.5, FBMC signal exhibits little spectral regrowth, which can increase the immunity of signals from out of band leakage.

## V. CONCLUSION

In this paper, we discussed and compared FBMC waveform with OFDM according to a set of future wireless systems requirements. FBMC transmission technique has been recently attracted to offer a competitive alternative to the long established OFDM as the modulation scheme of choice in future wireless communication networks. FBMC modulations have emerged as a definitely attractive solution since no redundancy is transmitted and subcarrier signals are shaped with very spectral efficient waveforms. Simulation results shown that the FBMC outperforms the CP-OFDM. FBMC techniques lead to an enhanced a new physical layer which offers performance gains and additional functionalities in conventional networks. Hence, FBMC technique is defined as an applicable candidate for future wireless Communications.

## REFERENCES

- [1] Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: Next-generation wireless broadband technology," *IEEE Wireless Commun.*, vol.17, no.3, pp.10–22, Jun. 2010.
- [2] M. Agiwal, A. Roy and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617-1655, 2016.
- [3] A. Barki, A. Bouabdallah, S. Gharout and J. Traoré, "M2M Security: Challenges and Solutions," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1241-1254, 2016.
- [4] G. Fettweis, "A 5G Wireless Communications Vision," *Microwave J.*, vol. 55, ed. 12; pp. 24–39, Dec. 2012.
- [5] Estella, A. Pascual-Iserte, and M. Payaró, "OFDM and FBMC performance comparison for multistream MIMO systems," in *Proc. Future Netw. Mobile Summit*, Jun. 2010, pp. 1–8.
- [6] V. Vakilian et al., "Universal Filtered Multi-Carrier Technique for Wireless Systems Beyond LTE," 9th Int'l. Wksp. Broadband Wireless Access, IEEE GLOBECOM '13, Atlanta, GA, Dec. 2013.
- [7] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. London, U.K.: Artech House, 2000.
- [8] G. Li and G. Stuber, *Orthogonal Frequency Division Multiplexing for Wireless Communications*, Springer, New York, NY, USA, 2006.
- [9] H. Zhang, D. Le Ruyet, D. Roviras, Y. Medjahdi, and H. Sun, "Spectral efficiency comparison of OFDM/FBMC for uplink cognitive radio networks," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 621808, 2010.
- [10] B. Farhang-Boroujeny and C. H. Yuen, "Cosine modulated and offset QAM filter bank multicarrier techniques: A continuous-time prospect," *EURASIP J. Adv. Signal Process.*, vol. 2010, Jan. 2010.
- [11] L. Lin and B. Farhang-Boroujeny, "Cosine-modulated multitone for very-high-speed digital subscriber lines," *Eurasip Journal on Applied Signal Processing*, vol. 2006, Article ID 19329, 16 pages, 2006.
- [12] R. Chang, "High-speed multichannel data transmission with bandlimited orthogonal signals," *Bell Sys. Tech. J.*, vol. 45, pp. 1775–1796, Dec. 1966.
- [13] B. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Transactions on Communication Technology*, vol. 15, no. 6, pp. 805–811, 1967.
- [14] R. P. Ramachandran and P. Kabal, "Synthesis of bandwidth efficient OQAM and VSB transmultiplexers," *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, Albuquerque, NM, 1990, pp. 1639-1642 vol.3.
- [15] S. Nedic, "An unified approach to equalization and echo cancellation in OQAM-based multi-carrier data transmission," *Global Telecommunications Conference, 1997. GLOBECOM '97., IEEE*, Phoenix, AZ, 1997, pp. 1519-1523 vol.3.
- [16] H. Nam, M. Choi, S. Han, C. Kim, S. Choi and D. Hong, "A New Filter-Bank Multicarrier System With Two Prototype Filters for QAM Symbols Transmission and Reception," in *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 5998-6009, Sept. 2016.
- [17] N. Benvenuto, S. Tomasin and L. Tomba, "Equalization methods in OFDM and FMT systems for broadband wireless communications," in *IEEE Transactions on Communications*, vol. 50, no. 9, pp. 1413-1418, Sep 2002.
- [18] Sun-Young Jung, J. Sang-Min and H. Sang-Kook, "AMO-FBMC for Asynchronous Heterogeneous Signal Integrated Optical Transmission Photonics Technology Letters, IEEE, 2015. 27(2): p. 133-136.
- [19] Schaich, F., T. Wild and Y. Chen, "Waveform Contenders for 5G - Suitability for Short Packet and Low Latency Transmissions. 2014 IEEE 79th VTC Spring, p. 1-5; Seoul, South Korea. 2014.
- [20] B. Farhang-Boroujeny, "OFDM Versus Filter Bank Multicarrier," in *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 92-112, May 2011.
- [21] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1170–1183, May 2002.
- [22] H. Saeedi-Sourck, Y. Wu, J. W. M. Bergmans, S. Sadri, and B. Farhang-Boroujeny, "Complexity and performance comparison of filter bank multicarrier and OFDM in uplink of multicarrier multiple access networks," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp.1907–1912, Apr. 2011.

# Optical Channel Analysis of Turbo coded MIMO-OFDM system for Visible Light Communication

Sabitha Gauni, Manimegalai C.T and T. Rama Rao.  
SRM University, Kattankulathur, Chennai, India

sabitha.g@ktr.srmuniv.ac.in,manimegalai.c@ktr.srmuniv.ac.in,ramaraotr@gmail.com

**Abstract**—Optical Wireless Communication (OWC) system uses Intensity Modulation with Direct Detection (IMDD) to modulate the intensity of the signal in order to transmit information. There is an increase in the exploration of the modulation techniques used in Visible Light Communications (VLC) in recent times. In this paper a power efficient modulation technique of Digital Pulse Interval Modulation (DPIM) and Dual-header Pulse Interval Modulation (DH-PIM) is compared for VLC systems. Further, power allocation is studied under AWGN and Rayleigh channel. It is seen that the results prove to be spectrally efficient in the analysis of MIMO-OFDM VLC systems.

**Keywords**—Optical wireless communication, Visible Light Communication, DPIM, DH-PIM, MIMO-OFDM

## I. INTRODUCTION

The research in the field of optical wireless communication (OWC) and visible light communications (VLC) is unending as the bandwidth varying from 430 THz to 750 THz is very attractive [1]. VLC in essence has a dual purpose of illumination and communication. There is an added advantage of having cheap end devices such as LED and photo-detectors. The modulation schemes used for Visible Light Communication are limited, for the reason that for intensity modulation with direct detection (IMDD) the modulated signal is to be unipolar and real. The analysis of OWC systems with modulation techniques like Digital Pulse Interval Modulation (DPIM) and Dual-header Pulse Interval Modulation (DH-PIM) have gained interest [2].

The COFDM technique is particularly suited to provide reliable reception of signals affected by strong distortions, as it can be the case for terrestrial broadcasting. Multipath propagation is known to limit the performance of high bit-rate modulation schemes. The principle of COFDM relies in splitting the information to be transmitted over a large number of carriers, in such a way that the signaling rate on each of them becomes significantly lower than the assumed channel coherence bandwidth [3]. The codes are constructed by using two or more component codes on different interleaved versions of the same information sequence. For a system with two component codes, the concept behind turbo decoding is to pass soft decisions from the output of one decoder to the input of the other decoder, and to iterate this process several times so as to produce more reliable decisions. Because of this advantage turbo codes are now part of a large number of communication standards.

The paper is sectioned as follows: Section 2 gives a brief description about the system model for Optical Wireless Communication. Section 3 then goes on to describe the proposed system model. Section 4 presents the performance results, followed by conclusion in Section 5.

## II. SYSTEM MODEL FOR OPTICAL WIRELESS COMMUNICATION

Optical wireless transceivers intended for mass-market applications are likely to have tight cost constraints imposed upon them [4]. Consequently, it is highly desirable that the chosen modulation technique is rather simple to implement achieving excellent power efficiency and/or bandwidth efficiency is of little use if the scheme is so complex to implement that cost renders it unfeasible. Thus the two techniques of DPIM and DH-PIM are explored [5].

### 2.1 Modulation techniques for VLC

**2.1.1 Differential Pulse Interval Modulation:** DPIM is an asynchronous modulation technique, in which each block of 'M' input data bits is mapped to one of 'L' possible symbols of different length. 'L' is varying. A symbol is composed of a pulse of one slot duration followed by a series of empty slots, the number of which is dependent on the decimal value of the M-bit data stream being encoded. The symbol length of the DPIM is unfixed [6]. Also because the symbol length is varying detection at the receiver is done through clock synchronization at the receiver and not symbol synchronization. This simplifies the system and is an advantage over PPM [7].

**2.1.2 Double Header Pulse Interval Modulation:** DH-PIM modulation is more complex, the time slot included by each symbol is also mutative, but the symbol adopts two kinds of starting pulse. The symbol  $s_k$  formed by a head slot and m empty time slots followed [8]. The head time slot is included by  $\alpha+1$  time slots ( $\alpha$  is integer). Considering two forms of head H1 and H2, the H1 initial pulse width is  $\alpha/2$  time slot, followed by  $(\alpha/2)+1$  protected time slots; H2 pulse width is a time slots, followed by one time slot. When  $k < 2^{M-1}$ , the head time slot of symbol  $s_k$  is H1, otherwise it is H2.

## 2.2 Advantages of Differential modulation techniques

2.2.1 Power Efficiency: The modulation techniques are compared in terms of average optical power required for a certain target bit error rate (BER) performance or signal to noise ratio(SNR). In other words, average power required to achieve given BER at a given data rate [9]. Power efficiency is given as the energy per pulse per average energy per bit.

DPIM's average power requirement and power efficiency is given by,

$$P_{avgDPIM} = \sqrt{\frac{2N_0R_b}{R^2L_{DPIM}\log_2L}} Q^{-1}(P_{ebitDPIM}) \quad (1)$$

$$\mu_{DPIM} = \sqrt{\frac{8}{(L+1)\log_2L}} \quad (2)$$

For DHPIM is the average power requirement and the power efficiency is given by:

$$P_{avgDH-PIM} = \sqrt{\frac{2N_0R_b}{16R^2L_{DH-PIM}M}} Q^{-1}(P_{seDH-PIM}) \quad (3)$$

$$\mu_{DH-PIM} = \sqrt{\frac{9\alpha^2}{2M(2^{M-1}+2\alpha+1)}} \quad (4)$$

2.2.2 Bandwidth Efficiency: The optical bandwidth is practically limitless. The modulation schemes which require high bandwidth are more susceptible to Inter Symbol Interference(ISI) and as result incurring more power penalty [10].Transmission Bandwidth is defined as minimum bit slot ( $T_{min}$ ).

$$B_{req} = \frac{1}{T_{min}} \quad (5)$$

DPIM:

$$T_{sDPIM} = \frac{T_bM}{L_{DPIM}} \quad (6)$$

$$B_{reqDPIM} = \frac{L+1}{2M} R_b \quad (7)$$

DH-PIM:

$$T_{sDH-PIM} = \frac{2M}{(2^{M-1}+2\alpha+1)R_b} \quad (8)$$

$$B_{reqDH-PIM} = \frac{(2^{M-1}+2\alpha+1)R_b}{\alpha M} \quad (9)$$

2.2.3 Transmission Reliability: A study based on packet transmission rate is more convenient as the symbol length varies in most of the modulation techniques. The size of packet is assumed to be constant of  $N_{pkt}$  bits [11]. The average slot length is given by:

$$L_{pkt} = \frac{\bar{L}N_{pkt}}{M} \quad (10)$$

$\bar{L}$  is the valid code combinations. Therefore the packet transmission rate,  $R_{pkt}$  is given by :

$$R_{pkt} = \frac{R_sM}{LN_{pkt}} \quad (11)$$

$R_s$  is the slot rate and  $T_s$  the slot duration.

$$\text{For DPIM, } R_{pktDPIM} = \frac{2MB_{req}}{N_{pkt}(2^{M+1})} \quad (12)$$

$$\text{For DH-PIM, } R_{pktDH-PIM} = \frac{\alpha MB_{req}}{N_{pkt}(2^{M-1}+1+2\alpha)} \quad (13)$$

## III. PROPOSED SYSTEM MODEL

The figure 1 shows the block diagram of the proposed Turbo coded OFDM system. Information bits are generated by a Random Data Generator and are coded at the Turbo encoder. These coded bits are then bit-interleaved and de-multiplexed into several sub-channels. The de-multiplexed bits are mapped to a constellation point for the given modulation scheme. This is then followed by Binary to Decimal conversion in preparation for modulation. The modulator performs IFFT on the encoded decimal bits to convert from frequency domain to time domain signal. In accordance with OFDM technique, a Cyclic Prefix is appended to the head of each symbol. The previous two steps of IFFT conversion and the appendage of cyclic prefix forms the core of OFDM Modulation. After the mapping and modulation of OFDM symbols, the transmitter sends the bits over the channel to the receiver. The channel is assumed to be a frequency-selective Additive White Gaussian Noise (AWGN) channel.

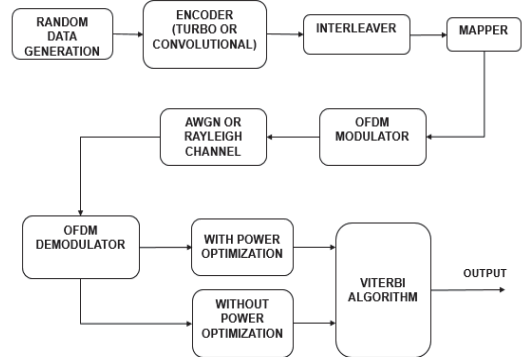


Fig. 1. System Model

At the receiver, the received signal is demodulated, converted back from decimal to binary, de-interleaved and decoded by a Turbo Decoder as follows. Discrete Fourier transform (DFT) is used to convert the time-domain signal to a frequency-domain signal. After the DFT, the received



signal for each sub channel is used to compute the soft metric of each coded bit at the de-mapper. The soft metric is de-interleaved at the de-interleaver and sent to the decoder. The receiver also estimates the gain and phase components of the channel response for coherent detection. This estimation is usually performed with pilot signals such as pilot tones and pilot symbols. A feedback from the receiver to transmitter will help the transmitter to allocate different levels of power to various sub channels but at the same time keeping the total transmit power constant. Therefore, the performances of coded OFDM systems are improved with efficient power allocation. The received data symbols at the decoder are used for computing the error rate in transmission and coding.

A. TURBO CODED OFDM SYSTEMS

Turbo coding involves the Parallel Concatenation of convolutional codes with message interleaving. The Turbo coded OFDM system has the advantages of less power consumption, less complexity and also shows better performance compared to convolutional codes. The Convolutional codes do not break the message stream like the block codes. The combination of turbo codes with the OFDM transmission is so called Turbo Coded OFDM (TC-OFDM) can yield significant improvements in terms of lower energy needed to transmit data, a very improvement issue in personal communication devices [12]. It is shown, however, that there is a large potential gain in using the iterative property of turbo decoders where soft bit estimates are used together with the known pilot symbols. The performance of such an iterative estimation scheme proves to be of particular interest when the channel is strongly frequency- and time- selective.

1. Turbo Encoder

The encoder for a turbo code is parallel concatenated convolutional code, the block diagram of which is given in Figure 2[14]. The binary input data sequence is represented by  $d_k = (d_1, \dots, d_n)$ . The input sequence is passed into the input of a convolutional encoder. ENC1 and a coded bit stream,  $X_{k1}$  is generated. The data sequence is then interleaved. That is, the bits are loaded into a matrix and read out in a way so as to spread the positions of the input bits. The bits are often out in a pseudo-random manner.

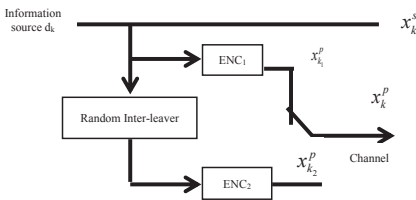


Fig. 2. Turbo Encoder

The interleaved data sequence is passed to a second convolutional encoder ENC2, and a second coded bit stream,  $X_{k2}$  is generated [13]. The code sequence that is passed to the modulator for transmission is a multiplexed (and possibly punctured) stream consisting of systematic code bits  $X_{k1}$  and parity bits from both the first encoder  $X_{k1}$  and the second encoder  $X_{k2}$ .

2. Turbo Decoder

A block diagram of a turbo decoder is shown in Fig. 3. The input to the turbo decoder is a sequence of received code values,  $R_k = \{Y_{sk}, Y_{pk}\}$  from the demodulator. The turbodecoder consists of two component decoder –DEC1 to decode sequences from ENC1, and DEC2 to decode sequences from ENC2. Each of these decoders in a Maximum-A-Posteriori (MAP) decoder [14]. DEC1 takes as its input the received sequence systematic values  $Y_{sk}$  and the received sequence parity values  $Y_{pk1}$  belonging to the first encoder ENC1. The output of DEC1 is a sequence of soft estimates EXT1 of the transmitted data  $d_k$ . EXT1 is called extrinsic data, in that it does not contain any information which was given to DEC1 by DEC2. This information is interleaved, and then passed to the second decoder DEC2 [15]. DEC2 takes as its input the (interleaved) systematic received values  $sk$  and the sequence of received parity values  $Y_{pk2}$  from the second encoder ENC2, along with the interleaved form of the extrinsic information EXT1, provided by the first decoder. DEC2 outputs a set of values, which, when de-interleaved using an inverse form of interleaver, constitute soft estimates EXT2 of the transmitted data sequence  $d_k$ . This extrinsic data, formed without the aid of parity bits from the first code, is feedback to DEC1. This procedure is repeated in an iterative manner.

The iterative decoding process adds greatly to the BER performance of turbo codes. However, after several iterations, the two decoders' estimates of  $d_k$  will tend to converge. At this point, DEC2 outputs a value ( $\hat{d}_k$ ); a log likelihood representation of the estimate of  $d_k$ . This log likelihood value takes into account the probability of a transmitted '0' or '1' based on systematic information and parity information from both component codes. More negative values of ( $\hat{d}_k$ ) represent a strong likelihood that the transmitted bit was a '0' and more positive values represent a strong likelihood that a '1' was transmitted. ( $\hat{d}_k$ ) is de-interleaved so that its sequence coincides with that of the systematic and first parity streams. Then a simple threshold operation is performed on the result, to produce hard decision estimates,  $d_k$ , for the transmitted bits.

The decoding estimates EXT1 and EXT2, do not necessarily converge to a correct decision. If a set of corrupted code bits form a pair of error sequence that neither of the decoders is able to correct, then EXT1 and EXT2 may either diverge, or converge to an incorrect soft value. In the next sections, the algorithms used in the turbo decoding process, within DEC1 and DEC2.

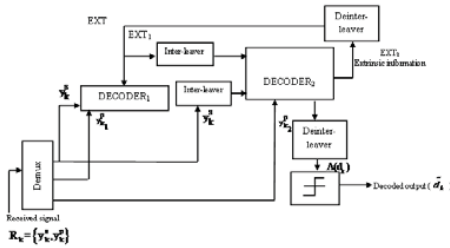


Fig. 3. Turbo Decoder

**B. SOFT OUTPUT VITERBI ALGORITHM (SOVA)**

The Conventional Viterbi Algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states – called the Viterbi path – that results in a sequence of observed events. It simulates encoder state, given the received values. SOVA is a slightly modified version of the conventional Viterbi algorithm. The algorithm is modified to provide a soft output for each decoded bit. SOVA has two modifications over the conventional Viterbi algorithm. First, the path metrics which are used to select the maximum likelihood path through the trellis are modified to take account of a-priori information. Secondly, the algorithm is modified to provide a soft output for each decoded bit.

The algorithm followed by the Viterbi Algorithm is shown in Figure 4. The arrow marks indicate the probabilities from one state to another. There are two states considered in this system: With Power Optimization and Without Power Optimization[6]. The transition probabilities from one state to another and itself are given in the figure, and these values are pre-set by the programmer. The arrows converging onto the states represent the start probabilities, the arrows between them indicate the transition probabilities, and finally the arrows emerging from the two states are called the emerging probabilities. These values are then used by the Viterbi Decoder to compute the error in the system.

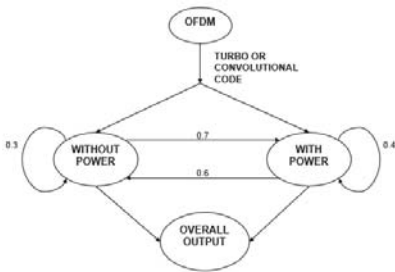


Fig. 4. The Viterbi Algorithm

**C. RAYLEIGH FADING CHANNEL**

The effect of the wireless channel on the transmitted signals is multiplicative, where the multiplicative term is a complex Gaussian random variable. If the channel coefficient has zero mean, then such a channel is considered as Rayleigh fading since the absolute value of the received amplitude is a Rayleigh random variable. The PDF of the received signal amplitude in a Rayleigh fading channel is given as

$$f(x) = \frac{x}{\sigma^2 \alpha} \exp\left(-\frac{x^2}{2\sigma^2 \alpha^2}\right) \quad \forall x > 0. \quad (14)$$

In the above equation, x has a Rayleigh Distribution,  $\sigma^2$  is the variance of received signal amplitude and  $\alpha$  is the fading parameter.

**IV. PERFORMANCE RESULTS**

The following are the simulation results of the proposed system.

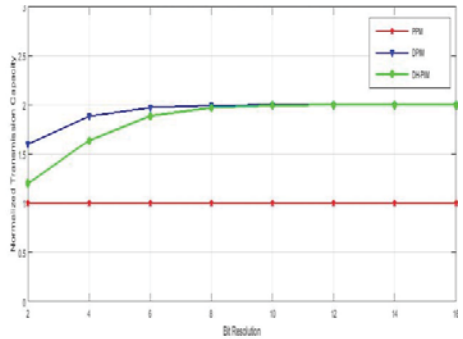


Fig. 5. Normalized transmission capacity versus 'M'

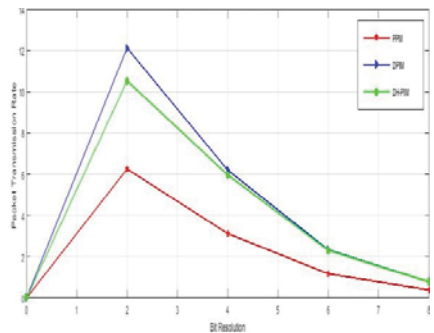


Fig. 6. Packet Transmission Rate versus 'M'

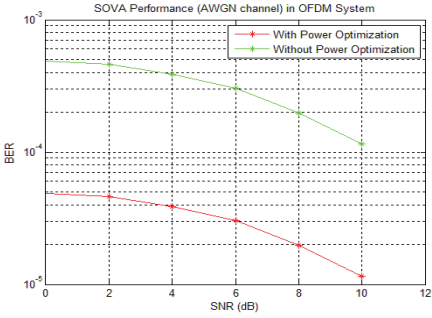


Fig. 7. Performance of SOVA (AWGN Channel) in OFDM system.

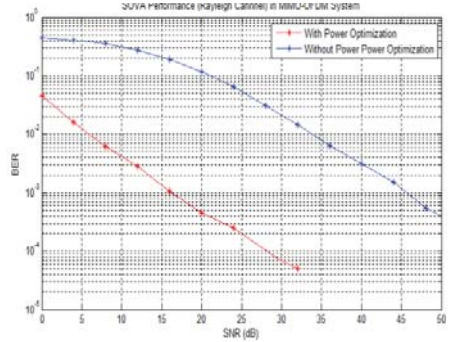


Fig. 10. Performance of SOVA (Rayleigh Channel) in MIMO-OFDM system.

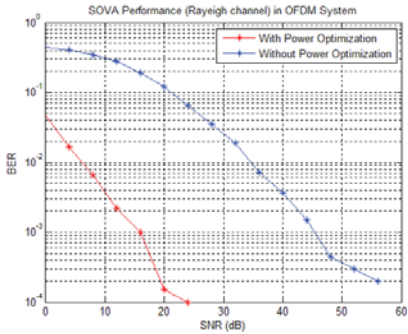


Fig. 8. Performance of SOVA (Rayleigh Channel) in OFDM system.

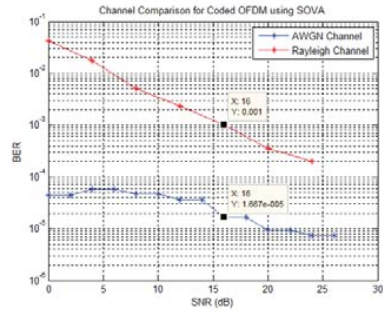


Fig. 11. Comparison of SOVA Performance in AWGN and Rayleigh Channels.

The figure 5 shows the same plot but with the transmission capacities normalized to the transmission capacity of PPM. DPIM offers higher capacity at lower values of  $M$  than DH-PIM but the difference between them is negligible at high orders. The transmission capacities of DPIM and DH-PIM are about four times as that of PPM for a particular value of  $M$ . The figures 7 and 8 show the performance of the SOVA algorithm in AWGN and Rayleigh channels respectively. It is clear from the plot that Rayleigh channel encounters more BER than AWGN, due to the presence of multiple paths.

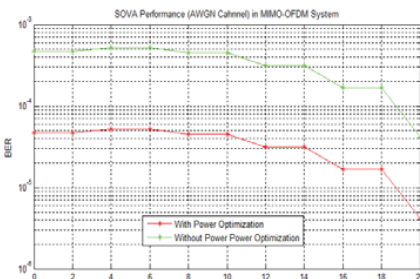


Fig. 9. Performance of SOVA (AWGN Channel) in MIMO-OFDM system.

The figures 9 and 10 depict the same algorithm performed in a MIMO-OFDM system for both AWGN and Rayleigh Channels respectively. Finally, Fig. 11 shows the comparison between the performances in AWGN channel and Rayleigh channel, when SOVA algorithm is applied to a MIMO-OFDM scenario.

## V. CONCLUSIONS

The performance of the optical communication system was compared with respect to the efficiency of modulation techniques used. For lower values of 'L' in DPIM and DH-PIM, they are not efficiency in terms of power and bandwidth usage. Hence we incorporate the MIMO-OFDM systems with the Turbo coding. Eventually for higher value of  $M$  the transmission rate for DPIM and DH-PIM are the same. By viewing the performance graphs of the algorithm in various scenarios, it is observed that when power is optimized, the results show reduced BER. When the existing system in OFDM is extended to MIMO, there is a significant increase in SNR as almost double the value when observed in an OFDM system. The SNR in an OFDM system is about 10 dB whereas in a MIMO-OFDM systems it is 20 dB, observed in AWGN channel. In

Rayleigh channel, the increase in SNR is about 2-3 dB. At around 15 dB, the error rate for AWGN is 0.00001, while the error rate for Rayleigh channel is 0.001, because of the multipath degradation in Rayleigh channel.

#### REFERENCES

1. Burchardt H., Serafimovski N., Tsonev D., Videv S., and Haas H., "VLC: Beyond point to point communication," IEEE Communications Magazine, pp. 98-105, July 2014.
2. Choubey A. and Sharma D., "Analyzing of Various Modulation Methods on Basis of Error Performance for VLC," IEEE EESCO Conference, pp. 1-5, Jan 2015.
3. Yu, Z., "Adaptive Modulation schemes for optical wireless communication systems", Ph.D. dissertation, School of Eng., Univ. Warwick, 2010.
4. Yan G., and Min W., "Performance Research of Modulation for Optical Wireless Communication", ISCST, pp. 357-360, August 2010.
5. C. Lethien et al., "Energy-autonomous picosecond remote antenna unit for radio-over-fiber system using the multiservice concept," Photonics Technology Letters, vol. 24, no. 8, pp. 649-651, Apr. 2012.
6. Motoharu Matsuura and Jun Sato, "Bidirectional Radio-Over-Fiber Systems Using Double-Clad Fibers for Optically Powered Remote Antenna Units", IEEE Photonics Journal, Vol. 7, No. 1, February 2015
7. Z. Jia, J. Yu, Y.-T. Hsueh, A. Chowdhury, H.-C. Chien, J. A. Buck, and G.-K. Chang, "Multiband signal generation and dispersion-tolerant transmission based on photonic frequency tripling technology for 60-GHz radio-over-fiber systems," IEEE Photon. Technol. Lett. **20**(17), 1470-1472, 2008.
8. Q. Chang, H. Fu, and Y. Su, "Simultaneous Generation and Transmission of Downstream Multi-band Signals and Upstream Data in a Bidirectional Radio over Fiber System," IEEE Photon. Technol. Lett. **20**(3), 181-183, 2008.
9. Y. Hsueh, Z. Jia, H. Chien, J. Yu, and G. Chang, "A Novel Bidirectional 60-GHz Radio-Over-Fiber Scheme With Multiband Signal Generation Using a Single Intensity Modulator," IEEE Photon. Technol. Lett. **21**(18), 1338-1340 2009.
10. L. Zhang, X. Hu, P. Cao, T. Wang, and Y. Su, "A Multiband Radio over Fiber System Using a Single-drive Mach-Zehnder Modulator," in *Proceeding of ACP 2010*, FP5.
11. Q. Chang, Y. Tian, J. Gao, T. Ye, Q. Li, and Y. Su, "Generation and Transmission of Optical Carrier Suppressed-Optical Differential (Quadrature) Phase-Shift Keying (OCS-OD(Q)PSK) Signals in Radio Over Fiber Systems," J. Lightwave Technol. **26**(15), 2611-2618, 2008.
12. Mohammad Salim, R.P. Yadav, S. Ravi Kanth "Performance Analysis of Log-map, SOVA and Modified SOVA Algorithm for Turbo Decoder", International Journal of Computer Applications, pp. 0975-8887, Volume 9- No.11, November 2010.
13. R.P. Yadav, Mohammad Salim, S.Ravikanth "Performance Analysis of Log-map, SOVA and Modified SOVA Algorithm for Turbo Decoder", International Journal of Computer Applications, pp. 0975-8887, Volume 9- No.11, November 2011.
14. Bernard Le Floch, Michel Alard, Member, IEEE, And Claude Berrou, Member, IEEE, "Coded Orthogonal Frequency Division Multiplex", vol. 30, no. 13, June 2001.
15. M. Cluzeau, M. Finiasz and J. P. Tillich, "Methods for the Reconstruction of Parallel Turbo Codes," Proceedings of the IEEE International Symposium on Information Theory (ISIT), Austin, 13-18 June 2010, pp. 2008-2012.

# A Hybrid Pedestrian Navigation Based on Activity Sequence Recognition, Map Matching and HMM

Beihang Zhang<sup>1</sup>, Zhuqing Jiang<sup>1</sup>, Guanghua Zhang<sup>3</sup>, Rui Qu<sup>1</sup>, Yang Liu<sup>1</sup>, Aidong Men<sup>1</sup>, Yi Zhao<sup>2</sup>, Shanbao He<sup>2</sup>

<sup>1</sup>Beijing University of Posts and Telecommunications, Beijing, China

<sup>2</sup>China Academy of Space Technology, Beijing, China

<sup>3</sup>Shenzhen Skycaster Micronics Inc, Shenzhen, China

**Abstract**—Nowadays, it is a fundamental need for tracking and navigating based on smartphone applications. The applications are used to support guidance for visitors in public buildings like museums and libraries. The smartphone equipped with better inertial sensors can be used for pedestrian dead reckoning(PDR). In the indoor environment, the human activities often change at the special points connected with the indoor map, such as turning at the corner, and walking stairs. Hence in this paper, an indoor localization system is proposed, using smartphone and indoor map matching algorithm. The system fuses dead reckoning data from the smartphone and location-related activities by a Hidden Markov Models(HMM). In other words, the location-related activities(walking, sitting, turning and up and down stairs) are linked by HMM. It is the core idea that exploits the position information and the available location-related activities to reduce the accumulative error and provide useful information for navigation system. The predicted location would converge on the true one after several detected activities, which is easy to implement. We also propose a method to eliminate interruptions of the random activities. The method is evaluated in field experiment and the result shows that it is robust and accurate.

**Keywords**—Indoor tracking;Pedestrian dead reckoning;Activity detection;Label Activity;Hidden Markov Model(HMM)

## I. INTRODUCTION

In these years, the service based on the indoor location is widely used. People spend most of their time in the indoor environment. The location service is applied in retail, healthcare, museum-visiting, etc. The GPS technology is used widely in the outdoor environment, but it is failed in the indoor environment.

With more accuracy sensors used in the smartphone, there has been much research in the field. Generally speaking, there are three ways in indoor navigation: the inertial sensor-based approach, the wireless radio fingerprint scheme and the visual navigation technology[1–3]. The inertial sensor-based approach is supported with high-precision sensor components like accelerometers, gyroscopes, etc. A pedestrian dead reckoning proposed by Schindhelm[4] has been widely researched. It works as a stand-alone method and is easily applied. But the inertial sensor-based method is disturbed by accumulative errors. The wireless radio fingerprint scheme can reach high accuracy with the light and radio wave to track the pedestrians position[5–7]. The scheme includes signal propagation models and signal matching strategies[8], but it requires a costly infrastructure and access point(AP) and fingerprint map, and it cannot adapt to the environment changes. The visual navigation technology is originated in robot navigation. The basic idea is constructing or updating a map of an unknown environment

while simultaneously keeping track of an agent's location. Some algorithms like SLAM(simultaneous localization and mapping)[9] and SURF(Speeded-Up Robust Features)[10] are widely used. In the recent study, Hardegger proposed a ActionSLAM[11] using Action to update the map and location.

Many fusion algorithms have been put forward to solve the problem that the cumulative location errors increase with time. The fusion technology with Wi-Fi and inertial sensors can reach high accuracy[8]. With the development of Google Indoor Maps or NAO CLOUD, the floor plan can be used to limit the error offset. Existing map matching techniques, such as HMM[3], Conditional Random Field(CRF)[12, 13], Kalman[8] and particle filters[1], have been successfully applied. Inspired by these techniques, a fusion scheme is proposed to track the pedestrian path.

The physical action recognition has been studied for many years. The original research focuses on the law of acceleration[14]. In the recent years, people concentrate more on the machine learning method. The research focuses on the selection of the action features, for example the mean value or the variance, the choice of the different classifiers. [15] uses the unsupervised learning for human activity recognition with smartphone sensors. In our action recognition method, we improve the precision by the use of the feature of the barometer. We not only classify the human action but also label the activity in a time window.

In this paper, the navigation model in the indoor environment such as museums and libraries is studied in detail. First, the improved PDR algorithm is introduced. Second, we propose our action recognition method and improve the recognition precision. To eliminate the random actions, we label the activities based on the sequence of the actions. Third, the ideal HMM strategy is proposed fitting with the pedestrian tracking. In the proposed HMM model, it is also taking the random actions into consideration. The location result would converge on the true path even without initial point. Through several experiments, our system is validated to be effective and practical providing accurate position in indoor environments.

The rest of the paper is organized as follows. In Section II, we provide a brief overview of our proposed system. In Section III, an improved PDR algorithm is detailed. And in Section IV, the action recognition algorithm and the method of activity labeling are described. Section V describes an ideal HMM based on map matching algorithm. A novel inertial position algorithm based on the HMM is proposed. In section VI, the experiment is implemented in an indoor environment and the results are analyzed. Section VII is the conclusions. And the section VIII is the acknowledgment.

## II. SYSTEM SOLUTION OVERVIEW

The overall framework of the system is shown in Fig.1. This proposed system consists of three parts including the dead reckoning, the activity labeling and the map matching algorithm based on the HMM. The input of the system is divided into four parts: acceleration, the value of the gyroscope, barometric pressure and the magnetic intensity. The map based on the activity mode has been loaded in the phone in advance.

The accelerometer is used for collecting the acceleration and calculating the step number and stride length. The gyroscope data is used for the calculation of angular displacement in real time. The barometric pressure is used for detecting the floor changes. And the magnetic and the acceleration together calculate the orientation. In our system, the four sensors data are also used for human motion mode recognition. In our map matching algorithm, we get a map fusing ground truth and possible motion state. And then the HMM-based human motion algorithm is used which accurately matches motion events to the right motion detection, resistance of false positive/negative detection. The system avoids the cumulative errors and is easy to implement.

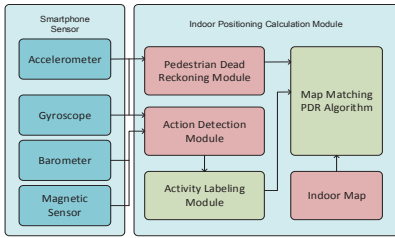


Fig. 1. The Proposed Indoor Positioning System(Improved PDR)

## III. PEDESTRIAN DEAD RECKONING

The pedestrian dead reckoning consists of two parts: distance estimation and direction estimation. We can get the absolute heading by the digital compass or get the relative heading by integrating the gyroscope measurements. To correct the error caused by the digital compass, the former study combines it with the gyroscope measurements by the Kalman filter[8] or PF filter[1]. Besides, in our approach, we use the barometer to detect the floor change. So we introduce the barometer feature in part C.

### A. Distance Detection

First of all, the acceleration is used to detect the step of human according to [16]. The total acceleration magnitude  $A = \sqrt{a_x^2 + a_y^2 + a_z^2}$  is used to detect the step and motion mode as well. It is regarded as a step when the acceleration exceeds the threshold which is adjusted dynamically. According to the laws of the human motion, the limit of walking speed is  $0.2s/step$ . It means that the interval between two steps

is above  $0.2s$ . The dynamic threshold is set as the following equation.

$$T_n = \alpha \cdot T_{old} + \beta \frac{Max_i + Min_i}{2} + \gamma \quad (1)$$

$T_n$  is the dynamic threshold.  $\alpha, \beta, \gamma$  are the pre-learned parameters.  $Max_i$  and  $Min_i$  are the maximum and minimum values of the acceleration signal in one period [16]. The walking distance is expressed as  $d = \sum_{i=1}^n l_i$ , and the  $d$  is the distance,  $n$  is the number of the steps and  $l_i$  can be expressed as followed.

$$l_i = a_{step} \cdot f_i + b_{step} \cdot v_i + c_{step} \quad (2)$$

$a_{step}, b_{step}, c_{step}$  are the pre-learned parameters obtained during training stage and  $f_i$  is the stride length and  $v_i$  is the acceleration variance of the  $i$ th step.

We can get the  $f_i$  and the  $v_i$  from the following equations.

$$f_i = \frac{1}{t_i - t_{i-1}} \quad (3)$$

$$v_i = \frac{1}{N} \sum_{t=t_{i-1}}^{t_i} (a_t - \bar{a}_i)^2 \quad (4)$$

$t_i$  describes the  $i$ th step,  $a_t$  denotes the acceleration at the time  $t$ .  $\bar{a}$  is the mean values and  $N$  is the number of samples in one step. We conduct an experiment that people put the smartphone in their pocket. The result is shown in the Fig.2. The error rate is less than 1%.

### B. Direction Estimation

We use the gyroscope in our system to get the relative angular variation and the digital compass to get the absolute orientation. To deal with the different ways the smartphone held, the acceleration and the gyroscope data are combined together. The  $XYZ$  axis system is applied for the pedestrians with  $X$ -axis pointing to the right hand, the  $Y$ -axis pointing to the forward and the  $Z$ -axis pointing to the sky while the  $xyz$  axis system is for the phones. In our paper,  $\bar{a}_x, \bar{a}_y, \bar{a}_z$  are the three values of the acceleration separately in each axis. And the three axes angular displacement is  $\theta_x, \theta_y, \theta_z$ . We can get the  $Z$ -axis' angular displacement by the following equation.

$$O_z = \frac{\bar{a}_x}{g} \theta_x + \frac{\bar{a}_y}{g} \theta_y + \frac{\bar{a}_z}{g} \theta_z = \frac{\bar{a}_x \theta_x + \bar{a}_y \theta_y + \bar{a}_z \theta_z}{g} \quad (5)$$

An experiment is conducted to verify the algorithm, which is shown in Fig.4. The user walked first in a straight line, took a  $90^\circ$  turn, and then walked on. The phone was put in the pocket unbending position. The result shows the angular displacement is  $91.54^\circ$ .

### C. The Barometer Feature

In order to get the accurate floor information, the barometer is used. The accuracy of the barometer in the smartphone is enough to distinguish the floor change. In this Section a brief introduction about the barometer is explained. According to the U.S. Standard Atmosphere, the relationship between the air pressure and the altitude can be expressed as follows.

$$P_r = P_0 \left(1 - \frac{H}{44330}\right)^{5.255} \quad (6)$$

In this paper,  $P_0$  means the standard atmospheric pressure, and  $H$  is the altitude in meters and  $P_r$  is the measured atmospheric pressure. But because of the different of the local atmospheric environment and the accuracy of the sensors, it is more reasonable to compare the air pressure to the preloaded data.

The  $H$  is the physical quantity relative to the sea level, but the actual we need is the local elevation. With the help of the pre-loaded local atmospheric pressure information, the floor information can be calculated. The experiment is conducted in a 4 floor building and the result is shown in Fig.3. It can be clearly seen that the air pressure changes between the different floors. Besides, in the following part, the pressure is used to distinguish the action of climbing stairs from walking.

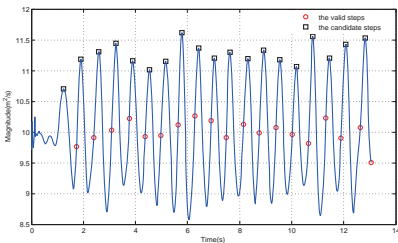


Fig. 2. Step Detection

#### IV. ACTION RECOGNITION AND ACTIVITY LABELING

It has been popular for decades that the researchers use the acceleration data for action recognition. The action defined as the minimum human motion. In this paper, we not only study on improving the accuracy of the human action recognition, but also try to label the actions sequence and use the labels as the landmarks in the map. The landmarks can overcome the mismatch problem of the unknown actions. The approach is introduced in the following parts.

##### A. Action Recognition

With the advancement in the mobile phone technology and emergence of smartphone containing lot of sensors, physical action recognition is realized by the usage of the acceleration sensor. Much research focuses on improving the precision of the action recognition. Machine Learning method includes Naive Bayes, Decision Tree, K-Nearest Neighbor(KNN) and Support Vector Machine(SVM) classifiers, etc. Different features are used including the mean value, the maximum and the minimum value, the standard deviation and the signal entropy, etc. Wu et al. [17] evaluated different classifiers on the three activities (Walking, jogging, using stairs) using mean, standard deviation and fast Fourier transform as features and obtained average accuracy of 90% using KNN classifier. Cesar and Marco [18] proposed a human action hierarchical recognition system based on time-domain features and neural networks without the need of the smartphones to be constrained to a single fixed body position, which achieved the accuracy above 92%. [15] proposes unsupervised learning methods for action

recognition even when the number of actions is unknown. The mixture of Gaussian exactly distinguishes those actions achieve above 90%.

In this paper, the mean and deviation of three axis acceleration value and the total acceleration magnitude  $A$  are chosen as the features and extracted using a window of  $w$  with  $w/2$  overlap value between consecutive windows. The sampling frequency of smartphone in this paper is 50Hz. According to the human motion rule, each window representing 2 seconds is proper. Though different overlap values can be used, but 50% overlap values has been shown to have produced reasonable results. We choose KNN(K=3) classifier with 10-fold stratified cross validation and the result is shown in the Table.I.

In the study, we can see that the upstairs and downstairs are

TABLE I.  
CONFUSION MATRIX FOR ACTIVITY RECOGNITION

	Walk	Sit	Stand	US	DS	Turn	U-Turn	True count
Walk	476	0	0	12	19	0	0	507
Sit	0	538	7	0	0	0	0	545
Stand	0	11	673	0	0	0	0	684
US	18	0	0	349	31	0	0	398
DS	11	0	0	24	402	0	0	437
Turn	0	0	0	0	0	50	0	50
U-Turn	0	0	0	0	0	0	50	50

promiscuous with the walking. The similar result also proven by the former study[18, 19]. The smartphone is integrated with the barometer as the technology development. Air pressure is inversely proportional to the height in the bottom layer of the earth's atmosphere, which means altitude increases as air pressure decreases. Therefore, barometer could not only be utilized to calculate the altitude of indoor building, but also can be used as the feature for the action recognition. With the air pressure added to the data set, the result shows that actions can be easily distinguished and is shown in the Table.II.

We use the data of the gyroscope and compass to detect the turning. The window is set to 3 seconds. Normal turn occurs at the corner and the U-turn occurs randomly. In our study, the range of the normal turn is set between 50 to 135, and if the value exceed the normal turn, it is regarded as U-turn.

In this part, we classify the different activities(turn, U-turn, climbing the stairs, walking, standing and sitting) with high classification rates, above 98%.

TABLE II.  
CONFUSION MATRIX FOR ACTIVITY RECOGNITION IN THE PAPER

	Walk	Sit	Stand	US	DS	Turn	U-Turn	True count
Walk	502	0	0	3	2	0	0	507
Sit	0	538	7	0	0	0	0	545
Stand	0	11	673	0	0	0	0	684
US	5	0	0	384	0	0	0	398
DS	9	0	0	0	428	0	0	437
Turn	0	0	0	0	0	50	0	50
U-Turn	0	0	0	0	0	0	50	50

##### B. Activity Labeling

The action of turning around the corners can be used as the landmarks in a single floor. When the system detects the turn, the data fusion algorithm would correct the cumulative errors. But if the turn does not happen around the corner, it will result in false match. So it is more reasonable to label

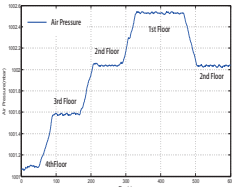


Fig. 3. Air Pressure and Floor Detection

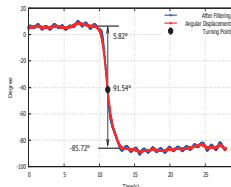


Fig. 4. Angular Displacement Around Z-axis

the human activities in the building and add the false match algorithm. So in the HMM part, a method is proposed to delete the false activity from the activity sequence. In this part, we consider the indoor position used in the public building such as museums, galleries.

In the indoor environment like museums, the activities are divided into two classes: location-related activity(LRA), and location-unrelated activity(LUA). We define the LRA that the activity may happen when the visitors pass through special points. For example, when the visitor turns at the corner or climbs the stairs, the activity will be detected. The LUA can happen at any time and any position. We use the LRA for map matching to eliminate the accumulation errors and discard the LUA. The LRA includes turning at the corner, climbing the stairs, standing in front of some landmarks and sitting in the chairs, etc. The LUA includes standing and U-turning. In the action recognition part, we classify the different actions. The results are prepared for labeling the activities.

In a single floor, turn is an iconic action. An activity sequence window is used and the turn is considered as the key action as shown in the Table.III. The standing in front of the displays(SFD), real-turn and sitting are considered as the landmarks. The turn following the first two activity would be deleted. Because the turn is redundant. Then in the single floor, the special points can be used.

In the multiple floor environment, with the help of the pre-loaded air pressure and the actions, it is easy to map the location to the special point. The end of the stairs is defined as the entrance(exit) of one floor.

 TABLE III.  
THE ACTION LABEL RULES IN SINGLE FLOOR

action-sequence	stand	real-activity
turn	stand	SFD
turn	sit	sit
turn	//	real-turn

## V. HMM MAP MATCHING ALGORITHM

In this part, we propose Hidden Markov Models based on indoor map and previous work, and describe how to model particular problem. The modeling will use the link-node model that represents the connectivity between feature locations. And then the true location is positioned during the navigation.

### A. Hidden Markov Model Introduction

Hidden Markov Model is a statistical model, which is used to describe a Markov process containing a set of hidden states and observations. It is represented by a finite

set of states and each state is connected to a probability distribution. The HMM focuses on analyzing correlation of spatiotemporal information. The transition generated between the hidden states is called **transition probabilities** and the probability between the observations and the hidden states is called **emission probabilities** in our paper and shown in the Fig.5. A formal characterization of HMM is represented as  $\lambda = \{S, V, A, B, \pi\}$  as follows:

1)  $\{S_1, S_2, S_3, \dots, S_N\}$  — A set of hidden states.  $N$  is the number of states in the model;

2)  $\{v_1, v_2, v_3, \dots, v_M\}$  — A set of distinct observations.  $M$  is the number of observation symbols per state;

3)  $A = [a_{ij}]$  — An  $N \times N$  matrix for the state transition probability distributions, where  $a_{ij}$  represents the probability of making a transition from state  $S_i$  to  $S_j$ ,  $a_{ij} = p(q_{t+1} = S_j | q_t = S_i)$ ;

4)  $B = [b_j(k)]$  —  $b_j(k)$  is the observation probability distribution in each of the states,  $b_j(k) = p(v_k | q_t = S_j)$ ;

5)  $\pi = [\pi_i]$  is the initial state distribution,  $\pi = p(q_1 = S_i)$ .

We use the  $\lambda = (S, v, A, B, \pi)$  to represent an HMM. Based on the sequence of the observation states, we want to find the most probability sequence of the hidden states, which is the sequence of the locations in the paper.

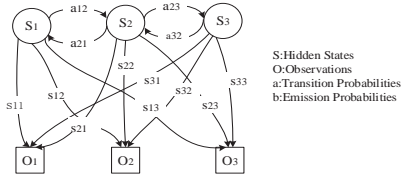


Fig. 5. the structure of the HMM

### B. HMM with Indoor Map Algorithm

The observations is the users's location getting from the PDR module and the hidden state is the real location. The node in the map is important. First, we define the node attribute as the follows:

- 1)coordinate of the node;
- 2)coordinate of the neighbor nodes,accessible direction(the direction of the neighbor nodes), and accessible distance(max distance of the accessible direction);
- 3)node type(the corresponding human activities happening at the node in the map).

For example, the attribute of node  $n$  may be:

- 1) $(x_n, y_n)$ ;
- 2) $\{l, m, n, p\}$  (neighbor nodes);  
 $\{w_l, w_m, w_n, w_p\}$  (accessible direction);  
 $\{d_l, d_m, d_n, d_p\}$  (accessible distance);
- 3) Corner (node type).

The HMM model based on the indoor map is defined as follows.

**Hidden states:** the location of the nodes in the map are the hidden states. The pedestrian is located at these nodes when the activities happens. It is a finite state set. Indoor map is used to estimate the transition matrix of the hidden states. The transition is uniform over the neighbor nodes. For example, the



given node contains  $n$  neighbors, then the transition probability to each neighbor is  $1/n$ .

**Observations:** The observations are the human activities which can be got from the activities label rules and the coordinate of the pedestrian which is calculated by the sensor data. The emission probability is defined as

$$p(O_t, m_t | S_i) = p(O_t | S_i) \cdot p(m_t | S_i) \quad (7)$$

The emission probability composes of the  $p(O_t | S_i)$  and  $p(m_t | S_i)$ .  $p(O_t | S_i)$  means the probability from the observation direction and distance to the current state. Then the observation probability is defined as

$$p(O_t | S_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(O_t - S_i)^2} \quad (8)$$

The  $\sigma$  is the standard deviation of the measured coordinate with the true coordinate.  $S_i$  is the  $i$ th neighbor node of the last detected node and  $O_t$  is the measured coordinate from the last detected node. For example, the last detected node is  $P\{x_{t-1}, y_{t-1}\}$  and  $S_i$  is one of the neighbor node. The PDR part gives the coordinate  $O_t$  which is calculated beginning with  $P$ . In the Eq.8,  $p(m_t | x_i)$  is the confusion matrix in the action detection which is given in the previous part.

If the start point is given, our method is effective. If the start point is unknown, we take all the nodes in the same node type as the initial point, and select the candidate from the neighbors, the algorithm is described in detail as the following.

The algorithm provide a possible solution for automatic startup positioning. Given a sequence of recent detection events, we can obtain the joint probability  $p_i$ . Each  $p_i$  corresponds to a sequence of the motion and we can calculate from the Markov process.

$$p_{t+1}(j) = p_t(i) \cdot a_{ij} \cdot b_j(O_{t+1}) \quad (9)$$

$p_t(i)$  is the probability of the sequence of the activity at time  $t$ .  $a_{ij}$  is the state transition probability from  $i$  to  $j$  and the  $b_j(O_{t+1})$  is the observation probability at state  $j$ . We pick the two highest matching probabilities, say  $p_{match}^1$  and  $p_{match}^2$  ( $p_{match}^1 \geq p_{match}^2$ ). Only when the  $p_{match}^1/p_{match}^2$ , called the distinguishing ration, exceeds a predefined threshold  $\eta$ , called the distinguishing threshold, the candidate node is taken as the current location.  $\Gamma$  represents the activity sequence and  $n_{landmark}$  represents the number of the landmarks in the floor. The detail of the algorithm is in Algorithm 1.

The algorithm which is used to judge whether the chain should be deleted shown as follows. It means that the coordinate of candidate point should be within the range of Eq 8. If the distance traveling after the last detected activity is not in the range of the  $3\sigma$  or the traveling direction is not in the current heading direction, the point chain is deleted.

## VI. EXPERIMENT AND RESULT

To evaluate the performance of the proposed map matching algorithm, we compared our method with the original PDR algorithm. The smartphone is xiaomi 4 which has various built-in sensors. The device is equipped with a MPU6500 accelerometer, a MPU6500 gyroscope sensor and a BOSCH barometer sensor. The frequencies are set to 50Hz for all sensors.

First, we conduct an experiment to verify the activity labeling algorithm. The volunteers walked in a long corridor. Three

---

### Algorithm 1 : The Activity Map Matching

---

**Definition:** AT=ActivityType

**Definition:** IPC=InitialPointCoordinate// given

**Definition:** CP=currentPoint; NP=neighborPoint;

**Input:** IRN = IndoorRoadNetworkoftheBuilding

**Input:** SD = SensorData

**Output:** The detected activity sequences: ActivitySequences

**Output:** The point chain of the activity: PointChain

---

```

1: for  $n \in \Gamma$  do
2:   Flag=0;//numberofNeighborsOutOfRange
3:   if  $n=1$  then
4:     currentPoint ← getInitialPoint(AT, IPC)
5:     addToThePointChain(CP)
6:   else
7:     if  $n_{activitySequence} > 0$  then
8:       for  $i = 1 : n_{PointChain}$  do
9:         if thePointOutOfRange(i, IRN, SD) then
10:          deleteThePointChain(i);
11:          Flag=Flag+1;
12:        end if
13:      end for
14:    end if
15:    if Flag =  $n_{chain}$  then
16:      deleteTheActivity
17:    else
18:      for  $i = 1 : n_{chain}$  do
19:        proMatrix = CalculatePro(NP, SD)
20:        CP = NP
21:        addToThePointChain(CP)
22:      end for
23:    end if
24:  end if
25: end for
    
```

---

### Algorithm 2 : thePointOutOfRange

---

**Input:** PointChain, IRN, SD

**Output:** chainState //if the PointChain of the range

```

1:  $\Delta = |theCoordinate - theCurrentCoordinate|$ 
2: if  $\Delta \cdot \sigma > 3\sigma \parallel CH \neq AD$  then
3:   chainState=1; //OutOfTheRange
4: else
5:   chainState=0;
6: end if
    
```

---

landmarks was set(two displays and one chair). The volunteers stopped before the displays and sat on the chairs. With the help of the sensors and the activity labeling algorithm, it is easy to label the activity. The result has been shown as follows in Figure. 6. AR and AL respectively represent action recognition and activity labeling.(W:walk,S:stand,T:turn,SIT:sit,SFD)

In the activity recognition part, the experiment is conducted in the five floor teaching building in Beijing University of Posts and Telecommunications. Data were collected by five volunteers with the phones on their hands. Each volunteer was asked to perform the seven activities. All individuals conducted the activities at the constant speed for a few minutes.

In another experiment, we performed our experiment in an office building to simulate the museum environment. The proposed system was implemented on the smartphone. The experiment was conducted with the smartphone on the hand.

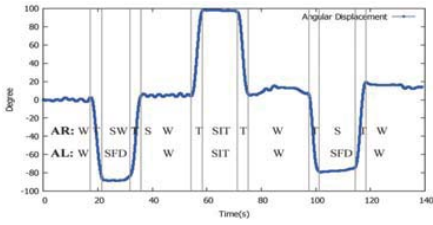


Fig. 6. Activity Labeling

It takes about 180s to finish this simulation. The simulation in the floor can be seen in Figure.7. The result of the simulation can be seen in Table.IV.

TABLE IV.  
ROUTE ESTIMATION RESULTS

Positioning Method	Avg(m)	Min(m)	Max(m)	< 1m(%)
Improved PDR only	3.31	0	7.04	43%
Improved PDR with map matching	1.27	0	2.65	69%

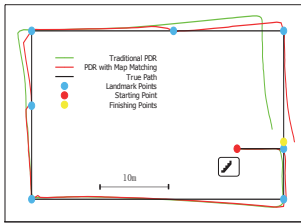


Fig. 7. Results of Routing Test in the Indoor Environment

VII. CONCLUSIONS

We have proposed an indoor position system using HMM theory and special activity matching to the special indoor location. The system use the sensor data and is independent from the other infrastructures. It is easy to implement and underspent. We improve the precision of the action recognition. In the activity part, we can avoid the random activity and false match. According to the experimental results and the comparison with the other indoor location methods, the proposed system perform well.

VIII. ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China(61671264 and 61671077).

REFERENCES

[1] A. Masiero, A. Guarnieri, F. Pirotti, and A. Vettore, "A particle filter for smartphone-based indoor pedestrian navigation," *Micromachines*, vol. 5, no. 4, pp. 1012–1033, 2014.  
 [2] K. Park, H. Shin, and H. Cha, "Smartphone-based pedestrian tracking in indoor corridor environments," *Personal and ubiquitous computing*, vol. 17, no. 2, pp. 359–370, 2013.

[3] X. Niu, M. Li, X. Cui, J. Liu, S. Liu, and K. R. Chowdhury, "Wtrack: Hmm-based walk pattern recognition and indoor pedestrian tracking using phone inertial sensors," *Personal and Ubiquitous Computing*, vol. 18, no. 8, pp. 1901–1915, 2014.  
 [4] C. K. Schindhelm, F. Gschwandtner, and M. Banholzer, "Usability of apple iphones for inertial navigation systems," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 1254–1258, IEEE, 2011.  
 [5] H. Liu, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Accurate wifi based localization for smartphones using peer assistance," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 10, pp. 2199–2214, 2014.  
 [6] Y. Kim, Y. Chon, and H. Cha, "Smartphone-based collaborative and autonomous radio fingerprinting," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 1, pp. 112–122, 2012.  
 [7] A. R. Jimenez, F. Zampella, and F. Seco, "Improving inertial pedestrian dead-reckoning by detecting unmodified switched-on lamps in buildings," *Sensors*, vol. 14, no. 1, pp. 731–769, 2014.  
 [8] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of wifi, smartphone sensors and landmarks using the kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, 2015.  
 [9] M. G. Puyol, D. Bobkov, P. Robertson, and T. Jost, "Pedestrian simultaneous localization and mapping in multistory buildings using inertial sensors," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 4, pp. 1714–1727, 2014.  
 [10] N. Mostofi, M. Elhabyby, and N. El-Sheimy, "Indoor localization and mapping using camera and inertial measurement unit (imu)," in *Position, Location and Navigation Symposium-PLANS 2014, 2014 IEEE/ION*, pp. 1329–1335, IEEE, 2014.  
 [11] M. Hardegger, D. Roggen, and G. Tröster, "3d actionslam: wearable person tracking in multi-floor environments," *Personal and Ubiquitous Computing*, vol. 19, no. 1, pp. 123–141, 2015.  
 [12] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, "Lightweight map matching for indoor localisation using conditional random fields," in *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on*, pp. 131–142, IEEE, 2014.  
 [13] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, "Indoor tracking using undirected graphical models,"  
 [14] B. Zhou, Q. Li, Q. Mao, W.-H. Tu, and X. Zhang, "Activity sequence-based indoor pedestrian localization using smartphones," 2014.  
 [15] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6067–6074, 2014.  
 [16] C. Huang, S. He, Z. Jiang, C. Li, Y. Wang, and X. Wang, "Indoor positioning system based on improved pdr and magnetic calibration using smartphone," in *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on*, pp. 2099–2103, IEEE, 2014.  
 [17] W. Wu, S. Dasgupta, E. E. Ramirez, C. Peterson, and G. J. Norman, "Classification accuracies of physical activities using smartphone motion sensors," *Journal of medical Internet research*, vol. 14, no. 5, p. e130, 2012.  
 [18] C. Torres-Huitzil and M. Nuno-Maganda, "Robust smartphone-based human activity recognition using a tri-axial accelerometer," in *Circuits & Systems (LASCAS), 2015 IEEE 6th Latin American Symposium on*, pp. 1–4, IEEE, 2015.  
 [19] M. Arif, M. Bilal, A. Kattan, and S. I. Ahamed, "Better physical activity classification using smartphone acceleration sensor," *Journal of medical systems*, vol. 38, no. 9, pp. 1–10, 2014.

# Personalized Learning in a Virtual Learning Environment Using Classification of Objective Distance

Sataworn Chaichumpa, Santichai Wicha and Punnarumol Temdee

School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand

E-mail: 5771501004@lamduan.mfu.ac.th, santichai@mfu.ac.th and punnarumol@mfu.ac.th

**Abstract**—Personalized learning helps students to satisfy their own learning. Although the overall individual performance is generally used to satisfy the course objective, to pass or to fail, all learning objectives in the course may not be satisfied. To promote successful personalized learning, this paper proposes the objective distance which is used for navigating appropriate learning objects to the students individually. Instead of selecting among all available learning objects, the selection with objective distance is done only among possible learning objects to ensure individual and course satisfaction. In this paper, the objective distance is constructed and evaluated to study whether or not it can be used for the assessment of student's performance. The study is conducted with 55 students at Mae Fah Luang University studying in a special learning module through a virtual learning environment. These students are free to select learning object and always have a performance evaluation at the end of their studies. The classification methods used for evaluating the proposed objective distance are K-Nearest Neighbor (K-NN) and Artificial Neural Network (ANN). The classification results show that the proposed objective distance can be used for personalized assessment. At the same time, K-NN and ANN provide 87.33% and 92.33% classification accuracy respectively.

**Keywords**—Personalized Learning, Classification, Objective Distance, K-Nearest Neighbor, Artificial Neural Network

## I. INTRODUCTION

Personalized learning has become very popular in the education area because it is based on students' attentiveness and their abilities. The diversity of each student [1-2] creates a sense of awareness of their real abilities, skills, and requirements so that the information can be utilized for designing curriculum or providing guidance in the right path to ensure learning efficiency [3-4]. There are many research works that have studied about personalized learning such as the development of English language recommendation systems for students to have the right instrument for their profile, thus to help to improving their English proficiency and to motivate them in language learning [5]. The category of the education goal is to deliver and create knowledge through enhanced learning [6-7]. Defending the learning objective is a very important feat in teaching. It is the focus of basic education that provides the expectation of teaching, learning, and evaluation [8]. The clear objective definition of the content and each unit score of the students will inform the real ability of each student.

This paper proposes the measurement for classifying the abilities of students from the distance of the objective in virtual learning environment. The proposed distance is the calculation of the current students' abilities with the expected competency that is specified in each objective. There are totally 6 objectives

used in this paper. The positive classifier result is shown in this paper.

## II. RELATED WORKS

Virtual Learning Environment (VLE) [9] is the environmental integration for students to study through the internet technology. It provides the space to enable the participation among teachers and classmates with the support of personalized learning. VLE provides the tools to work for the administrators, teachers and students. VLE provides the course for the students to register, examines their grade report, and creates a report. VLE also acts as the platform for students to attend the lecture and access into the educational material for the learning content. The student will get the test with limited time, whereas the question and the answer will generally be automatically sent to the teacher for making feedback and comment and return to the students. Every VLE provide communication tools such as e-mail, chat, and multimedia conference, etc. Moreover, it provides some equipment and recommendation for the teachers to develop the material for the curriculum that will also be presented through VLE [9]. Nowadays, computer technology has been highly developed. It can support from K12 education to lifelong learning [10].

Learning with e-learning is one of the famous virtual learning environments [11-12] in the current time. It requires some new ideas to learn in e-learning. Such ideas can bring critical techniques to the classroom by emphasizing on higher skills, so students will quickly get feedback to solve problems [13]. By managing the curriculum for the students in the right way through considering the levels of their knowledge, the student can be encouraged to learn individually [14-15]. For traditional learning, personalize learning is the process that the teachers will enhance the students to learn what they need according to their strength and their requirement. Then the teachers will provide the curriculum depend on the students' interesting and the academic standard [16-18]. The personalized learning focuses on the learning environment that the students will be the leader and the teachers are as the facilitator. The students' abilities will be adjusted to the learning details [19-20], whereas the suggested learning levels are all about the students' abilities [21]. For VLE, the personalized English learning recommendation system have been developed to enhance the students' abilities to read the interest subjects in various ways. Also, this system makes the students to be engaged with the learning system. Through analysis of the detail that follows the co-operative working model and data mining, the recommendation is based on the selection of lessons that can be matched with the students'

skills and abilities [22]. Most researchers use the final score to define the student performance [23] or course objective. For this paper, the score on the students' ability of each objective is used to measure the distance between the current statuses of student's ability to the expected course objectives. It is classified to define their real individual performance. It is thus called objective distance. The detail of how to measure objective distance is shown in the next section.

There are many research works applying Artificial Neural Network (ANN) to classify the problems in different way, such as in the industrial field, financial business and banking, and for operational diagnosis [24]. Especially in education, this method has been applied for a long time. Moreover, it has also been applied to predict the students' learning success [23, 25] and to predict the students' GPA in their curriculum [26-27], as well as to predict the applicants' potential for being accepted in the educational institution [28].

K-Nearest Neighbor (K-NN) is also widely applied in the education domain [29]. Many studies focus on the analysis of the students' learning outcome according to the curriculum – to pass or fail [30]. Similar studies try to estimate the final test of the students who registered in the Web-Based curriculum [31]. By showing the students' learning outcome including weak, good and very good, K-NN can evaluate the learner progression in the University [23].

From the related works, this study focuses on how to classify the proficiency of learning skills in virtual learning environment by classifying the proposed objective distance with different classification methods including K-NN and ANN.

III. RESEARCH METHODOLOGY

Fig. 1 shows the research methodology. This paper presents the Experimentation and Data Gathering, Model Constructing and Model Testing and validating respectively.

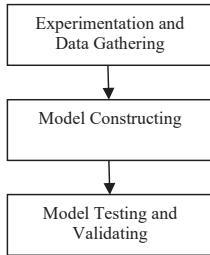


Fig. 1. Research Methodology

A. Experimentation and Data Gathering

This paper conducts an experiment with 55 students who are first year students from School of Information Technology, Mae Fah Luang University, Thailand. A calculation of the distance of students' abilities from their each objective score with the desirable of the objectives is shown in (1), (2), (3).

$$Mob_{sc} = \sqrt{(Mob_j - Mob_{jg})^2} \tag{1}$$

$$STobj_t = \sqrt{(Mob_j - STobj_j)^2} \tag{2}$$

$$Mob_{sc} = Mob_{jg} - STobj_t \tag{3}$$

Where,  $Mob_{sc}$  is the distance between total score of each objective and the goal satisfaction score,  $Mob_j$  is the total score of each objective,  $Mob_{jg}$  is the goal satisfaction score,  $STobj_j$  is the students' real score of each objective,  $STobj_{dt}$  is the distance between each objective of each student  $STobj_{dt}$  is the difference between students' real score to the goal satisfaction score.

The pre-defined total score and satisfaction score of each objective used in this paper is shown in Table I. The satisfaction score means the least score to satisfy each individual objective. The satisfaction portion is pre-defined by the instructor following the objectives to clarify the minimum scores required to accomplish each objective.

TABLE I. SATISFACTION PORTIONS FOR EACH LEARNING OBJECTIVE

Objective Score	Application Using	Document Creating	Formatting	Object Manipulating	Mail Merging	Output Preparing
Score (100 %)	15%	20%	27%	25%	5%	8%
Satisfaction Score	11.25%	15%	20.25%	18.75%	3.75%	6%

The study is done to find the objective distance of each student to each objective. For example, the student No. 1 has total score which is equal to 9 and the satisfactory score which is equal to 7. By calculating from 75% of satisfaction, if that student receives 6 scores, the distance value will be -1. Some examples of objective distances from some students are presented in Table II.

TABLE II. OBJECTIVE DISTANCES

Student	Using application	Objective Distance	Document Creation	Objective Distance	Formatting	Objective Distance	Objects	Objective Distance	.....	Classified Group
	9	2(7)	12	3(9)	16	4(12)	15	4(11)	.	..
1	6	-1	8	-1	11	-1	8	-3	.	satisfactory
2	5	-2	6	-3	7	-5	9	-2	.	Ordinary
3	8	1	10	1	15	3	12	1	.	satisfactory
8	3	-4	1	-8	6	-6	1	-10	.	Unsatisfactory
5	9	2	12	3	15	3	13	2	.	satisfactory
..	..	..	..	..	..	..	..	..	.	..

B. Model Constructing

After obtaining objective distances of each student, they are modeled by using K-NN and ANN to classify into three groups of students including satisfactory, ordinary and unsatisfactory.

1) K-NN Classification Model

K-NN has been generally applied in many applications of supervised learning [23]. The Euclidean Distance is also generally used to measure the distance between known and unknown examples. Therefore, this paper employs Euclidean Distance to measure objective distance of known and unknown student as shown in equation (4).

$$D_s = \sqrt{\frac{(STobj\ 1_{Tl} - STobj\ 1_{Tr})^2 + (STobj\ 2_{Tl} - STobj\ 2_{Tr})^2 + (STobj\ 3_{Tl} - STobj\ 3_{Tr})^2 + (STobj\ 4_{Tl} - STobj\ 4_{Tr})^2 + (STobj\ 5_{Tl} - STobj\ 5_{Tr})^2 + (STobj\ 6_{Tl} - STobj\ 6_{Tr})^2}{2}} \quad (4)$$

Where  $D_s$  is the objective distances between 2 groups of students.  $STobj$  from 6 distances are used. The distance between each objective of each student,  $T_l$  and  $T_r$  are the distance from known group and unknown group respectively. The data will be measured against 3 groups with different groups to the nearest will be chosen.

2) ANN Classification Model

ANN [32] is used to create the model in mathematics for copying the brain system. In general, ANN has been copied when some inputs have been inside. The inputs will be multiplied with their associated weight that is replaced the significance for input. Every node that had been proceeded to become motivated by the activate function that has an output from equation (5) are applied. The concept of neural network is shown Fig 2. In this paper, there are 6 distances as the inputs and 3 classes of student's groups as the outputs. Additionally, Backpropagation neural network is used in this paper.

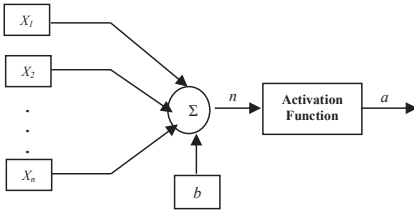


Fig. 2. Neural Network Concept

$$n = \sum_{i=1}^z x_i w_i + b \quad (5)$$

Where,  $n$  is the output of the sum function,  $x_i$  is the input  $i$ ,  $w_i$  is associated weight of input  $i$ ,  $z$  is numbers of total inputs and  $b$  is the biasing weight.

The Backpropagation process comprises two parts - forward pass and backward parts. The forward part is the passing data into the input layer through the output layer. The

backward part adjusts the weight values to be matched with the error correction that is the difference between the actual response and the target response. It is the error signal that is passed to the neural network in a different way of the connection.

IV. RESULTS AND DISCUSSION

This section is to evaluate the accuracy of classification models comparing between two classifiers.

A. Classification Ratio

The comparison of each group of student classified by K-NN and ANN is shown in Table III. The result shows that ANN has obtained more classification accuracy comparing to that is obtained by K-NN. However, the proposed objective distance is suitable for being used to classify the students' proficiency. For future work, more studies with many different contents will be conducted to verify the classification accuracy. Additionally, an effort should be made to bring the result of personnel abilities that have been classified into the different objectives to make the curriculum more suitable for the students to learn.

TABLE III. CLASSIFICATION RATIO

Type of Skill (A,B,C)	Pre-defined Classification result	K-NN		ANN	
		Classification result	% of accuracy	Classification result	% of accuracy
Unsatisfactory(A)	22	22	100%	22	100%
Ordinary(B)	18	11	61.11%	14	77.78%
Satisfactory(C)	15	15	100%	15	100%
<b>Total</b>	<b>55</b>	<b>48</b>	<b>87.04%</b>	<b>51</b>	<b>92.59%</b>

V. CONCLUSION

This paper has proposed the objective distance for the classification of student's performance by using K-NN and ANN. The objective distance is measured from the current performance status of each student to the expected performance accordingly to the course objective. For this paper, the study is conducted with 55 students from School of Information Technology, Mae Fah Luang University, Thailand studying in particular online course. The students' performance has been classified into three different categories including Unsatisfactory (A), Ordinary (B), and Satisfactory (C). K-NN and ANN are used as the classifiers in this paper. The classification results from both K-NN and ANN have suggested that the proposed objective distance can be used effectively for classifying students' performance of online learning in VLE. Additionally, K-NN has obtained 87.04% accuracy and ANN has obtained 92.59% respectively.

## REFERENCES

- [1] M. Alian, R. Jabri, "A Shortest Adaptive Learning Path in eLearning Systems:Mathematical View," Isra Private University, Amman, Jordan, Journal of American Science, pp. 32-42, 2009.
- [2] S. Graf, "Adaptivity in learning management systems focusing on learning styles," Vienna University of Technology, Austria, 2007.
- [3] A. Dwi C, A. Basuki, "Personalized Learning Path of a Web-based Learning System", Informatics Engineering Trunojoyo University Indonesia, International Journal of Computer Applications (0975 – 8887) vol. 53, no. 7, 2012.
- [4] I. A. Alshalabi, S. Hamada, K. Elleithy, "Automated Adaptive Learning using Smart Shortest Path Algorithm for Course Units," Department of Computer Science and Engineering University of Bridgeport Bridgeport, CT, USA, pp. 1-5, 2015.
- [5] G. Hwang, H. Sung, C. Hung and I. Huang. "A Learning Style Perspective to Investigate the Necessity of Developing Adaptive Learning Systems," Educational Technology & Society, vol.16, no. 2, pp. 188-97, 2013.
- [6] BS. Bloom, "Taxonomy of educational objectives: The Classification of Educational Goals.Handbook 1: Cognitive Domain. 2nd ed. Boston: Addison Wesley; 1984.
- [7] J. Conklin, LW. Anderson, D. Krathwohl, P. Airasian, KA Cruikshank, RE. Mayer, P. Pintrich, J. Raths and MC. Wittrock, "A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives Complete Edition," 2005.
- [8] R.G. Carroll, "Design and evaluation of a national set of learning objectives: the medical physiology learning objectives project," Advances in physiology education, vol 25, no.2 ,pp. 2-7, 2001.
- [9] A. Kumar, R. Pakala, RK Ragade and JP Wong, "The virtual learning environment system," InFrontiers in Education Conference, FIE'98. 28th Annual 1998, vol. 2, pp. 711-716, IEEE, 1998.
- [10] A. Tucker, "A Model Curriculum for K-12 Computer Science: Final Report of the ACM K-12 Task Force Curriculum Committee," 2003.
- [11] P. Brusilovsky, J. Eklund and E. Schwarz, "Web-based education for all: a tool for development adaptive courseware," Computer networks and ISDN systems, vol. 30, no. 1, pp. 291-300, 1998.
- [12] T. Murray, S. Blessing, and S. Ainsworth. "Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive and intelligent educational software," Springer Science & Business Media, 2003.
- [13] M. Vargas-Vera, M. Lytras. "Personalized learning using ontologies and semantic web technologies." World Summit on Knowledge Society. Springer Berlin Heidelberg, 2008.
- [14] M. Yarandi, H. Jahankhani and AR. Tawil, "A personalized adaptive e-learning approach based on semantic web technology," Webology, vol. 10, no. 2, 2013.
- [15] S. Gauch, M. Speretta, A. Chandramouli and A. Micarelli, "User profiles for personalized information access," InThe adaptive web, Springer Berlin Heidelberg, pp. 54-89, 2007.
- [16] Basham, D. James , E. Tracey. Hall, A. Richard. Carter, and M. William. Stahl. "An Operationalized Understanding of Personalized Learning." Journal of Special Education Technology 31, no. 3, pp. 126-136., 2016.
- [17] CM. Chen, H.M. Lee, and Chen, Y.H., "Personalized e-learning system using item response theory". Computers & Education, 44(3), pp.237-255., 2005.
- [18] C. Chiu-Jung. and L.Pei-Lin., "Personalized computer-assisted mathematics problem-solving program and its impact on Taiwanese students". The Journal of Computers in Mathematics and Science Teaching, 26(2), p.105., 2007.
- [19] E. Sangineto, "An Adaptive E-Learning Platform for Personalized Course Generation," Architecture Solutions for E-Learning Systems, pp. 262-281, 2008.
- [20] CM. Chen, "Intelligent web-based learning system with personalized learning path guidance," Computers & Education, vol. 51, no. 2, pp. 787-814, 2008.
- [21] CM. Chen, C.J. Chung, "Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle," Computers & Education ,vol.51, no. 2, pp. 624-647, 2008.
- [22] MH. Hsu, "A personalized English learning recommender system for ESL students," Expert Systems with Applications, vol.34, no. 1, pp. 683-688, 2008.
- [23] AM. Shahiri, W.A Husain, "Review on Predicting Student's Performance Using Data Mining Techniques," Procedia Computer Science, vol. 72, pp. 414-422, 2015.
- [24] GP. Zhang, "Neural networks for classification: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30 ,no. 4, pp. 451-62, 2000.
- [25] BC. Hardgrave, RL. Wilson and KA. Walstrom, "Predicting graduate student success: A comparison of neural networks and traditional techniques," Computers & Operations Research, vol. 21 Number 3, pp. 249-63, 1994.
- [26] M. Paliwal, U.A. Kumar, "A study of academic performance of business school graduates using neural network and statistical techniques," Expert Systems with Applications, vol. 36, no. 4, pp. 7865-72, 2009.
- [27] TA. Rashid, NK. Aziz, "Student Academic Performance Using Artificial Intelligence," ZANCO Journal of Pure and Applied Sciences, vol.28, no. 2, 2016.
- [28] VO. Oladokun, AT Adebajo and OE. Charles-Owaba, "Predicting students' academic performance using artificial neural network: A case study of an engineering course," The Pacific Journal of Science and Technology, vol. 9, no. 1, pp.72-79, 2008.
- [29] JE. MacLeod, A. Luk and DM. Titterington, "A re-examination of the distance-weighted k-nearest neighbor classification rule," IEEE Transactions on Systems, Man, and Cybernetics, vol. 17, no. 4, pp. 689-696, 1987.
- [30] T. Tanner, H. Toivonen, "Predicting and preventing student failure- using the k-nearest neighbour method to predict student performance in an online course environment," International Journal of Learning Technology, vol. 5, no. 4, pp. 356-377, 2010.
- [31] B. Minaei-Bidgoli, DA. Kashy, G. Kortemeyer and WF. Punch, "Predicting student performance: an application of data mining methods with an educational web-based system," InFrontiers in education, FIE 2003 33rd annual, vol. 1, pp. T2A-13, IEEE , 2003.
- [32] M. Kubat, "Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.", pp. 409-412, 1999.

# Maritime Wireless Multimedia Communication System Based on Software Defined Radio

Zhenjia Chen<sup>1,2</sup>, Yonghui Zhang<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, Haikou 570228, China

<sup>2</sup>College of Information Science & Technology, Hainan University, Haikou 570228, China

E-mail: zhyhemail@163.com

**Abstract**—This paper proposes an long distance maritime wireless multimedia communication system between ship and ship, ship and land. The system transmits data through TV White Spectrum (TVWS) which has many characteristics of long range propagation, not be absorbed easily by the water, carrying large amount of data and so on. Using TVWS communicate improves the communication bandwidth for the maritime users. Software Defined Radio (SDR) is used to switch communication parameters to optimize the quality of wireless communication and reduces communication channel jam phenomenon to appear. The system in this article realizes multimedia communication like audio communication, video communication, network monitoring between the ship and coastal, deep sea fleet. It is a kind of maritime wireless multimedia communication solution which is convenient, cheap and effective.

**Keywords**— TVWS; SDR; end-to-end; Multicast; Socket

## I. INTRODUCTION

At present, the offshore operation is more and more frequently. Deep sea fishing, submarine energy exploitation, marine biology research, ecological environment protection and other activities are increasing. Maritime transportation is increasingly busy and human activities have expanded to the ocean. People's demand for communication is no longer entirely confined to simple voice communication. During business growing, large amount of wireless data transmission is put forward. Compared with land and coastal activity environment, maritime communication is complicated. It is more complex and has high operation risk. So it is especially important to guarantee the reliability and the timeliness of marine communications and the safety of offshore operation. Nowadays, offshore vessels are equipped with the basic satellite communications equipment such as satellite phones. This way implements deep sea communication between ship and offshore station, but the price of equipment and the working cost is expensive. The data transmission of satellite communication has high delay. In addition, coastal ships use ultrashort wave radio or short wave radio communication. Those method unable to realize flexible networking and only suitable for voice transmission. The maritime user urgently need flexible and convenient multimedia communication.

Nowadays, army is the main user of the maritime network deployment. Maritime wireless networks are also applicable to other maritime users, such as the maritime mesh network based on satellite which applied to commercial fishing, emergency rescue, or coast guard [1]. Heterogeneous Wireless Network (HWN) infrastructure integrates maritime wireless network and local wireline network. It consists of maritime wireless mesh

network, satellite mobile network and Internet. This system is put forward for ship-to-shore communication. Mobile terminal access to the wireless mesh network, then through the shipborne satellite gateway access satellite mobile network [2]. In order to integrate heterogeneous wireless and wireline network, people suggest that ship-to-ship and ship-to-shore through VoIP calls traverses different network path. Adopting dual IP address and dual SIP to manage mobile users and VoIP session management, [3] chooses the best path to decrease satellite link. It reduces the data transmission delay and the cost of data interaction.

In this paper, the system uses TVWS as the communication frequency band between multiple maritime communication access nodes (AP). It uses the common WIFI frequency (mainly for 2.4GHz) as access frequency for general mobile terminals and IP Camera and establishes the maritime HWN. The system without server to save communication data and terminals transmit data directly by end-to-end packets. The system is suitable for use on sea with flexible networking. Every AP can acts as a gateway connecting mobile satellite network and reduces the complexity of the satellite mobile network access. The system uses SDR as long-distance RF communication module between AP and AP. SDR can realizes frequency bandwidth adjust, modulation mode setup, system programming, source code control. It can realizes efficient frequency utilization, and reduces the situation of spectrum occupy or signal interference.

Spectrum is an important strategic for many countries. It is a scarce, non-renewable resources. However, the contradiction between the accelerated development of information technology and the scarcity of spectrum is increasing day by day. At the same time, it is also has the problem that spectrum resource utilization efficiency is low. Maritime communication has the requirement of long-distance communication, light rain fade. TVWS can meet the requirement of maritime traffic. The traditional radio communication system is consist of antenna, modulation module, detector, amplifier, filter, mixer. Most of these modules are usually realize in hardware. In addition to the antenna, SDR can implements the function of these components by software. Based on its reconfigurability and flexibility, it has been paid great attention in the army and disaster early warning application. The channel of wireless communication will be taken up in accidentally or intentionally, SDR can adjust those parameters to maintain the basic communication[4].

The paper is organized as follows. Section II presents the structure of AP description. SDR realizes virtual Ethernet

interface for maritime communication nodes. The virtual Ethernet interface is implemented by processing unit through the virtual network devices top/tap. Section III describes the system network based on SDR in detail. Section IV focuses on the end-to-end communication process on the basis of OSI communication protocol using socket communication protocol between terminal and terminal. We conclude in Section V with a description of future work.

II. COMMUNICATION ACCESS NODE

As shown in Fig. 1. APs communicate using TVWS and SDR as the communication interface. AP has routing capabilities and transfers data between communication interfaces. There are three main communication interface: WIFI interface, Ethernet interface, SDR interface. WIFI interface provides access for the mobile terminal; Ethernet interface provides access for the IP Camera, to provide security monitoring; SDR interface for data transfer between communication nodes bridge, and can realizes frequency point flexible switch in white spectrum, and avoid channel blockage. AP assigns a unique IP address for each communication interface. The processing unit realizes the packet forwarding between three communication interface by maintaining the routing table.

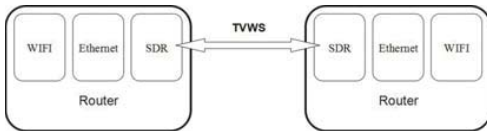


Fig. 1. Structure diagram of communicate between APs

SDR is based on the software programming and control to realize wireless communication platform. Its system hardware structure and function are relatively independent. Its broadband AD converter and DA converter as far as possible close to the RF module, and is an open source, modular, standardization of general hardware platform. SDR defines a series of software and hardware technology. It modifies the software or a programmable firmware, to change the function of the radio [5]. According to the principle of modular design, SDR has the open OSI architecture, the function of good programmability, software portability, support broadband, high rate, multiple modal wireless communication. SDR as the RF module of the system, have the data link layer and physical layer with controlled. Processing unit abstracts network layer, and realizes standard IP data packets.

The system realizes virtual Ethernet interface through the TUN/TAP. It realizes the function of the virtual network interface card, binding hardware of SDR and local interface of processing unit together. The TAP is equivalent to Ethernet device, and operates Ethernet data frame from layer two. TUN simulates network layer equipment, and operates IP packets from layer three. TUN/TAP provides packet receive and transmit server for the user space process. The OS kernel looks it as Ethernet equipment. It transmits and receives Ethernet frame to the user space process. User space processes can be regarded the IP Client, send IP packet to TUN/TAP device for

Ethernet frame encapsulation [6]. The structure of the virtual interface is shown in Fig. 2.

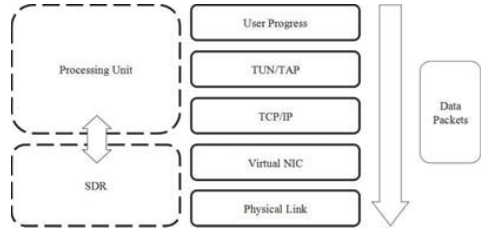


Fig. 2. Structure of virtual interface

Processing unit is a Linux operating system of embedded system with the TUN/TAP drive. It can realizes data interaction through create special socket as the channel, and uses the TUN/TAP as interface. TUN/TAP drivers is using a device file data interaction of user mode and kernel mode. Processing unit transfer data to physical link through TUN/TAP driver, and transmit network grouping package by TCP/IP protocol stack.

Processing unit through local interfaces transmit script to the SDR module. It assigns IP address for the virtual Ethernet interface and updates the local routing table. The protocol data unit (PDU) from TUN/TAP is the data packet in third OSI layer. The SDR module receives the PDU, and converts into data stream through the code package encoder. Then, the stream as base band signal be transmitted to the GFSK modulation modulator. After resampling, the signal that more flexible. The base band signal is set to the carrier frequency, and achieves spectrum shift. Finally, the antenna radiates the electromagnetic signal. The receiver and transmitter are similar, but the receiver add received signal filtering to remove noise. Packet sending and receiving process is shown in Fig. 3.

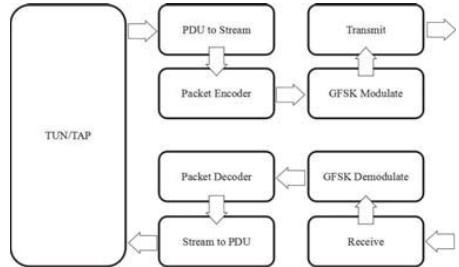


Fig. 3. The flow chart of data transceiver

RF module selects TVWS as the communication frequency between APs. Some countries are developing and utilizing this frequency actively. At present, the international has generally released ground analog TV signal. They distributed frequency (700MHz) to 4G. And the local spectrum cleaning work is being actively conducted by the State Administration of Radio Film and Television (SARFT). It would have very extensively applied prospect in the future. Through the spectrum monitoring activity, we found that the local 700MHz band had not been used. The experiment selected 780MHz frequency



point as a carrier frequency of the RF module. We used professional spectrum analyzer on the frequency spectrum monitoring. Its nearby original spectrum is shown in Fig. 4.

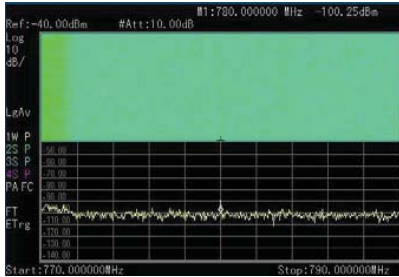


Fig. 4. The original spectrum of 780MHz

Processing units mounted and initialized the SDR. It send scripts to the virtual Ethernet interface of SDR, set 780MHz as baseband signal carrier frequency and startup RF module. As shown in Fig. 5, for the initialization had completed, the spectrum without data.

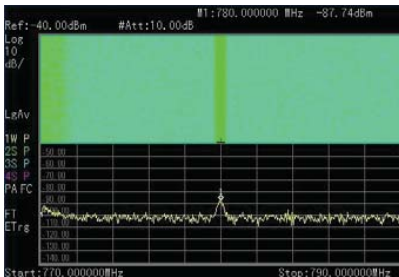


Fig. 5. The spectrum of the initialization has completed

After equipment stability, processing unit send control script to SDR. We send ARP broadcast continuously. Data packets were the Ethernet broadcast packets containing local IP address. The continuous pulse signal can be measured by spectrum analyzer. As shown in Fig. 6, the terminal send ARP broadcast continuously. The RF module emitted signal without causing interference to other spectrum signal nearby.

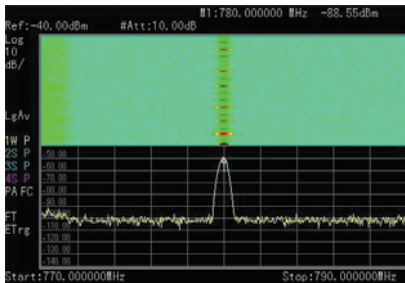


Fig. 6. The spectrum of send ARP broadcast

### III. SYSTEM NETWORK

Maritime communication network mode need to be flexible, and be as the basis of IP network. The wireless equipment between ships can set up their own wireless local area network (LAN), which is simple and convenient communication network. Maritime WLAN also requires for network has good reliability and stability, and has the ability to self-management. As shown in Fig. 7, as mentioned in this article maritime instant communication system network diagram. AP forms the main framework of the wireless mesh network, mainly to achieves two different frequency band network data routing. Every AP in the network could be used as mobile satellite network access gateway. As long as there is access to a AP, other nodes by connecting to the network can access the Internet. This way can greatly reduces the complexity of the access to the Internet and the cost of maritime user to access the Internet.

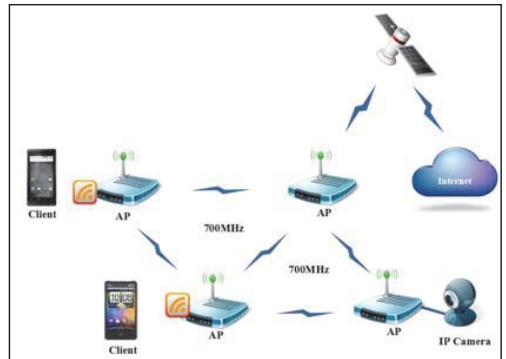


Fig. 7. Maritime instant communication system network diagram

AP realizes data exchange between 2.4GHz spectrum and TVWS. It has function of routing in communication network. 2.4GHz spectrum is ISM frequency band spectrum. Mobile terminal can frequently access to this spectrum, and its related applications and platform has highly plentiful. The system mainly considering the compatibility with existing equipment. By that can reduces the development and the application cost. TVWS available width is different in each place. In the sea system uses SDR to sense spectrum environment, and selects suitable frequency band to form communication network. When the distance between two APs are beyond the signal coverage, APs automatically choose the intermediate node as a relay. Through multiple hops to establish long distance point-to-point wireless communication network.

### IV. TERMINAL COMMUNICATION

In order to avoid the maritime server building and information transit storage, adapt to the liquidity characteristic of maritime instant messenger communication, the system use socket communication protocols as the bridge realizes end-to-end communication directly between mobile terminals. Terminal through the multicast technology to send broadcast in the network. Other terminal in this network receive packets and parsing out the IP address and the user information. Terminal

establishes the connection through the IP address and service port. Terminal of the end-to-end communication process as shown in Fig. 8.

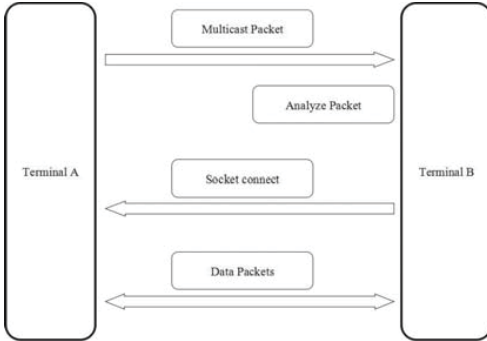


Fig. 8. End-to-end communication process between terminal and terminal

**A. Multicast**

Multicast implements broadcast form of one transmitting to multiple receiving, one-way network connection. The sender set up the target set of host addresses, and send broadcast packets in the RF channel. All the host of the entire group can receive the datagram. Multicast technology can effectively solves the problem of single point send multiple point receive mode, and realizes efficient data transfer by one-to-multiple point mode. It saves vast network bandwidth, and reduces the network load, and improves the efficiency of data transmission, and reduces the possibility of backbone network congestion.

**B. Socket Connection**

The socket is used to describe the IP address and port, is a communication chain handle. It can be used to realize communication between different virtual machine or computer. In order to be able to communicate application layer and network layer, system uses socket communication protocol which is the software abstraction layer of TCP/IP protocol communication. In the design model, the socket is a facade pattern. It hides the complicated TCP/IP protocol family behind. For users, a set of simple interface can easily organize data, in order to conform to the specified protocol.

The socket is divided into client and server. The terminal access to AP at the beginning as the server side. It start to open the port to monitor local related services. The process is in a state of waiting for the connection at this time, in real-time monitoring network state. The terminal request connection as the client side. Its target is the server socket. Client first extracted address and port number of the server socket, and then to the server socket connection requests are put forward. The server socket to receive the client's connection request, and waits for the connection state conversion to respond to the request. The server send the description of socket to the client. After the client confirms description, connection is established.

**C. Data Communication**

In the system, AP distributes fixed IP address for each terminal. AP in this pattern can only positioning mobile terminals. Cross between AP, they are as a routing transit through communication. Terminal uses socket communication protocol to realize the end-to-end communication. The system supports characters, pictures, audio and video communication between mobile terminals. Before sending a variety of services packets, terminal transforms data into the data flow, and writes the data flow in the pipe. The data receiving terminal according to the data flow service port to receive various services. And the received data stream according to different types of data format, it is transformed into the corresponding data information. Similar to multimedia communication between terminals, IP Camera access to AP, namely the released its own unique IP address to the network. The mobile terminal in the network can access through TCP packet network security cameras. Fig. 9 for the mobile terminal network monitoring diagram.



Fig. 9. The mobile terminal network monitoring diagram

As is known to all, when at sea is without any reference to subjectively to observe their location, only by latitude and longitude to locate. Nowadays, the majority of ships are equipped with basic theodolite to record their own location. The military ships are using radar scanning system for ships near the trend. But civilian vessels can only through the outlook and the wireless interphone learned that the location of the other ships, to realize the coordination and cooperation between the deep-sea fishing fleet. Ship records the sail set route according to the situation of location. It hard to avoid emergency in sudden incidents or night sailing. Because can't directly observed near the location of the ship, for maritime navigation is inconvenient. In order to solve this problem, the system in this paper join the geographical position information (GPI) in the process of communication, and realizes ship position share and navigation path set in the communication network, and other functions. The data frame format with GPI is shown in Fig. 10.

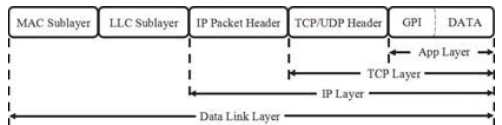


Fig. 10. The data frame format with GPI

## V. CONCLUSION

With the shortage of spectrum resources, using spectrum resource efficiently becomes more and more important. Digital radio and television, provide a more free frequencies. The application of TVWS is increasing, and also is faced with various challenges in technology. Using white spectrum is emerging immeasurable commercial value. At present, the bottleneck of its development lies in the lack of real network model, and be used in specific communication applications rarely. The spectrum of the maritime environment is more simple. There has more available spectrum resource. Marine geographic environment, however, is much more complicated than the land. Maritime user is a very special target group. This article put forward this communication system according to the characteristics of the instant messaging for maritime.

This system highlights the high software portability and networking of simplicity. It uses TVWS to realize long distance communication, and provides high communication bandwidth to meet user demand trend of multimedia communication at sea. The system adjusts communication frequency on the basis of demand through SDR. It can reduces the channel interference, and uses white spectrum more effectively. Using low-cost communication structures builds self-organizing network, the node equipment expands communication network through the signal multiple jump forward. Terminal accesses to AP and communicates to others through end-to-end communications link directly. The software is compatible with the existing mobile terminal equipment, and implements information exchange between every two users in the network. It can quickly implements long-distance multimedia communication

at sea. Beyond the maritime on the communication, the system also can be used in remote wireless broadband connection communication technology to provide a relatively low cost solution. In the open desert or in places such as grassland, this system also has the certain practical value.

## ACKNOWLEDGMENT

In this paper, the research was sponsored by the National Natural Science Foundation of China (Grant No. 61561018), the Application Technology Research and Development and Demonstration of Hainan Province(Grant No. ZDXM2015104).

The corresponding author is Yonghui Zhang.

## REFERENCES

- [1] M. Manoufali, H. Alshaer, P.-Y. Kong, and S. Jima, "Technologies and networks supporting maritime wireless mesh communications," in 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC), pp. 1–8, Apr. 2013.
- [2] Wencai Du, Zhengxin Ma, Yong Bai, Chong Shen, Youling Zhou, Baodan Chen "Integrated Wireless Networking Architecture for Maritime Communications," pp.134-138, SNPD 2010.
- [3] Yong Bai, Wencai Du, "VoIP Services for Ocean Fishery Vessels over Integrated Wireless and Wireline Networks," pp.3461-3465, PIMRC 2013.
- [4] Nguyen B.Truong, Young-Joo Suh, Chansu Yu, "Latency Analysis in GNU Radio/USRP-based Software Radio Platforms," pp.305-310, MILCOM 2013.
- [5] Rajeshree D. Raut and Dr. Kishore D. Kulat, "SDR Design for Cognitive Radio" IEEE 4th International Conference on Modeling, Simulation and Applied Optimization, 2011, pp 1-8.
- [6] KALLGREN(D.G.), SMAAL(J.G.), IP unicast/multicast operation over stanag 5066, IEEE Milcom, 2001.

# Characterizing the Service Usage of Online Video Sharing System: Uploading v.s. Playback

Chenyu Li, Jun Liu, Shuxin Ouyang

Beijing Key Laboratory of Network System Architecture and Convergence  
 School of Information and Communication Engineering  
 Beijing University of Posts and Telecommunications, Beijing, China  
 Email: lichenyu@126.com, liujun@bupt.edu.cn, ouyangshuxin@gmail.com

**Abstract**—Online video sharing system is becoming increasingly prevalent among the general public recently. Understanding its service usage is crucial for allocating the network resources and adjusting the service design. In this paper, we focus on a leading online video sharing system in China, namely Youku. We collect over 12 billion traffic traces from the network operator and crawl video meta-data for 30 consecutive days from Youku. Based on those data, we present the first comparative characterization work on two key usages of online video sharing system: uploading and playback. We examine the different temporal patterns and intensities of user activity between uploaders and viewers. We also compare the properties of the uploaded videos and the watched videos in Youku, and reveal various user preferences behind those two service usages. Eventually, we study the popularity dynamics in terms of view count distribution and active period for both video uploading and video playback. Our analysis results can be utilized by network operators, service providers, online advertisers and content creators, to obtain potential revenues, improve service performance and provide a better user experience.

## I. INTRODUCTION

Online video sharing system generates massive Internet traffic and consumes a large part of network bandwidth. As the white paper of Cisco Systems notes, the global video traffic accounted for 64% of all consumer Internet traffic in 2014, and will be up to 80% in 2019 [1]. A better understanding of the service usage of large online video sharing systems can provide direct help to network operators, service providers, online advertisers and content creators, to adjust their resources, services and strategies for larger revenues and better user experience.

In this paper, we characterize the service usage of a leading online video sharing system in China, namely Youku ([www.youku.com](http://www.youku.com)). Youku achieves more than 500 million monthly active users and 800 million daily video views [2]. It offers a comprehensive type of video service, providing both the user-generated-content (UGC) sharing content and the video-on-demand (VoD) copyrighted content. In our analysis, we focus on two essential service usages of the online video sharing system: video uploading and video playback. Based on over 12 billion large-scale traffic traces collected from the network operator and 30-day long-term meta data crawled from the Youku website, we provide a detailed and in-depth analysis on how Youku is utilized by uploaders and viewers. To the best of our knowledge, this is the first comparative

study to analyze the dynamics of video uploading and video playback conjointly for the online video sharing system. The contributions of our work are summarized as follows:

- We collect large-scale traffic traces and crawl long-term video meta-data for the analysis of online video uploading and playback conjointly. We further release our whole dataset to the public for research purpose ([https://github.com/lichenyu/Datasets/tree/master/Youku\\_Uploading\\_Playback\\_151212\\_151218/](https://github.com/lichenyu/Datasets/tree/master/Youku_Uploading_Playback_151212_151218/)).
- We provide a detailed and comparative characterization of user activities, video properties and video popularity for the uploading and playback service usages of online video sharing system.
- To measure the dynamics of video service usage, we propose several novel metric notions, such as the category entropy and the video lifetime.
- Together with the characterization results, we also discuss how network operators, service providers, and other interested parties can utilize our observations to design and adjust their resources and strategies to gain larger revenues and improve the service performance.

There have been some existing efforts on characterizing the online video service:

**Traffic traces based analysis:** Based on traffic traces, Gill *et al.* [3] examined the user access patterns, file properties, video accessing and referencing behaviors of YouTube in a campus network. Zink *et al.* [4] analyzed the properties, population and access patterns of YouTube videos based on the traffic collected from a university network. Ben Abdesslem *et al.* [5] collected video requests of YouTube in a cellular network, and studied the characteristics of users, videos and trends of popularity.

**Video meta-data based analysis:** Based on crawled video meta-data, Cha *et al.* [6] characterized the popularity distribution, popularity evolution and content properties of YouTube and a Korean video service provider Daum. Abhari *et al.* [7] crawled data from YouTube website for five months, and analyzed the popularity distribution and access pattern. Cheng *et al.* [8] also performed a long-term crawling, and studied static properties, access pattern, popularity distribution, popularity trend and social network of YouTube videos

Our study complements these existing works by analyzing and comparing the uploading and playback of online video

service at the same time, based on both traffic traces and crawled data. The remainder of this paper is organized as follows. The details of our dataset are described in Section II. Section III provides the characterization study and discusses the implications. Conclusions are drawn in Section IV.

II. DATASET

The data used in this paper consists of two parts: i) **traffic traces** collected from a network operator for the analysis of video playback and ii) **video meta-data** crawled from the Youku website for the analysis of video uploading.

A. Large-scale traffic traces

Large-scale traffic traces were collected at the export interfaces of a leading Internet service provider and network operator which covering a whole province in northeastern China. A traffic trace contains the URL of an HTTP transaction, timestamp and anonymized user identifier. Video requests of Youku can be identified by their specific formed URLs, which match the regular expressions `.*\,youku\.com/v_show/id_[A-Za-z0-9=]{17}.*` or `.*\,youku\.com/.videos/[A-Za-z0-9=]{13}/(comments|download)\?.*`. Meanwhile, video ID, a distinct 17-digit identifier for each video, can be extracted from the URL of the video request. The collection period lasted from December 12th, 2015 to December 18th, 2015. Overall, we collected 12,106,121,482 traffic traces, with 155,991 Youku video requests covering 49,252 viewers and 101,639 videos. We further collected the meta-data of those videos, by sending the video ID into the open API (`cloud.youku.com`) provided by Youku. The meta-data contain the static and dynamic statistic information of a video, including category, duration, uploaded date, uploader, view count and etc.

B. Long-term video meta data

To study the dynamics of the video uploading in Youku, we crawled the meta-data of a set of newly uploaded videos with a two-step data collection procedure. First, on a certain date, we initially crawled the video ID list from the portal's "most recently uploaded videos of the day" section. Video meta-data of these newly uploaded videos were retrieved at the same time. Then, we tracked the daily popularity of these videos for 30 consecutive days. We retrieved the video meta-data every day and extracted the updated view counts. Thus, for each tracked video, we could obtain its daily view counts for a month since the released date. The crawling lasted from December 12th, 2015 to December 18th, 2015. And during those days, 29,916 videos were either deleted by the uploaders or blocked by Youku. Therefore, we exclude those videos from our dataset. Eventually, our long-term crawling dataset consists of 144,457 videos uploaded by 80,929 uploaders.

III. ANALYSIS

A. User Activity

**Per-day Activity:** We first look through the daily user activity in Youku. Fig. 1 shows the daily numbers of uploaders,

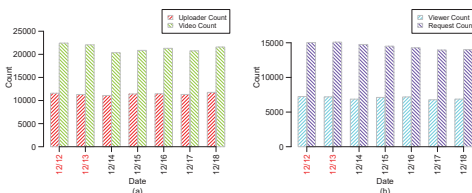


Fig. 1. User activity over different days for (a) uploading and (b) playback.

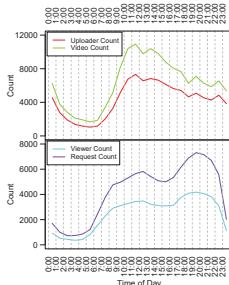


Fig. 2. User activity over different time of day.

uploaded videos, viewers and video requests in our dataset. It can be noticed that the video (request) number each day is generally larger than the uploader (viewer) number, indicating users may upload (watch) multiple videos within one day. For the uploader count and the viewer count, there is no apparent difference between weekdays and weekends (shown as red text in the figure). And for the video count and the request count, the numbers on weekends are slightly larger than those on weekdays. In summary, although the scale of users is quite steady among those days, the activity of those users can be different. Network operators and service providers should allocate their resources for those days with high user activity in advance, in order to provide a satisfied user experience.

**Per-hour Activity:** Next, we examine the circadian pattern of video uploading and video playback for the Youku service. Fig. 2 illustrates the uploader count, uploaded video count, viewer count and video request count during each hour in our dataset. It can be noticed that those numbers vary dramatically within a day, with low values at late night and high values during the daytime. Generally, the video (request) count is proportional to the uploader (viewer) count. For the *uploaded videos*, the number increases rapidly in the morning from 7:00 to 10:00, and reaches its peak at 11:00. During the time from 11:00 to 22:00, the uploading activity decreases steadily with two remarkable drops appearing at 12:00 and 18:00 (lunch and supper time). Then, there is a major decrease in the uploader activity from 23:00 to 6:00. And for the *video requests* from viewers, there is also an increase in the morning lasting from 5:00 to 8:00, which shifts two hours earlier comparing to the playback. During the daytime from 9:00 to 16:00, the viewer activity is relatively steady, with a peak appearing at 12:00. Then request count increases since 17:00, touches the top at

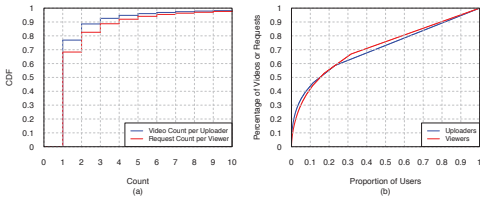


Fig. 3. (a) CDF of per-user service usage intensity. (b) Percentage of usage against proportion of users.

20:00, and keeps relatively large until 22:00. Note the *playback user activity* keeps increasing at lunch time, supper time and night, during which the uploading user activity decreases. At last, the video playback decreases dramatically at late night and touches the bottom around 3:00.

Those variances in per-hour user activity between the video uploading and video playback can be explained by the purposes of different users in utilizing the online video service. For the viewers, they are just for entertainment and relaxation. They can initial requests whenever they want (e.g. as soon as they wake up in the morning), and prefer to watch videos at their spare time (e.g. meal breaks and late night). While for the uploaders, they need to do some preparations before uploading a video, and usually take it as a work to accomplish in the working hours. Fig. 2 shows that the rush hours of video uploading and video playback are seldom overlapped. A good advice to uploaders is to adjust their uploading time according to the viewer count at different time of the day. That is, videos should be released at or even slightly before the rush hours for viewers, in order to obtain more potential views and thus make their videos popular.

**Per-user Usage:** At last, we focus on the intensity of service usage for each user. Fig. 3 (a) shows the cumulative distribution function (CDF) of the video number for each uploader and the request number for each viewer. Similar heavy-tail scenarios can be observed from the figure. For instance, within one day, 76.93 % of the uploaders upload only 1 video and 92.62 % of them upload less than 4 videos. As for the viewers, 68.31% of them watch only 1 video and 88.72% of them initiate less than 4 requests. Meanwhile, there are around 2% “heavy” uploaders (viewers) achieve more than 10 videos (requests) in one day. We further plot the percentages of total uploaded videos and video requests over the proportion of users (in decreasing order) in Fig. 3 (b). It can be noticed that the top 20% uploaders and viewers both account for around 55% of the service usage. Therefore, the famous Pareto principle [9], which states that roughly 80% of the effects come from 20% of the causes, does not hold for the per-user uploading and playback of the Youku service.

**B. Video Property**

**Category:** In this subsection, we focus on the categories of the videos, to figure out what kinds of videos are widely uploaded or watched by Youku users. Youku categorizes each of its video into one of 25 predefined categories. In Table.

TABLE I  
RANK LIST OF VIDEO CATEGORY AND CORRESPONDING DURATION

Uploading		Category	Duration (mins)	
Rank	Name	% of Videos	Median	Mean
1	Life	16.45	4.28	9.49
2	HowTo	11.68	8.15	17.07
3	Music	10.86	3.71	7.40
4	Game	9.32	20.27	25.15
5	Family	7.14	1.93	3.84
6	TV Episode	7.00	17.49	20.27
7	News	6.02	2.26	6.13
8	Sport	5.44	3.97	11.40
9	Creativity	4.12	3.80	5.38
10	Amateur	4.02	1.85	5.63

Playback		Category	Duration (mins)	
Rank	Name	% of Videos	Median	Mean
1	TV Episode	26.45	44.85	43.50
2	Animation	10.20	13.00	15.59
3	Movie	8.07	87.21	68.60
4	Life	6.89	4.39	9.00
5	Entertainment	6.04	5.04	9.04
6	Variety Show	5.94	63.46	53.73
7	Music	5.64	4.03	6.23
8	Amateur	3.84	4.31	7.06
9	HowTo	3.57	4.67	16.50
10	Humor	3.45	3.65	5.64

I, we list the Top 10 most frequently uploaded and watched video categories and their video proportions in our dataset. Those categories cover 82% of the total uploaded videos and 80% of the total watched videos. As shown in the table, the rank list varies significantly across service usages. For the uploaded videos, the top category is “Life”, followed by “HowTo”, “Music”, “Game” and “Family”. Most of the videos in these categories correspond to the user-generated-content (UGC), such as self-made cooking videos, makeup tutorials, game replays, singing recordings, home videos and etc. While for the watched videos, over 1/4 of them are TV episodes, and the Top 3 most popular categories (“TV Episode”, “Animation”, and “Movie”) all correspond to the copyrighted VoD content. This means although most of the videos uploaded to Youku every day are the UGC videos, the key component that attracts users to watch videos on Youku are the copyrighted VoD videos. Moreover, from the perspective of service purposes for different users, we can infer that a large part of the uploaders utilize Youku to save their private videos and share with others, while most of the viewers use Youku for entertainment and relaxation. The analysis of video categories demonstrates promising practical applications. For instance, the service provider could further adjust the content purchasing and advertisement placement strategies according to the video category. And the network operator should consider deploying dedicated cache servers for the hot VoD videos.

Another interesting fact we find about video category is that some users may only upload (watch) some specific category of videos, while others may upload (watch) videos in a wide range of categories. To measure the heterogeneity of user preference on different video categories in uploading and playback, we propose the notion of category entropy. Let  $c_u(i)$  be the number of videos in category  $i$  uploaded (watched) by

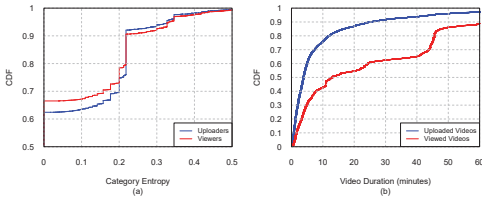


Fig. 4. (a) CDF of category entropy. (b) CDF of video duration.

user  $u$ , and  $n$  be the total number of video categories. The normalized category entropy of user  $u$  is defined as  $H_u(C)$ :

$$H_u(C) = -\frac{1}{\ln(n)} \sum_{i=1}^n \frac{c_u(i)}{\sum_{j=1}^n c_u(j)} \ln \frac{c_u(i)}{\sum_{j=1}^n c_u(j)} \quad (1)$$

$\ln(n)$  is the maximum entropy that can be achieved within  $n$  categories, and the normalized value should be in between 0 and 1. If the videos spread evenly across categories, the normalized entropy will be very close to 1. Otherwise, if the videos are concentrated in a few categories, the value will be close to 0. Fig. 4 (a) shows the CDF of normalized category entropies for the uploaders and viewers with multiple videos in our dataset. It can be noticed that the category entropies are generally very small: for 70% of the uploaders and 73% of the viewers, the entropy values are less than 0.2. This means both the uploaders and the viewers tend to focus on only a small handful of video categories. And for as many as 62% uploaders and 67% viewers, their category entropies are even equal to 0, indicating they would only upload or watch one specific category of videos in Youku. In addition, an obvious increase in the entropy distribution at around 0.22 can be observed in the figure for both uploaders and viewers. It corresponds to the users with two videos in different categories. As analyzed in Section III-A, the proportion of users who upload or watch two videos is relatively large, leading to this concentration in the category entropy distribution.

**Duration:** Next, we investigate how long the videos uploaded and watched by users in Youku are. Fig. 4 shows the CDF of video duration for the uploaded videos and watched videos in our dataset, respectively. The median and mean of video durations in each category are also shown in Table. I. It can be noticed that generally the uploaded videos are quite short. 76% of them are less than 10 minutes, and over 90% of them are less than half an hour. In comparison, the percentages for the watched videos are 43% and 62%, respectively. That is, the watched videos are comprised of more long videos. This is as expected, since a large part of the watched videos are the copyrighted VoD videos such as TV Episodes, Movies, Variety Shows and Animations, which are usually long videos referred to Table. I. In particular, an obvious increase in the duration distribution of the watched videos can be observed at around 45 minutes from Fig. 4. According to Table. I, this increase, which accounts for around 20% videos, is caused by the large number of TV Episodes and Variety Shows. The observations of video duration above can provide useful applications in

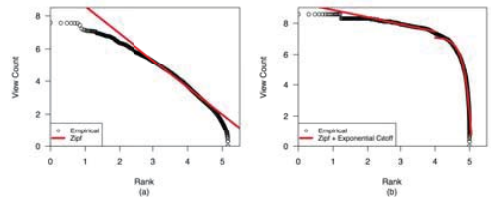


Fig. 5. View count against rank on a log-log scale. (a) uploading, (b) playback.

practical. For instance, the network operators can manage the uplink and downlink bandwidth accordingly, and design better caching and replication mechanism for the video traffic.

### C. Video Popularity

**View Count:** In this subsection, we analyze popularity characteristics of the uploaded videos and watched videos in Youku. We use *view count*, how many times a video is watched by users, as the metric of video popularity. For the uploaded videos, we examine their view counts after a month since the uploaded date; and for the watched videos, we investigate their view counts at the time of requests. It is widely observed that the popularity of online content follows the Zipf's law [10], with the access frequency  $f$  of the rank  $k$  content following:

$$f(k; s, N) = \frac{1/k^s}{\sum_{n=1}^N (1/n^s)} \quad (2)$$

where  $N$  is the total number of the content, and  $s$  is the value of the exponent characterizing the distribution. This function follows a straight line when plotted on a log-log scale axis.

$$y = \alpha \cdot x + \beta \quad (3)$$

To check whether the popularity of Youku uploaded videos and watched videos follows the Zipf's law, we plot the view counts against ranks on a log-log scale in Fig. 5. For the uploaded videos in Fig. 5 (a), an approximate linear relationship can be observed between the rank and the view count. We further run a linear regression with our dataset, and plot the fitted line as the red line ( $\alpha = -1.6510$ ,  $\beta = 10.2300$ ) in the figure. Comparing with the perfect Zipf distribution, there are less extremely popular and unpopular videos in the empirical data. As for the watched videos in Fig. 5 (b), we find the curve is strongly skewed with a heavy tail decreasing dramatically. Clearly, a straight line can not fit the distribution well. This is because in the case of video playback, there are far less unpopular videos watched by users than the Zipf model predicted. In our analysis, we try to fit the distribution with a piecewise function: we keep the linear Zipf model as the beginning part of the distribution when the view count is large, then use an exponential distribution to fit the heavy tail.

$$y = \begin{cases} \alpha \cdot x + \beta & x \leq r \\ a \cdot e^{b \cdot x} + c & x > r \end{cases} \quad (4)$$

We run a regression with empirical data, and get the estimated parameters as  $\alpha = -0.4570$ ,  $\beta = 9.3037$ ,  $a = -3.1225 \times 10^{-11}$ ,  $b = 5.1560$  and  $c = 7.1043$ . The fitted curve is shown

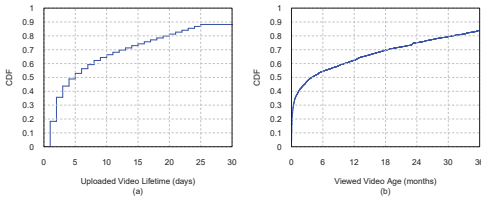


Fig. 6. (a) CDF of video lifetime. (b) CDF of video age.

as the red line in Fig. 5 (b). It's obvious that our model can fit the practical distribution quite well. Overall, for the videos uploaded by users every day, after a month, their popularity will approximately follow the Zipf's law. And for the videos watched by users every day, their popularity can be fitted by a Zipf model with an exponential cutoff.

**Active Period:** In this subsection, we further analysis video popularity in the temporal dimension. For the video uploading, we propose the notion of video *lifetime*, to analyze for how long a new uploaded video can be actively watched among the viewers. More specifically, we check the daily view count of a video since its uploaded date. If the video fails to receive adequate daily views starting from the  $k$ -th day for over  $T$  consecutive days, we regard the lifetime of the video ends on the  $k$ -th day. In other words, the lifetime of this video is  $k - 1$  days. We define the adequate daily views  $ADV_v$  as:

$$ADV_v = \min(\delta, \lambda \frac{V_v}{n}), \quad (5)$$

where  $n$  is the number of the observation days, and  $V_v$  is the total view count that video  $v$  receives during the observation period.  $\delta$  and  $\lambda$  are two thresholds to judge whether the video is active or not. That is, if the daily view count of a video is less than  $\delta$  or less than  $\lambda$  times of the average view count increase, the video will be treated as non-active. In our analysis, we set  $\delta = 100$ ,  $\lambda = 1.5$  and  $T = 5$ . Fig. 6 (a) shows the CDF of lifetime for all the uploaded videos in our dataset. It can be noticed that most of the videos have rather short lifetime. For 66% of the videos, their lifetime is less than 10 days. Meanwhile, there are indeed some videos keep attracting viewers. Around 10% of the videos are with lifetime longer than our observation period (one month). The average lifetime for all videos is 9.75 days. The analysis of video lifetime can provide direct help to the service adjustment and resource allocation. For instance, we can conclude that the golden days to place advertisements or provide caches for the newly uploaded videos are the first couple of days. After only 5 days since the uploading, over half of the videos will be not actively watched by users anymore and thus become less valuable for service providers and network operators.

And for the video playback, we focus on the *age* of the videos watched by users. Video age is defined as the number of days between the released date and the viewed date. Fig. 6 (b) shows the CDF of age for all the watched videos in our dataset. It can be noticed that the watched videos are comprised of both young and old videos. The relative young videos account for a

large fraction. 37% of the videos are less than 30 days old, and over half of the videos are less than 4 months old. This can be explained by the fact that the popular categories of video, such as TV episodes, animations, variety shows (referred to Section III-B) will update periodically and keep attractive to users only for a short period of time. Meanwhile, there are indeed some old videos popular among users. For instance, 16% of the videos watched by users are more than 3 years old, and the oldest video in our dataset (a short humor video named "What makes a man excited") has an age of over 9 years. Understanding the characteristics of video age is crucial for content creators and service providers to learn the watching behavior of their users, and adjust the content they provide to attract more users. And network operators can also redesign the delivery mechanism for video streaming accordingly, to improve the service quality and user experience.

#### IV. CONCLUSION

In this paper, we have presented a detailed characterization on two major online video service usages: uploading and playback, of a leading online video sharing system in China, Youku. Based on the datasets containing over 12 billion large-scale traffic traces and 30 days long-term crawling video meta-data, three key aspects have been analyzed: user activity, video properties and video popularity. Our measurements shed light on the uploading and playback characteristics in the Youku service, and got promising implications in practical. The studies presented in this paper are crucial and reliable for all interested parties related to the online video service such as network operators, video portals, online advertisers and content creators, to improve their service performance and user experience.

#### REFERENCES

- [1] "Cisco visual networking index: Forecast and methodology, 2014-2019," [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white\\_paper\\_c11-481360.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html).
- [2] "Youku tudou partners with xiaomi to accelerate multi-screen ecosystem development," [http://ir.youku.com/phoenix.zhtml?c=241246&p=irol-newsArticle\\_print&ID=1988630](http://ir.youku.com/phoenix.zhtml?c=241246&p=irol-newsArticle_print&ID=1988630).
- [3] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *ACM SIGCOMM Conference on Internet Measurement*, 2007.
- [4] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of youtube network traffic at a campus network—measurements, models, and implications," *Computer networks*, vol. 53, no. 4, pp. 501–514, 2009.
- [5] F. Ben Abdesslem and A. Lindgren, "Large scale characterisation of youtube requests in a cellular network," in *International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2014.
- [6] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1357–1370, 2009.
- [7] A. Abhari and M. Soraya, "Workload generation for youtube," *Multimedia Tools and Applications*, vol. 46, no. 1, pp. 91–118, 2010.
- [8] X. Cheng, J. Liu, and C. Dale, "Understanding the characteristics of internet short video sharing: A youtube-based measurement study," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1184–1194, 2013.
- [9] J. M. Juran, "Universals in management planning and controlling," *Management Review*, vol. 43, no. 11, pp. 748–761, 1954.
- [10] D. M. Powers, "Applications and explanations of zipf's law," in *the joint conferences on new methods in language processing and computational natural language learning*, 1998.



# Characterizing the Content Popularity of Online Video Service

Chenyu Li, Jun Liu, Shuxin Ouyang

Beijing Key Laboratory of Network System Architecture and Convergence  
 School of Information and Communication Engineering  
 Beijing University of Posts and Telecommunications, Beijing, China  
 Email: lichenyu@126.com, liujun@bupt.edu.cn, ouyangshuxin@gmail.com

**Abstract**—Given the large volume of video content and the high variability of user attention, it is of great importance to understand the characteristics of online video popularity for technological, economic and social reasons. In this paper, based on data collected from a leading online video service provider in China, namely Youku, we analyze the video popularity in-depth from three key aspects: long-term view count, growth pattern and early view count. First, we propose the Log-Normal distribution to fit the long-term popularity for a set of videos published on the same date. We also analyze how some impact factors, such as video category and video tags, influence the long-term video popularity. Then, we investigate the popularity evolution of an individual video over time, and find a number of patterns that can describe the evolution trend. Furthermore, the number and temporal locations of popularity bursts are considered in the popularity growth pattern analysis. At last, we check the linear relationship between the early video popularity and the long-term video popularity for the videos with different popularity growth patterns. We find the linear relationship appears at very early stage for the videos with a popularity burst at the beginning; while for the videos starting with a slow popularity growth, it usually takes more than 15 days to obtain the influential view count.

## I. INTRODUCTION

With the emergence of Web 2.0 services, a huge amount of video content has been brought into the Internet. There are nearly a million minutes of videos across the network every second, and it will take an individual over 5 million years to watch all the videos across the global IP network each month [1]. Meanwhile, the user attention on those videos is allocated in a rather asymmetric way leading to the “rich-gets-richer” effect: a small part of videos receive most of the views whereas most of the videos remain unnoticed [2]. Hence, a better understanding of online video popularity is critical to content publishers, service providers, online advertisers and network operators. For instance, with the popularity information of the newly published content, effective information services (e.g. video recommendation and searching) can be better designed and supported. And the revenue of advertisements in the online marketing can be maximized through planning campaigns and estimating costs according to the video popularity information. At the same time, network infrastructures such as cache servers can be deployed in advance for better video delivery. And in the social aspect, a thorough understanding of popularity evolution will uncover the rules governing collective human behavior. That is, to reveal where the user view growth of an

individual video comes from: caused by random user choices, endogenous effects (e.g. listed on the home page) or exogenous effects (e.g. shared on other websites) [3].

In this paper, we characterize the content popularity of a leading online video service provider in China, namely Youku ([www.youku.com](http://www.youku.com)). To understand the nature and dynamics of video popularity, we collect the meta-data of 19,629 newly published videos from Youku website for 30 consecutive days. The popularity of those online videos is measured by *view count*. We analyze the characteristics of video popularity from three key aspects: long-term view count, growth pattern and early view count. The main contributions of our work are summarized as follows: 1) We analyze the distribution and impact factors of the long-term video popularity. The Log-Normal distribution is proposed to fit the long-term view counts. We find that the video popularity is largely influenced by video category, and certain keywords in the tag list would also affect the video popularity. 2) We investigate the popularity growth trend of an individual video, and reveal a small number of growth patterns. In particular, the number and temporal locations of sudden popularity bursts are taken into account in our analysis of growth patterns. 3) We shed light on the correlation between early popularity and long-term popularity. We utilize linear regression models on a logarithmic scale to model the correlation, and find the regression variables and the model parameters depend on the popularity growth patterns.

The rest of this paper is organized as follows. Section 2 briefly discusses the related work. Section 3 describes the dataset and collection method used in our study. The main analysis results are presented in Section 4. In section 5 we conclude the paper.

## II. RELATED WORK

**Static content popularity of online video service:** Gill *et al.* [4] collected traffic in a campus network and characterized the usage patterns, video file properties, video popularity and transfer behaviors of the YouTube service. Zink *et al.* in [5] also collected traffic in a university campus network, and analyzed the session duration, session data rate, video popularity and access patterns of YouTube videos. Cha *et al.* in [6] analyzed the video popularity characteristics of two video sharing systems, YouTube and Daum. They examined

the distribution and evolution of video popularity, and investigated mechanisms to improve the video delivery. Those works provide valuable insights into the popularity of video contents. However, their datasets were collected on only one or several days. Our study complements these works by tracking the video popularity for 30 consecutive days. Moreover, by doing this we can study the growth patterns of video popularity for an individual video.

**Dynamic content popularity of online video service:** Crane *et al.* [3] studied the video popularity with an epidemic model. They found the relaxation process followed a power law, and classified the popularity evolutions into four patterns according to the peak fraction. Szabo *et al.* [2] found early access of the content could reflect long-term user interest and presented a log-linear model to predict the popularity of online content. Figueiredo *et al.* [7] [8] analyzed the characteristics of popularity evolutions and referrer types of YouTube videos. Our study complements those previous works by proposing more proper popularity growth patterns that can describe the whole observation period rather than only the peak days. What's more, we take the number and temporal locations of the popularity bursts into consideration. Based on the analysis of popularity evolution, we further correlate early view counts with long-term view counts with different growth patterns.

### III. DATA COLLECTION

The data used in our study are collected from Youku via its open API provided for developers [9]. By sending a specifically formatted request to the API, we can get a list of VIDs for the newly published videos on that day. The VID is a distinct 17-digit identifier for each video, composed of 0-9, a-z, A-Z and =. For each video (VID), we can collect its meta-data via the API, which includes: duration, category, tags, resolution, UID, current view count and etc. The UID is a distinct numeric identifier for the video uploader. We can also obtain meta-data of the uploader (UID), such as the number of videos uploaded by him and the number of his followers.

We developed a crawler in Python to automatically retrieve the data. First, on October 17th, 2015, we initially collected 19,629 newly published VIDs. Video and user meta-data were grabbed at the same time. Then we run the crawler every day to retrieve the daily view count for each video in the dataset. The whole collection procedure had lasted for a month, from October 17th to November 15th, 2015. During the collection period, 740 videos were deleted by the uploaders and 227 videos were blocked by the video portal. We exclude those videos from our dataset. Eventually, we collect the 30-day time series of view counts for 18,662 videos.

### IV. ANALYSIS

#### A. Long-term View Count

First, we analyze the video popularity after a month since the publication date. Figure 1 (a) shows the cumulative distribution function (CDF) of the Day 30 view count for each video in our dataset. The x-axis of the figure is logarithmic for a better visualization. It can be noticed that the user interests

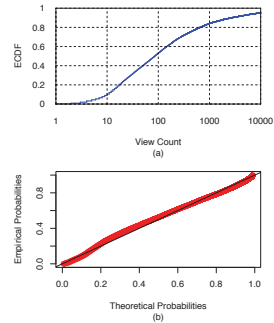


Fig. 1. (a) CDF of long-term view count. (b) P-P plot of long-term view count and log-normal distribution.

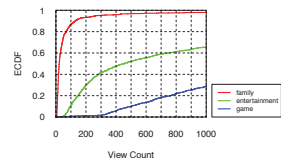


Fig. 2. CDF of long-term view count for videos in different categories.

are allocated in a rather asymmetric way. For instance, over half of the videos are watched for less than 100 times, and 84% of them are watched for less than 1,000 times. Meanwhile, the view counts for some popular videos can be very large (over 10,000), and the most popular one is watched for 14,381,973 times. We use the Log-Normal distribution to fit the long-term popularity, with a probability density function (PDF):

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] \quad (1)$$

By running a regression with our data,  $\mu$  and  $\sigma$  are estimated:  $\mu = 4.66606$  and  $\sigma = 2.06004$ . A P-P plot is shown in Figure 1 (b) to visualize the fitting performance. It can be noticed that most of the dots in the scatter plot follow a straight line with the slope equals to 1. This indicates the Log-Normal distribution is indeed a good approximation for the long-term view count. Such an uneven distribution of video popularity indicates that although a lot of videos are published every day, only a small part of them can actually become popular. We note our finding is different from the previous studies [10] [6] [11], which fit the video popularity of YouTube with Gamma, power law with an exponential cut-off and Weibull distributions. Although these distributions vary over service providers and datasets, they are all heavy-tail distributions.

We further investigate the impacts of different video properties on the long-term popularity. Some Pearson's correlation coefficients with the Day 30 view counts are as followed: video duration (0.034), tag number (0.003), video resolution (0.137), uploader video number (0.003) and uploader follower number (0.002). No significant correlations can be observed between these properties and the long-term view count.

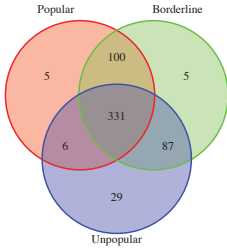


Fig. 3. Venn diagram of tags for videos with different popularity levels.

Nevertheless, we find the category of a video has a great impact on its view count. Figure 2 shows the CDF of long-term view count for three representative categories: “family”, “entertainment” and “game”. It is apparent from the figure that significant differences of popularity lie between different video categories. On average, videos of the three categories are watched for 865, 9,786 and 21,780 times, respectively. Videos in the “family” category tend to have very small view count. 93% of them are watched for less than 200 times. While the percentage of the “entertainment” videos and the “game” videos are around 30% and lower than 5%, respectively. Videos in the “entertainment” category have a relatively uniform popularity distribution. And for the “game” videos, the view counts are usually very large. Over 71% of them are watched for more than 1,000 times. The differences of view counts reflect the differences of user interests in these videos. Videos in the “family” category are usually private videos uploaded by users to share with their close friends. Other users have limited interests in those UGC (User-Generated-Content) videos. Most of the videos in the “entertainment” category are VoD (Video-on-Demand) contents released by the video portal to attract various users. And the game replays and commentaries account for a large proportion of the “game” category. These videos are very popular among the young users, who contribute to the majority of the view counts.

Then we look through how the video tags influence the popularity. We first extract the “hot” tags which appear in more than 10 videos. In total, there are 563 hot tags, account for 2.5% of all the tags. It can be noticed that most tags only appear in several videos. Then, based on the 30-th day view count of a video  $v(30)$ , we define 3 *popularity levels* for the videos in our dataset. **popular**:  $v(30) \geq 1000$ ; **borderline**:  $100 \leq v(30) < 1000$ ; and **unpopular**:  $v(30) < 100$ . Thus, each hot tag can be assigned to one or more sets (popular, borderline and unpopular) according to the popularity level of each video that contains the tag. Figure 3 demonstrates the Venn diagram of the three tag sets. It can be noticed while most tags are common among videos on different popularity levels, there are indeed some tags that only appear in the videos on certain popularity level. For instance, the video tag “Survivor Games” appears in 18 videos in our dataset. All of the 18 videos are at the *popular* level, with the average view count equal to 50,833 and the largest view count equal to 123,943.

Hence, the presence of certain tags in the video tag list can affect the video popularity.

### B. Growth Pattern

In this subsection, we investigate how the popularity of an individual video evolves over time since its publication. Instead of a simple logarithmic distributed popularity growth, as previous research concluded [11], we find the popularity evolutions of YouTube videos are rather complicated. Slow and steady growths of view counts can be noticed for some videos, while others may exhibit dramatic rises (i.e. bursts) in the popularity growth. To measure the growth trend, we focus on the daily increase in view count and propose the notion of view count growth rate. The view count growth rate  $r(i)$  of a video on the  $i$ -th day since the publication is defined as:

$$r(i) = \begin{cases} \frac{v(i)}{v(n)} & i = 1 \\ \frac{v(i) - v(i-1)}{v(n)} & 1 < i \leq n \end{cases} \quad (2)$$

where  $v(i)$  is the view count of the video on the  $i$ -th day and  $n$  is the total number of days in the observation period. And we define three *states* of the daily growth trend according to  $r(i)$ . That is, **slow**:  $r(i) < 0.02$ ; **steady**:  $0.02 \leq r(i) < 0.1$ ; and **fast**:  $r(i) \geq 0.1$ . Hence, for each video a sequence of 30 growth states can be observed. We denote the sequence as  $S = \langle s_1, s_2, \dots, s_{30} \rangle$ . We set  $s_i = 0$  for a slow state,  $s_i = 1$  for a steady state and  $s_i = 2$  for a fast state. Then, by merging the consecutive states with same values, we try to generate the final growth pattern of view count. However, we find there are some stochastic disturbances of the states in the sequence. For instance, a steady growth pattern may result in a sequence like  $\langle 1, 1, 1, 1, 1, 0, 1, 1, \dots, 1 \rangle$ , with a one-day “noise” state 0 (*slow*). If we directly merge the sequence for final pattern, the result will be “101” (*steady-slow-steady*) instead of “1” (*steady*). To tackle this problem, we apply a smoothing algorithm on the state sequences before the merge step, to filter out the disturbances. The algorithm iterates the state sequence with a window of specified length, and leverages the output of the last state in the window according to other state values in the window. The pseudo code of our algorithm is shown in Algorithm 1. Our definition of the popularity growth pattern is different from the previous work [3]. They focus more on the popularity growth shape only near the (single) peak day, while we consider the popularity evolution pattern during the whole observation period.

As analyzed in the previous subsection, many videos in the dataset just get limited views during the 30 days, and their growth trends can be greatly volatile. Hence, we exclude the inactive videos, whose view counts are less than 90, in the growth pattern analysis. These videos get less than 3 views per day on average, and have limited research and commercial values. And for the rest active videos, we calculate their growth rates, smooth the state sequences (with  $\omega = 5$  and  $\delta = 0.675$ ) and finally extract their growth patterns. Table I shows the top 8 view count growth patterns, which cover 83.64% of the active videos in our dataset. Some representative

**Algorithm 1** Smoothing algorithm for state sequence

---

**Input:**  $S$  = Sequence of the growth states,  $\omega$  = Size of the smoothing window,  $\delta$  = Factor of the smoothing  
**Output:**  $P$  = Smoothed sequence of the growth states

```

1: procedure SMOOTHFILTERING( $S, \omega, \delta$ )
2:    $P = \{\}$ 
3:    $index = \{\}$ 
4:   for  $i = 1$  to  $length(S)$  do
5:     if  $S_i = 2$  then
6:       add  $i$  into  $index$ 
7:       remove  $S_i$  from  $S$ 
8:     end if
9:   end for
10:  for  $i = 1$  to  $\omega - 1$  do
11:    if  $sum(S_1 : S_{\omega-1}) < (\omega - 1)/2$  then
12:       $P_i = 0$ 
13:    else
14:       $P_i = 1$ 
15:    end if
16:  end for
17:  for  $i = \omega$  to  $length(S)$  do
18:     $k = sum(S_{i-\omega+1} : S_{i-1})/\omega * \delta + S_i * (1 - \delta)$ 
19:    if  $k \geq 0.5$  then
20:       $P_i = 1$ 
21:    else
22:       $P_i = 0$ 
23:    end if
24:  end for
25:  for all  $i$  in  $index$  do
26:    insert 2 into  $P_i$ 
27:  end for
28: end procedure
    
```

---

 TABLE I  
 TOP 8 GROWTH PATTERNS

Growth Pattern	# of Videos	% of Videos
fast-slow	3748	40.92
fast-steady-slow	1267	13.83
fast-slow-fast-slow	1096	11.97
fast-slow-steady-slow	697	7.61
fast-steady-slow-steady-slow	423	4.62
fast-steady-slow-fast-slow	214	2.34
slow-fast-slow	114	1.24
fast-steady-fast-steady-slow	102	1.11

patterns are illustrated in Figure 4. It can be noticed that for most of the videos their growth patterns start with the *fast* state. Meanwhile, there are indeed some videos that get few views at the beginning and become popular in the middle of the lifetime (e.g. *slow-fast-slow*). And some videos may experience more than one burst of popularity during the observation period.

Understanding the growth pattern of view count is of great importance to derive the user interactions and predict the popularity evolution. The propagation process of a video can be reflected by the number and temporal locations of bursts in the view counts (i.e. *fast* growth). A burst followed by a steady growth is usually related to the word-of-mouth recommendations of the video in the social networks. And a dramatic increase in popularity in the middle of the growth is usually caused by external events such as: the video is listed on the first page of Youku; or the video is retweeted by a famous person on the OSN (Online Social Network).

### C. Early View Count

Early patterns of user access can reflect the long-term video popularity [2]. On a logarithmic scale there is an approximately linear relationship between the two:  $\log v(n) \propto \log v(i)$ . When determining how early the view count should be used for the model (i.e. the value of  $i$ ), previous works treat all the videos as a whole at the same time.

In this subsection, we further analyze the correlation between early view count and long-term view count, and complement the previous work [2] by combining the popularity

 TABLE II  
 REGRESSION PARAMETERS AND THE R-SQUARED VALUES

Growth Pattern	$\alpha$	$\beta$	$R^2$
Type1	0.9877	0.3709	0.90
Type2	0.9101	0.5328	0.90
Type3	0.7364	1.3131	0.75
Type4	0.9252	0.5332	0.78
Type5	0.9893	0.3495	0.97

growth patterns of individual videos with the linear model. We find the growth pattern of view count has a great impact on the choice of  $i$  in  $v(i)$ . For the videos which experience a burst of popularity at the beginning of the lifetime, they soon obtain the majority of the long-term view counts. Hence, the value of  $i$  for them can be very small. On the contrary, for the videos which start with a *slow* popularity growth, much more days are needed to get the influential amount of view counts. Hence, the  $i$  in  $v(i)$  for these videos should be much larger. Meanwhile, for those videos with more than one burst, the value of  $i$  should also be larger, as each burst accounts for a smaller proportion of the long-term view count compared to the videos which only experience burst once. Figure 5 shows the average normalized view count on 30 days for videos with 5 different kinds of popularity growth patterns:

**Type1:** Videos with only one burst of popularity, and the burst appearing at the beginning of the lifetime.

**Type2:** Video with multiple bursts of popularity, and one burst appearing at the beginning of the lifetime.

**Type3:** Videos with only one burst of popularity, and the burst appearing in the middle of the lifetime.

**Type4:** Videos with multiple bursts of popularity, and all the bursts appearing in the middle of the lifetime.

**Type5:** Videos without any popularity bursts.

It's apparent from the figure that the curves of videos starting with bursts are much higher than others. It takes only 1 and 2 days for the first two types of videos to obtain 50% of the total view counts. While for the rest three types of videos, they get the equivalent view count fractions on the 15th, 17th and 15th day, respectively. Hence, in the correlation analysis between early and long-term view count, we choose  $v(1), v(2), v(15), v(17), v(15)$  to model the relationship for each type of videos.

Figure 6 shows the long-term view counts v.s. the corresponding early view counts for the five types of videos on a log-log scale. The scatter plots of the data indicate strong linear correlations between the early and long-term view counts, with the average correlation coefficient exceeding 0.91. We describe the correlation with a linear model:

$$\log v(n) = \alpha \log v(i) + \beta \quad (3)$$

We further run regressions with the early view counts we choose for each type of the videos. The results are shown by the red lines in Figure 6. And Table II gives the estimated model parameters and the R-squared values of the regressions. It can be noticed that the observation durations of early view count we choose are proper for modeling the linear correlation with long-term view count. And the estimated parameters vary between different popularity growth patterns.

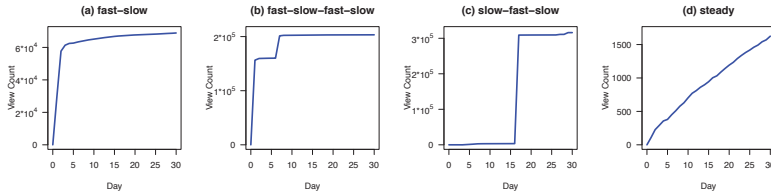


Fig. 4. Illustrations of representative growth patterns.

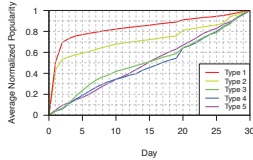


Fig. 5. Average normalized view counts for videos with different types of growth patterns.

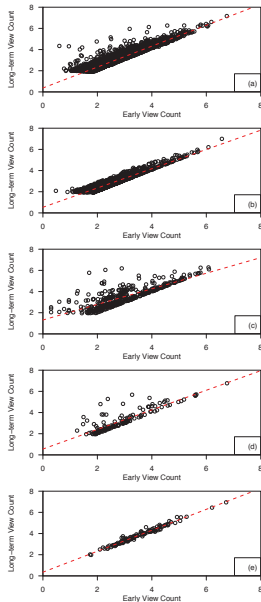


Fig. 6. The long-term view counts vs. the corresponding early view counts for the five types of videos on a log-log scale.

Overall, the early view count  $v(i)$  indeed has a linear correlations on a log-log scale with the long-term view count  $v(n)$  ( $n = 30$  in our case, same as [2]). However, the extent of early view count, that is the value of  $i$  for  $v(i)$ , and the model parameters depend on the popularity growth pattern of the individual video. Generally, for the videos experiencing a burst at the beginning of the lifetime, early view count on the first or second day can well reflect the long-term view count. However, for the videos starting with a slow growth

of popularity, it usually takes over 15 days to obtain the influential view count. Considering to the practical value of the model, such as predicting further view count,  $v(i)$  should be as early as possible while ensuring the performance of the model. Overall, with the specific models based on growth patterns, long-term view count will be better derived.

V. CONCLUSION

In this paper, we have thoroughly characterized the content popularity of a leading online video service in China, Youku. Based on the dataset crawled from Youku website for a month, three key aspects have been analyzed: long-term view count, popularity growth pattern and early stage view count. Our measurements shed light on how the popularity of video content distributes and evolves over time, as well as what influences and reflects the popularity. The findings uncovered in this paper are crucial and reliable for all interested parties of online video service such as content publishers, service providers and network operators, to improve the service design and video delivery for better user experiences.

ACKNOWLEDGMENT

This work was supported by 111 Project of China (B08004).

REFERENCES

- [1] C. Systems, "Cisco visual networking index: Forecast and methodology, 2014-2019 white paper."
- [2] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, 2010.
- [3] R. Crane and D. Sornette, "Robust dynamic classes revealed by measuring the response function of a social system," in *Proceedings of the National Academy of Sciences*, 2008.
- [4] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *ACM SIGCOMM conference on Internet measurement*, 2007.
- [5] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of youtube network traffic at a campus network—measurements, models, and implications," *Computer Networks*, 2009.
- [6] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Transactions on Networking*, 2009.
- [7] F. Figueiredo, F. Benevenuto, and J. M. Almeida, "The tube over time: characterizing popularity growth of youtube videos," in *ACM international conference on Web search and data mining*, 2011.
- [8] F. Figueiredo, J. M. Almeida, M. A. Gonçalves, and F. Benevenuto, "On the dynamics of social media popularity: a youtube case study," *ACM Transactions on Internet Technology*, 2014.
- [9] "Youku api," <http://open.youku.com>.
- [10] X. Cheng, C. Dale, and J. Liu, "Statistics and social network of youtube videos," in *International Workshop on Quality of Service*, 2008.
- [11] X. Cheng, J. Liu, and C. Dale, "Understanding the characteristics of internet short video sharing: A youtube-based measurement study," *IEEE Transactions on Multimedia*, 2013.

# Evaluating QoS in Large-Scale Wireless Sensor Networks in Context of Smart Homes

Ana Koren, Dina Šimunić

University of Zagreb, Faculty of Electrical Engineering and Computing  
Unska 3, 10000  
Zagreb, Croatia

E-mail(s): {ana.koren, dina.simunic} @fer.hr

**Abstract**— As the number of smart homes is rising exponentially, along with its abilities and services invariably expanding, reevaluating the concept itself constantly along the way, performance and quality of service (QoS) of wireless sensor networks comes into question, especially in the context of energy efficiency. In this paper analysis of the performance of IEEE 802.15.4 MAC operating in beacon-enabled mode is given, with all the nodes being within the hearing range of each other in a smart home operating in a way specialized to help its users with their Activities of Daily Living (ADL). Then, an analysis of the same network on a large scale – network consisting of two dozen smaller WSN networks interconnected is given. The goal is to develop a better understanding of energy-related issues, performance and QoS in wireless sensor networks, and to help find novel approaches to networking in the future that would exploit the characteristics of wireless media which will be of importance in building the concept of Smart Cities.

**Keywords**— *Ambient Assisted Living; eWALL; Wireless Sensor Networks; Smart Home*

## I. INTRODUCTION

Projections show that over 20% of Europeans will be 65 or over by 2025 [1]. Many, due to various impairments (physical or cognitive) need assistance in their Activities of Daily Living (ADL). Thus, several solutions from the field of smart homes have been developed, including eWALL framework (project no. 610658: "eWALL for Active Long Living"), eWALL is a "Caring Home" environment consisting "senses" intuitively the wishes and needs of the person that lives in the house, providing unobtrusive daily support, notifying informal and formal caregivers when necessary and serving as a bridge to supportive services in areas such as risk management and home safety, e-Health and lifestyle management and ADL [2]. It is IoT-based (Internet of Things) as it connects various smart objects, such as Plugwise [3], Beddit [4], Fitbit [5], Philips Hue [6], etc. A vital part of the system is the sensing environment consisting of wireless sensor nodes which collect the information which they then forward into sinks to be processed into relevant data. Performance and quality of service (QoS) of wireless sensor networks comes into question, especially in the context of energy efficiency. In this paper analysis of the performance of IEEE 802.15.4 MAC operating in beacon-enabled mode with all the nodes being within the hearing

range of each other in a smart home is given. Afterwards, the network is scaled up and the same evaluation of the large-scale network is performed. This will ensure important insight and conclusions, resulting in better overall understanding of energy-related issues, QoS and general WSNs' performance.

## II. NETWORK MODEL

The proposed model for a smart home includes variety of sensors (measuring temperature, illuminance, humidity, etc.) as well as wearable sensors, all realized through XBee [7] (ZigBee) and Arduino [8] devices. ZigBee [9] is a standard base network protocol supported by the ZigBee alliance. It uses the transported services of the IEEE 802.15.4 network specification. Thus, the IEEE 802.15.4 standard specifies the physical layer and MAC sub-layer for Low-Rate Wireless Personal Area Networks while the ZigBee standard specifies the network and application layers. There are three types of nodes in ZigBee wireless sensor networks: coordinator, routing node and end device or remote transceiver node. The coordinator selects suitable channels and adds child node to the already established network making it solely responsible for the network intelligence. In the first case, small-scale scenario of a single Smart Home (40 square meters), it is defined that all the nodes are within the hearing range of each other (40 m) which means there's no need for route nodes. The most relevant metrics for measuring performance of a wireless sensor network are packet delivery ratio, average end-to-end delay and power consumption [10]. Energy efficient protocols can massively improve the network's lifetime. Thus, design of MAC layer in particular has a major role in WSN's energy efficiency.

## III. SIMULATION

The following analysis of the small-scale network is based on the model presented in the Figure 1 and consists of a grid of 25 equally spaced and numbered nodes with the sink in the center. Furthermore, it is presumed that all nodes are within the hearing range from each other (distance up to 40 m, which means that there are no route nodes necessary). This type of wireless sensor network represents a single Smart Home environment. Network has been simulated for

a period of one minute (60 seconds). A wireless sensor network consists of a sink node and a number of small wireless sensor nodes. Sensor nodes collect sensory information and send it to the sink node through a wireless hop-by-hop transmissions.

Table 1. Characteristics of ZigBee sensor nodes used in simulations

Channel type	Channel/WirelessChannel
Radio-propagation model	Propagation/TwoRayGround
Network interface type	Phy/WirelessPhy/802_15_4
MAC type	Mac/802_15_4
Interface queue type	Queue/DropTail/PriQueue
Antenna model	Antenna/OmniAntenna
Max packet in IFQ (Interface priority queue)	50
Routing protocol	AODV (Ad hoc On-Demand Distance Vector Routing)

The full characteristics of ZigBee nodes used are given in the table above (Table 1).

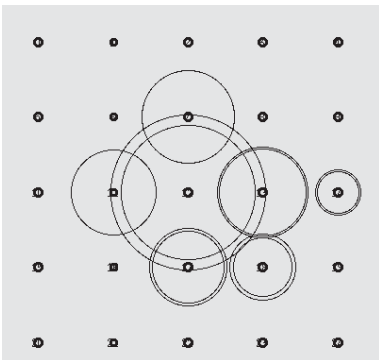


Figure 1. Topology of WSN in a single Smart Home (25 nodes)

Network congestion reduces the quality of service and it is the result of a network node carrying more data than it can handle. This typically results in queuing delay, packet loss or blocking the new connections [13].

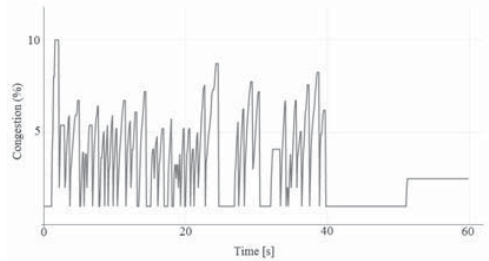


Figure 2. Congestion levels (%) for 25-node WSN

In the given network, congestion (shown in Figure 2) is low - under 10% at all times, which means it's not affecting the QoS negatively. The graph also shows the bursty nature of traffic generated by the nodes.

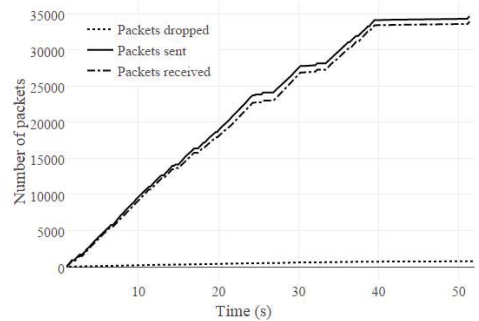


Figure 3. Packets sent, received and dropped in a 25-node network

The most relevant metrics, however, of a WSN performance are packet delivery ratio, average end-to-end delay and power consumption [14]. Having Packet Error Rate (PER) between 0% and 15% does not deteriorate notably the availability of the network, because MAC layer retransmissions are enough to compensate the packet errors and links that break rarely [15]. Using Ad hoc On-Demand Distance Vector routing protocol (AODV) a small-scale ZigBee network consisting of 25 nodes has given the results pictured in Figure 3. Packet error rate amounts to 2.6% with 34804 packets sent, 33921 packets received and 883 packets dropped.

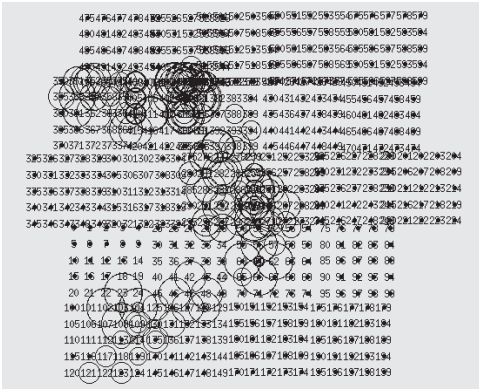


Figure 4. 600-nodes WSN topology

The projections show an exponential growth of IoT objects in the near future, with the main focus being smart home and smart cities. Experts estimate the number of IoT objects will reach almost 50 billion by the year 2020 [16]. Energy-efficient smart homes with sensing environments bring many benefits to its users and the environment so the task is to find a scalable solution which would help pave the way towards smart cities. In that context, the following simulations have been performed. Figure 4 shows a large-scale wireless sensor network, consisting of a grid of 600 nodes spaced within a one square kilometer. In this scenario, the nodes are static (not moving).

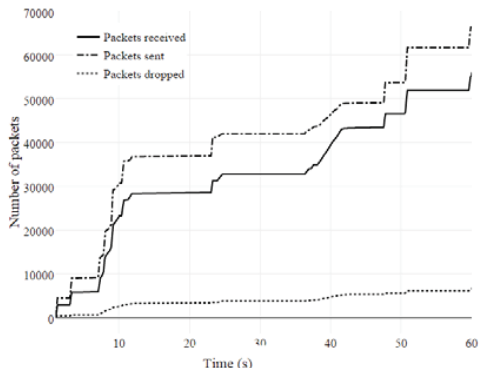


Figure 5. Packets sent, received and dropped in a 600-node network

Following the necessary routing of packets as well as heavier traffic (larger number of nodes), QoS is affected negatively. As shown of Figure 5, in this case there were 62777 packets sent, 56018 packets received and 6759 packets dropped, resulting in a packet drop rate of 10.8%.

Packets can be dropped due to the queue being full or errors during routing of packets. This results in a lower availability of the network.

Considering that Smart Home environments often make use of wearable sensor technology, simulating static nodes isn't enough to predict QoS of WSNs. Thus, taking the previously built network (Figure 4) and defining the randomized movement of a portion of nodes (simulating the wearable sensors in Smart Home environments) should give us more relevant results.

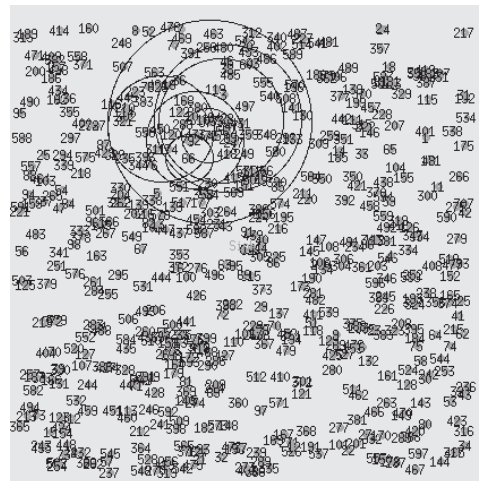


Figure 6. 600-nodes WSN topology with moving nodes (e.g. wearable sensors)

Figure 6 shows a large-scale wireless sensor network consisting of 600 nodes with randomized movement (velocity 0-5 m/s) inside the defined area (one square kilometer).

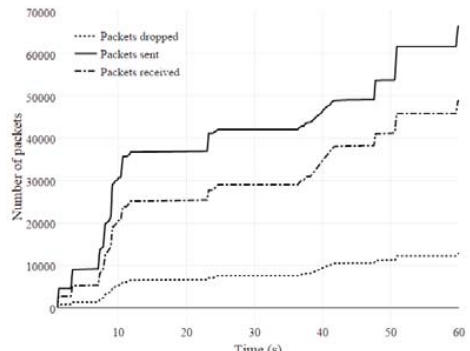


Figure 7. Packets sent, received and dropped in a 600-nodes moving network



Using AODV protocol in this scenario results in a packet error rate of 25.6%, which heavily influences quality of service. There have been 66438 packets sent, 52919 packets received and 13519 packets dropped (results shown on Figure 7). Such a high packet drop rate deteriorates the network's performance and needs to be addressed.

Table 2. Number and percentage of packets dropped in static and dynamic networks through time

Time (s)	Network static / dynamic	Number of packets dropped	Percentage (%) of packets dropped
10	Static	521	8.6
10	Dynamic	4852	12.6
30	Static	3951	10.9
30	Dynamic	7925	18.3
60	Static	6759	10.8
60	Dynamic	13519	25.6

Table 2 (pictured above) shows number and percentage of packets dropped in both static and dynamic networks through time. It's important to note that in static network, the percentage of packets dropped stabilizes at around 11 percent and doesn't continue to grow after the initial 30 seconds. Thus, the network is able to process the messages at the rate they're coming with acceptable performance and drop rate being less than 15 percent. On the other hand, in the dynamic network, the drop rate continues to increase and reaches unacceptable levels after the first minute, with over 25 percent of total messages being dropped. This also indicates that congestion levels are too high in the network and, as a consequence, the network isn't able to process messages in time resulting in continuous increase in percentage of packets that are being dropped. Since the drop rate is over 15%, the nodes in network cannot retransmit the packets at a fast enough rate and cannot compensate the delay in packet delivery which means the availability of the network deteriorates notably. Analysis of the cause of the packets dropped is given in the Table 3. Excluding the drops of packets at the end of the simulation, the majority of packets have been dropped due to the link quality at MAC layer. This is due to the link quality of the packets arriving at the MAC layer is not within acceptable threshold [15] and hence they get dropped. Link Quality Indicator (LQI) in ZigBee networks indicates how strong the communications link is. It is a computed value, based on the received signal strength as well as the number of errors received.

Table 3. Drop of packets cause

Number of packets dropped	Percentage of packets dropped	Cause of packet drop
24	0.2%	ARP (dropped by ARP)
1192	8.8%	END (dropped packets due to end of the simulation)
342	2.5%	DUP (drop MAC duplicate)
99	0.7%	CBK (drop RTR MAC callback)
10545	78%	LQI (Link Quality Indication packet drop)
15	0.1%	NRTE (dropped due to no route being available)
1302	9.7%	TTL (dropped because time to live has reached zero)

Other issues include TTL reaching zero which means maximum number of hops wasn't enough for reaching the destination. A small portion of packets dropped due to no route being available. Identifying these reasons will help designing the MAC layer in an optimal way. Major packet drops occur at the point when nodes start to move dynamically as opposed to being static: the rate of packets dropped rises to 25.6% which deteriorates the performance of the wireless sensor network as it exceeds tolerable losses that can be compensated by retransmissions on the MAC layer (between 0% and 15%).

IV. CONCLUSION

The analysis of a 600-node WSN shows some of the troubles with performance when scaling up a number of nodes, especially in a dynamic topology (moving nodes). The model allows the evaluation of the statistical distribution of the traffic generated by the nodes. Results show the rate of dropped packets becomes unacceptable (over 15% PER) when the nodes are moving. Static topology consisting of 600 nodes has resulted in 10.6% of packets being dropped while the dynamic one with the same number of nodes has 25.6% drop rate. Especially high (78%) is the portion of packets dropped in the dynamic scenario because of the Link Quality Indication (LQI). The LQI is computed from the raw Received Signal Strength Index (RSSI) by linearly scaling it between the minimum and maximum defined RF power levels for the radio. This

is due to the link quality of the packets arriving at the MAC layer is not within acceptable threshold and hence they get dropped. Thus, it is necessary to redesign the MAC layer taking into consideration node movement, which will be further explored in future work. Routing in such a topology is of crucial importance as well and should be one of the main focuses of further research. The goal was to develop a better understanding of energy-related issues, performance and QoS in wireless sensor networks, and pave the way for novel approaches to networking in the future that would exploit the characteristics of wireless media which will be of importance in building the concept of Smart Cities.

**Acknowledgement** - We thank our colleagues from the eWALL project who provided insight and expertise that greatly assisted the work presented in the paper.

#### REFERENCES

- [1] Eurostat - European Commission, "Population structure and aging", June 2015
- [2] The eWALL Consortium, "D2.9 Task forces COPD, MCI and ARI", eWALL for Active Long Living, November 2014
- [3] Plugwise, <https://www.plugwise.com/>
- [4] Beddit, <http://www.beddit.com/>
- [5] Fitbit, <https://www.fitbit.com/eu/chargehr>
- [6] Philips Hue, <http://www2.meethue.com/en-xx/>
- [7] XBee, <http://www.digi.com/lp/xbee>
- [8] Arduino, <https://www.arduino.cc/>
- [9] The ZigBee Alliance, <http://www.zigbee.org/>
- [10] Ramachandran I., Das A. K., Roy S., "Analysis of the Contention Access Period of IEEE 802.15.4 MAC", ACM Transactions on Sensor Networks (TOSN) (Vol. 3 Iss. 1), April 2007
- [11] Fragkiadakis A., Charalampidis P., Tragos E., "Adaptive Compressive Sensing for Energy-Efficient Smart Objects in IoT Applications", 2014 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory, and Aerospace and Electronic Systems (VITAE), May 2014, pp. 1-5.
- [12] Kim C., Talipov E., Ahn B., "Reverse AODV (R-AODV) Routing Protocol in Mobile Ad hoc Networks", The 2006 IFIP International Conference on Embedded and Ubiquitous Computing (EUC'06), LNCS 4097, August 2006., pp. 522-531.
- [13] Al-Bahadili H., "Simulation in computer network design and modeling: Use and analysis". Hershey, PA: IGI Global, 2012
- [14] Kim C., Talipov E., Ahn B., "Reverse AODV (R-AODV) Routing Protocol in Mobile Ad hoc Networks", The 2006 IFIP International Conference on Embedded and Ubiquitous Computing (EUC'06), LNCS 4097, August 2006., pp. 522-531.
- [15] Suhonen J., Hämäläinen T. D., Hännikäinen M., "Availability and End-to-end Reliability in Low Duty Cycle Multihop Wireless Sensor Networks", Sensors 2009 (Vol. 9, Iss. 3), 2009, pp. 2088-2116.
- [16] Evans D., "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything", Cisco White paper, April 2011.
- [17] Holger K., Wilig A., "Protocols and Architectures for Wireless Sensor Networks", John Wiley & Sons, 2005.
- [18] Evans D., "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything", Cisco White paper, April 2011.
- [19] Yang G., Zhang G., "An Energy-Efficient Cross-Layer MAC for Wireless Sensor Networks", 2011 International Conference on Electronic and Mechanical Engineering and Information Technology, August 2011, pp. 683-686.
- [20] Yanping Yu, "A Cross-layer Approach for Energy-efficient MAC Layer in Wireless Sensor Networks", 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08), October 2008, pp. 1-3.
- [21] Sadik K. Gharghan, Rosdiadee Nordin, Mahamod Ismail, "Energy-Efficient ZigBee-Based Wireless Sensor Network for Track Bicycle Performance Monitoring", Sensors 2014 (Vol. 14 Iss. 8), August 2014
- [22] Shaikh F. K., Zeadally S., Exposito E., "Enabling Technologies for a Green Internet of Things", IEEE Systems Journal (Vol. PP, Iss. 99), April 2015, pp. 1-12.
- [23] Kaushik P., Patel N. K. R., Singhai J., "Energy Efficient Clear Channel Assessment for LR-WPAN", IJCSI International Journal of Computer Science Issues (Vol. 8, Issue 3, No. 2), May 2011, pp. 387-393.

# Load-balanced Clustering Approach for Cooperative D2D Communications in Cellular Networks

Junhyuk Kim

School of Computing  
KAIST

Daejeon, Korea  
jhkim@nslab.kaist.ac.kr

Sunwoo Kim

School of Computing  
KAIST

Daejeon, Korea  
furyt4nk@gmail.com

Hyunsoo Yoon

School of Computing  
KAIST

Daejeon, Korea  
hyoon@nslab.kaist.ac.kr

**Abstract**—With their high power processors, current smartphones generally support simultaneous LTE and Wi-Fi connectivity. Moreover, using hybrid communication with clustering has significant potential to increase the overall performance of cellular networks. In this paper, a load-balanced clustering algorithm is proposed that maximizes the per-user throughput using clustering while assuring fair burdening of the offloaded packet transmissions for all user devices. The performance analysis is conducted using an ns-3 simulator and is compared with the legacy transmission mechanism of LTE, and the greedy clustering algorithm that selects the best-channel quality device as the cluster head(CH). The simulation results demonstrate that the proposed clustering algorithm effectively increases the per-user throughput and fair transmission of the offloaded packets.

**Keywords**—D2D communication, Wi-Fi-Direct, Clustering, Cooperative communication

## I. INTRODUCTION

Recently, many people use the cellular infrastructure due to the mobile communication standard established by the Wireless Standard Community. In addition, the current smartphone subscription totals 3.4 billion in 2016 and is expected to grow to 6.3 billion by 2021[1]. Driven by this large number of cellular users, the data traffic per base station (BS)(eNB in the LTE terminology) has increased every year. Therefore, in a BS-centralized cellular infrastructure, the increased data traffic causes problems for both service providers and cellular users.

From the service provider perspective, due to the wireless interference caused by the large data traffic at the BS, they cannot provide appropriate Quality-of-Service (QoS) to the users. From the users' perspective, the user equipment (UE) becomes more battery consuming as the UE follows the 4G communication standard [2]. Furthermore, if the number of devices that share the limited channel resources in the existing cellular infrastructure increases, a significant energy inefficiency problem will occur due to significant signal interference at the eNB. For example, devices that have various channel qualities are distributed in cellular networks, and these devices communicate using the resources allocated according to the BS channel policies. However, as the number of users for each BS increases, all UEs receive fewer channel resources than before. Moreover, at least one resource block is allocated to all UEs regardless of their channel quality. In this situation, although an UE that has bad channel quality receives a channel resource, the UE is not able to fully use the allocated resource and may repeats re-transmissions with high packet

drop rates due to the UE's bad channel quality. Whereas even an UE that has good channel quality re-transmits several times because the UE has fewer channel resources in comparison with the UE's communication capacity. As a result, the energy consumption of all UE increases and the throughput of the network decreases.

Because the inefficient transmissions of cellular traffic cause the above situations, cooperation between the cellular link and device-to-device (D2D) communication is needed for relaying cellular traffic. As illustrated in Fig. 1, devices on the right side communicate with the device on the left of the cluster (which is a cluster head) through Wi-Fi Direct communication, and the cluster head relays the offloaded packets from the devices on the right to the BS. In this way, the BS's burden can be reduced due to the D2D communication between the cluster head and the devices. Moreover, D2D communication provides improved channel quality and transmission rates due to the shorter communication range than the 4G communication standard. Therefore, UE that has bad channel quality can obtain high QoS and return allocated channel resources of the cellular networks to the BS when the UE uses D2D communication. In addition, another UE can benefit from using the returned channel resources. Thus, the energy efficiency and network throughput of the network system are improved.

Moreover, a clustering scheme can be used to further increase energy efficiency and throughput [3]. Therefore, the cluster head (CH) communicates with a BS as a representative, while a cluster member (CM) communicates with the CH using D2D communication. However, it should be noted that the CH consumes more battery than the CMs because the CH transmits more data traffic due to the packets collected from the CMs. In [3], a CH selection method that selects the UE that has the best channel quality in the cluster as the CH is used. Not only does the CH selection method burden the UE, but it also incurs unfair burdening of the offloaded traffic transmission on the specific CH.

In this study, the objective was to improve the throughput to be better than legacy transmission mechanism in LTE as well as to resolve the unfair burdening of the offloaded traffic transmission to a specific CH when using clustered D2D offloading in cellular networks. Accordingly, a load-balanced clustering scheme is proposed borrowing hybrid energy efficient distributed (HEED) protocol[4], which is well-known clustering protocol that alleviates the burden of CH for wireless sensor networks.

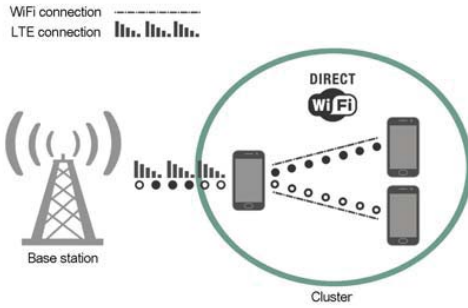


Fig. 1. Overview of a cellular network with clustering user devices. In a cluster, all CMs and one CH are connected over a Wi-Fi network, and two CMs transmit their packets using Wi-Fi Direct communication to the CH. Then, the CH forwards the CM's packets to the BS using LTE communication.

The remainder of this paper is organized as follows. Section II describes our system model. The proposed clustering protocol is presented in Section III. Section IV evaluates the proposed clustering protocol in packet-level simulation. Section V summarizes our contributions and concludes the article.

## II. SYSTEM MODEL

As discussed in [5], the 4G communication system benefits from using D2D communication among the user equipment. The D2D communication boosts energy efficiency and channel utilization, which achieves high fairness.

The proposed basic model is built on [3], which utilizes the clustering scheme for D2D communications in cellular systems. With the proposed system, LTE mobile devices create clusters through forming a Wi-Fi Direct group. In the group, the cluster head (CH) is selected according to the clustering parameters. Moreover, other mobile devices send their traffic to the CH through Wi-Fi Direct communication. The CH takes the role of communication bearer, which transmits offloaded packets from the cluster members (CMs) through the LTE communication to the BS.

The following properties of cellular communication systems are assumed in this model.

- The inter-packet processing overhead between a Wi-Fi packet and a LTE packet is negligible.
- An eNB has channel quality information of all devices.
- Only uplink transmissions of a user device is considered.
- Resource scheduling is not considered.

## III. LOAD-BALANCED CLUSTERING

The overarching goal of the proposed approach is to increase the fairness of burdening the offloaded packet transmissions and the per-user throughput for user devices. Therefore, the CH selection is primarily based on the modulation

and coding scheme (MCS) index of each device. In general, mobile devices and the BS exchange their channel quality information (CQI) before transmitting their packets in order to determine modulation scheme. Moreover, according to the above assumptions, the CQI can be acquired from the BS. In order to increase the fairness of each user device, the communication overheads due to transmitting the offloaded packets from the CMs of the user devices are also considered as a secondary clustering parameter. Thus, the MCS index relevant to the channel quality information is used as the primary parameter to select the initial CH and the secondary parameter to resolve the unfairness burden of the CH.

In the following, the clustering parameters of the proposed clustering protocol and the conditions of executing the proposed algorithm are defined. Moreover, the proposed load-balanced clustering algorithm (LBC algorithm) and the algorithm's pseudocode are presented.

### A. Clustering Parameters

First, the primary parameter refers to the MCS index of the existing LTE system. In general, the signal-to-interference-plus-noise ratio (SINR) value between the user and base station fluctuates due to the fading effect in cellular networks. However, the MCS index is determined according to the corresponding range of SINR values. Thus, using the MCS index value instead of the SINR value as the primary parameter is more stable and does not generate floating point comparisons. In the existing LTE system, the user device must measure the SINR value between the user device and the corresponding BS in order to determine an appropriate MCS index. Therefore, acquiring the MCS index of the user device is not complicated.

Second, the secondary clustering parameter is used to break the connection with the CH and rotate the role of the CH, and is defined as in eq.(1):

$$CH_{prob} = \max\left(\frac{k}{insteadPacket_i}, p_{min}\right) \quad (1)$$

where  $k$  is the constant,  $p_{min}$  is the minimum value of the secondary parameter,  $insteadPacket_i$  is the number of transmitted offloaded packets when the  $i$ th device is the CH. In the LBC algorithm, the secondary clustering parameter of the CM increases in multiples of 2. When the secondary parameter is above 1, the CM breaks the connection with its CH and makes itself a CH. If  $node_i$ , which was a CH, sent numerous offloaded packets from the CMs to the eNB (i.e. the  $insteadPacket_i$  is high), the  $CH_{prob}$  of  $node_i$  is set to a relatively low value. This means that  $node_i$  becomes a CH later than the other devices.

### B. Conditional Execution

Frequent clustering procedures generate extra overheads. Thus, the clustering is restricted to the following three situations after the first execution of the LBC algorithm on all nodes.

First, the clustering procedure is executed when the fairness index  $J$  is below the threshold  $th_f$ . That is, the first condition is expressed as follows:

$$J = \frac{(\sum_{i=1}^n x_i)^2}{n * \sum_{i=1}^n x_i^2} \quad (2)$$

$$J < th_f \quad (3)$$

where  $J$  is the Jain's fairness index [6],  $n$  is the number of devices, and  $x_i$  is the  $i$ th device's transmitted packet from the CMs to the eNB when the  $i$ th device is the CH. Because a small  $J$  indicates an unfairly large number of offloaded packet transmissions of several CHs, triggering the procedure and changing the CH can ease the unfair offloaded packet transmissions between each user device.

Second, the clustering procedure is executed when the Quality of Service (QoS) is below the QoS threshold. In most wireless networks, the QoS functionality is provided based on the wireless standard (e.g. IETF RFC, IEEE 802, 3GPP, etc.). Moreover, the LTE system is garnering greater demand for high QoS because users are charged for their usage. Therefore, the QoS of user devices should be managed carefully, even when offloaded packets are transmitted using Wi-Fi Direct communication technology, which does not have usage charges. Through executing the clustering procedure and seeking another CH that satisfies the QoS requirement, the proposed scheme can manage the QoS issues.

Third, the clustering procedure is executed when the CH changes. Because the clustering protocol is independently executed on each device, the user devices should update their information and synchronize their cluster status. In addition, they should determine whether to leave the cluster or not according to the surrounding devices.

### C. Protocol Operation

As discussed in Section II.B, clustering is triggered when any of the above three conditions are satisfied. At each node, the clustering procedure requires a number of iterations. As described in Algorithm 1, the LBC algorithm consists of two parts: the initialization and the main process. Before a device begins to execute the main process, each device sets their  $CH_{prob}$  according to  $insteadPacket_{NodeID}$ . Later, the LBC algorithm repeats the main process until  $CH_{prob}$  is below 1. During the iteration, the highest MCS value of the neighboring CHs is assigned as the  $max\_mcs\_index$ .

If a device is the CM, the *CheckQoSRequirement* function verifies whether the CH of the device meets the second condition described in Section II.B or not. It returns a false when the MCS index value of the CH of the device is below both  $MCS_{threshold}$  and the MCS index value of the device. After the *CheckQoSRequirement* function returns a false, the current CH must be changed due to the dissatisfaction of the QoS requirement. Thus, the *AttachCond* function confirms that the  $max\_mcs\_index$  satisfies the CH requirement that the given MCS value exceeds the  $MCS_{threshold}$  or MCS value of the device and verifies that the  $max\_mcs\_index$  of a neighbor CH is suitable for the device's CH. If the neighbor CH of the  $max\_mcs\_index$  satisfies the CH requirement, the neighbor CH is assigned to be the CH of the device. Otherwise, the device declares itself as a CH. Then, the node doubles its  $CH_{prob}$ .

However, if a device is a CH, the device verifies the result of the *AttachCond* function and whether the node has a child or not. This condition indicates that a better CH candidate exists and that the device belongs to an one-member cluster (i.e. the number of devices in the cluster is one). Thus, changing the

---

### Algorithm 1 LBC algorithm

---

```

procedure LBC(NodeID)
    I. Initialization
     $S_{nbr} \leftarrow \{v: v \text{ lies within my communication range}\}$ 
    if ( $insteadPacket_{NodeID} \neq 0$ ) then
         $CH_{prob} \leftarrow \max(\frac{k}{insteadPacket_{NodeID}}, P_{min})$ 
    else
         $CH_{prob} \leftarrow CH_{init}$ 
    end if
    II. Main Process
    while  $CH_{prob} < 1$  do
         $max\_mcs\_index \leftarrow \max(getMCS(S_{nbr}))$ 
        if ( $NodeID \neq CH$ ) then
            if (CheckQoSRequirement(CH) = false) then
                if (AttachCond( $max\_mcs\_index$ ) = true) then
                     $CH \leftarrow ID_{max\_mcs\_index}$ 
                else
                     $CH \leftarrow NodeID$ 
                end if
            end if
             $CH_{prob} * = 2$ 
        else
            if (AttachCond( $max\_mcs\_index$ ) = true) and
            hasChild(NodeID) = false then
                 $CH \leftarrow ID_{max\_mcs\_index}$ 
            end if
        end if
    end while
     $previous\_CH \leftarrow CH$ 
     $CH \leftarrow NodeID$ 
    BreakConnection( $previous\_CH$ )
    LBC( $previous\_CH$ )
    for ( $elem$  in getChilds( $previous\_CH$ )) do
        LBC( $elem$ )
    end for

```

---

CH of the device does not generate additional clustering to with other devices and it obtains get better communication quality through joining to another cluster.

After the iteration is completed, the current CH of the device is backed up to  $previous\_CH$  and then the device declares itself as a CH because completing this step results in it becoming the device's turn to be the CH of the cluster. Then, the *BreakConnection* function disconnects all  $previous\_CH$ 's connections. Finally, because the status of  $previous\_CH$  is changed, the LBC algorithm is triggered on all CMs of the  $previous\_CH$  as well as the  $previous\_CH$  in order to synchronize the current state of the cluster.

## IV. EVALUATION

In this section, the performance of the proposed LBC algorithm is evaluated through packet-level simulations using the open source simulator ns3. The evaluation primarily focused on demonstrating a reasonable trade-off between the per-user throughput and fair transmission of the offloaded packets in terms of each mobile device. Furthermore, the scenario was simulated for two hours and the user device's LTE channel quality changed randomly according to the simulation environment.

TABLE I. NS-3 SIMULATION PARAMETERS

Parameter	Value
Common parameters	
Mobility model	Gauss-Markov mobility model
System area	Within a range of 400m from the eNB
Mean velocity	1m/s to 5m/s speed
Simulation time(total)	2 Hours
LTE	
Propagation loss model	Rayleigh propagation loss model
Resource scheduler	Round-Robin scheduler
UIBandwidth	20MHz
DIBandwidth	20MHz
UIEarfcn	18100
DIEarfcn	100
UE TxPower, noiseFigure	10dBm, 9dB
eNB TxPower, noiseFigure	30dBm, 5dB
AMC model	PiroEW2010
WiFi	
WifiPhy	802.11a standard
Propagation loss model	Rayleigh propagation loss model
Datamode	Ofdm54Mbps

### A. Simulation Environment

The simulations were conducted in an ns-3 simulator and the corresponding simulation model is presented here. Table I presents the simulation parameters and models used. The Rayleigh fading model was used as the pathloss model for both LTE and Wi-Fi Direct transmissions. Furthermore, Round-Robin (RR) schedulers were selected as the LTE base station resource scheduler. Moreover, the packet transmissions occurred according to an exponential distribution, and each transmission sent a 2 KB packet to the eNB.

In the simulation, multiple users were randomly located within the communication range of the eNB and Wi-Fi Direct communication range. In order to generate the random movement of the users' devices, the Gauss-Markov mobility model described in [7] was adopted and users moved randomly at speeds from 1 m/s to 5 m/s. In addition, all nodes were dispersed into a field with dimensions 400 x 400 m, with the eNB located in the center of the field. As depicted in Fig.1, users may organize clusters according to the proposed LBC algorithm, and the CH of each cluster sends their packets to the eNB, which is located on the left of Fig.1. Unless otherwise specified, the  $p_{min}$  for  $CH_{prob}$ , which is the minimum probability for becoming a CH, was set to 0.05 and the  $MCS_{threshold}$  was set to 14. Initially,  $CH_{init}$  was set to 0.1 for all nodes. Therefore, the performance of the proposed LBC algorithm was measured against the SINR-based greedy clustering algorithm that selects the device that has the best channel quality in a cluster as the cluster head (CH) (hereafter referred to as the *greedy algorithm*) and legacy transmission mechanism without clustering in a single cell network.

### B. Channel State Probabilities

Before evaluating the LBC algorithm, the channel qualities of the user devices are evaluated first because they can affect the throughput performance. Furthermore, the CH's MCS index value is used as each CM's MCS index value, i.e. if the CH's MCS index value is 26, the MCS index value of each CM is also 26. The clear effect of clustering the user devices

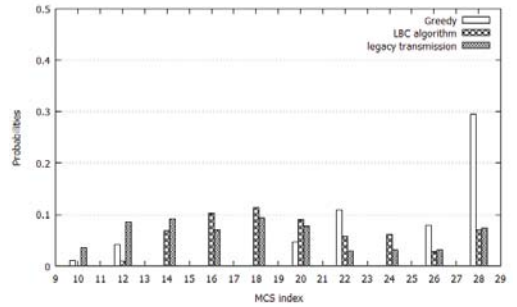


Fig. 2. Channel state probabilities during the simulation.

is the increases in the probability of transmissions with higher MCS values. As depicted in Fig. 2, the *greedy algorithm* uses higher MCS with high probabilities because the CH has the best channel quality in the cluster and all of its CMs' channel qualities refer to the CH's channel quality. In contrast, the channel state probabilities of the LBC algorithms are almost uniformly distributed above  $MCS_{threshold}$  because the LBC algorithm's CH selection criteria is more general. Naturally, the channel state probabilities of the legacy transmissions exhibit the normal channel state probability distributions of the cellular network.

### C. Throughput

The evaluation of the throughput began with the simple scenario of communicating clusters and eNB as described in Fig. 1. Fig.3 illustrates the throughput of the different communication algorithms. In Figure 3, it can be observed that the *greedy algorithm* exhibited the highest throughput performance, while the legacy transmission without clustering exhibited poorer performance than the other two algorithms. Moreover, the proposed clustering algorithm exhibited approximately 15% lower throughput performance than the *greedy algorithm* because the proposed algorithm changes the CH more frequently than the *greedy algorithm* and the SINR value of the CH is not always the best in the cluster. Moreover, as the number of user devices increased, the proposed clustering algorithm and the *greedy algorithm* exhibited a saturation effect [3], which had small successive improvements from some point of N.

### D. Fairness Index for Offloaded Packet Transmission

A simulation to measure the fairness index regarding how evenly the offloaded packet transmissions were distributed in the cluster was also conducted. Because the CH must receive the offloaded packets from cluster members through Wi-Fi Direct communication and transmit the packets to the base stations through LTE communication, the CH consumes more battery than the other cluster members. The battery consumption of the CH due to the offloaded packet transmissions from CMs should be considered because the CH user pays an equal charge for their communication. Therefore, how devices in a cluster share their offloaded traffic was measured using Jain's fairness index [6].

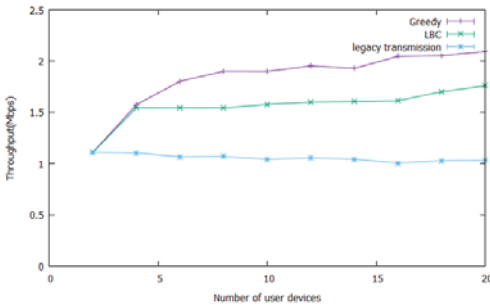


Fig. 3. Performance results for the per-user throughput in the cellular network.

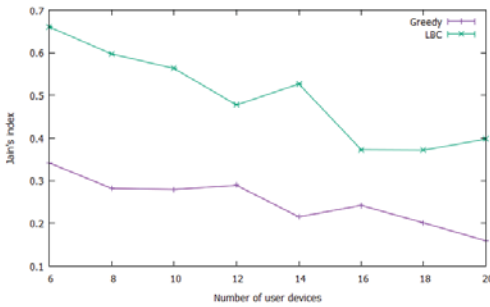


Fig. 4. Jain's fairness index for greedy algorithm and our LBC algorithm.

Fig. 4 depicts that the proposed LBC algorithm achieves a high fairness index value. As the number of user devices increases, the fairness index value has a tendency of general decline because the increased number user devices affects the denominator of eq. (2). Because the proposed LBC algorithm changes the CH when the overall communication has a low fairness index value and it generates more CHs than the greedy algorithm, the high fairness index value can be maintained for CHs as well as cluster members. However, the greedy algorithm has a poor fairness index value due to its CH selection according to its policy that prefers the best channel quality; thus, a significant burden of transmitting the offloaded packets from CMs is focused on specific CHs. That is, the greedy algorithm burdens specific devices that have better channel quality than other devices for the throughput of entire networks.

V. CONCLUSION

In this paper, a new clustering algorithm has been proposed for LTE and Wi-Fi Direct communications in cellular networks. The proposed LBC algorithm creates clusters that consist of user devices using Wi-Fi Direct communication and each cluster member offloading their traffic to the cluster head. In addition, the role of the cluster head is probabilistically rotated during the algorithm for load-balanced cluster communication. The proposed LBC algorithm exhibits a significant

improvement in terms of throughput as well as fairness of user devices. The simulation results indicate that the LBC algorithm achieves approximately 50% higher throughput than the legacy transmission. Moreover, fair offloaded packet transmission is achieved in the LBC algorithm while the throughput performance of the LBC algorithm exhibited only 15% lower speed than the greedy algorithm. Therefore, it is expected that the LBC algorithm will be beneficial to mobile users through increasing the per-user throughput and ensuring fair burdening of the offloaded packet transmissions.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2014R1A2A2A01006957).

REFERENCES

- [1] E. C. Communications. (). Ericsson mobility report: On the perse of the networked society, [Online]. Available: <https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf>.
- [2] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4g lte networks," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, ACM, 2012, pp. 225–238.
- [3] A. Asadi and V. Mancuso, "Dronee: Dual-radio opportunistic networking for energy efficiency," *Computer Communications*, vol. 50, pp. 41–52, 2014.
- [4] O. Younis and S. Fahmy, "Heed: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *Mobile Computing, IEEE Transactions on*, vol. 3, no. 4, pp. 366–379, 2004.
- [5] A. Asadi and V. Mancuso, "On the compound impact of opportunistic scheduling and d2d communications in cellular networks," in *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems*, ACM, 2013, pp. 279–288.
- [6] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984, vol. 38.
- [7] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless communications and mobile computing*, vol. 2, no. 5, pp. 483–502, 2002.

# Energy-efficient two-hop OFDM DF relay system

Enis Kocan, Milica Pejanovic-Djurisic  
Faculty of Electrical Engineering  
University of Montenegro  
Podgorica, Montenegro

Sofoklis Kyriazakos  
Faculty of Engineering and Science  
Aalborg University  
Aalborg, Denmark

**Abstract**—We examine solutions enabling reduction of power transmitted from both relay station (R) and base station (S - Source) on downlink in dual-hop OFDM decode-and-forward (DF) relay system, while keeping end-user capacity on the same level. Three different solution are examined, assuming implementation of ordered subcarrier mapping (SCM) at R station, then applying transmit antenna selection (TAS) on subcarrier basis at R station, and combination of these two solutions. We developed simulation model with parameters from LTE-Advanced reference scenario in rural environment and for the cases of indoor and outdoor placed user. Based on the comparison of achievable mean data rates per subcarrier in each considered system, we obtained data on percentage of achievable transmit power reduction in these systems, compared to OFDM DF baseline system. The derived results have shown that TAS solution brings significant power saving, while implementation of SCM cannot be considered as justifiable for the considered scenarios.

**Keywords**—*relay, OFDM, energy-efficient, mean data rate, antenna selection, subcarrier mapping*

## I. INTRODUCTION

Mobile cellular networks are facing ever increasing demands for higher data rates and system capacities, especially in dense populated urban areas. One of the possible solutions for these issues is shrinking the base station (BS) coverage area. On the other side, this causes implementation of greater number of base stations, which is not cost-effective solution. Thus, in the fourth generation (4G) of these systems, known as LTE-Advanced systems, implementation of relay stations (R) arose as the solution for extension of coverage area, and/or enhancement of the overall system capacity, [1]. The adopted solution assumes orthogonal frequency division multiplexing (OFDM) based R station, implementing decode-and-forward (DF) relaying in dual-hop relaying scenario, [2]. In such scenario, end-user terminal (D – destination) is out of reach of base station (S – source), and complete communication process between S and D is performed through R station.

OFDM based relay systems have attracted great research interest in the last decade, and researches on different solutions for their performance improvement is still ongoing, [3]-[11]. As one of the interesting solution for capacity enhancement, and/or bit error rate (BER) improvement is implementation of ordered subcarrier mapping (SCM) at R station in dual-hop OFDM relay systems, [3]-[8]. Ordered SCM is technique implemented through pairing subcarriers from the first hop (S - R) link, to subcarriers on the second hop (R - D link), all in

accordance to their channel transfer functions. This means that the subcarrier from the first hop with the best CTF is mapped to subcarrier with the best CFT on the second hop, then subcarrier with the second best CTF from the first hop, to subcarrier with the second best CTF on the second hop, etc. In this way, capacity enhancement is attained in all type of OFDM relay systems, where average signal-to-noise ratios (SNRs) on both hops do not differ too much, but at the same time, in OFDM DF relay systems BER performance improvement is attained, [8]. Another interesting solution for performance improvement of dual-hop OFDM DF based relay systems, in scenarios where R station has more than one transmit antenna, is transmit antenna selection (TAS) on subcarrier basis, proposed in [9]. TAS can be combined with SCM for further performance enhancement, and as it is shown in [9], for attaining the best BER performance, “modified” SCM with TAS should be applied. This modified SCM assumes that the subcarriers having the worst CTF on both hops are omitted, and then SCM is realized.

Besides system capacity and BER, a very important measure for assessment of certain relay solution is its energy efficiency. In general, it is not always the case that one-way relaying systems are more energy efficient than direct communication systems, [12], but in the assumed scenario where there is no possibility of direct communication between S and D, this should not be raised as an issue, [13]. In literature, energy efficiency of OFDM based relay systems usually considers different power allocation and relay selection algorithms for minimizing energy consumption, [10], [11]. Here, we take a different approach. Namely, we assess the level of transmit power reduction achievable with different solutions for performance improvement in OFDM DF relay systems, while the system capacity, as quality of service (QoS) measure, is preserved on the same level as in baseline OFDM DF system.

In this paper we compare energy efficiency of OFDM DF relay systems implementing SC, implementing TAS, and implementing TAS jointly with SCM. Using the simulation model with parameters for LTE-Advanced relay system in rural reference scenario, [2], we first compare the achievable average end-user capacities of the considered systems. Based on the obtained results, we managed to derive data on the level of transmit power reduction, from both S and R stations, in the proposed systems, while the system capacity is kept at the same level like in the baseline OFDM DF relay system. At the end, we derive some conclusions on the optimal energy-efficient OFDM DF relay system in the considered scenario.



The paper is organized as follows. Section II presents the system model. The approach undertaken for assessment of energy efficiency of the analyzed systems, and details of parameters used in simulation model, are given in Section III. Section IV contains obtained results and comments on the results, while concluding remarks are given in Section V.

## II. SYSTEM MODEL

We consider the same system model proposed in [9], where we have shown that through implementation of modified SCM and TAS at R station, OFDM DF relay system may attain significant BER performance improvement. This actually means that this system can be considered as an energy efficient solution, which enables energy saving if the BER is kept at the same level as in baseline OFDM DF relay system. However, in this paper we examine the possibility for transmit power reduction from both base station and R station on downlink (DL), while keeping the same end-user capacity system as the quality of service (QoS) performance measure. Thus, DL communication process is taken into account, and it is assumed that end-user (D) is not in the range of base station (S - source). The complete communication process on DL between S and D is performed through the R station, which operates in half-duplex mode. Such a communication system is denoted as dual-hop relay system. OFDM DF relay station is considered, because it is adopted as relay solution for LTE-Advanced system, which should primarily enable coverage extension in this system, [1], [2]. R station uses single antenna for receiving information from S and  $T_R$  antennas for transmitting information to D terminal (Fig. 1). It implements transmit antenna selection on subcarrier basis jointly with ordered SCM (in the following, for notation simplicity, "ordered SCM" will be denoted just as "SCM"). The simplified block-scheme of the R analyzed in this paper is given in Fig. 2.

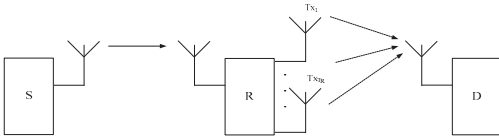


Fig. 1. Dual-hop relay system with TAS at R station

R station presented in Fig. 2, but without block for Subcarrier mapping, would be OFDM DF relay station with TAS. On the other hand if R station would have only one transmit antenna, it could not implement TAS (i.e. block for "Subcarrier comparison and selection" should be omitted). This type of R station would be OFDM DF relay station with SCM.

The signal received from S is OFDM demodulated through inverse fast Fourier transformation (IFFT) and then symbol demapping is performed (Fig.2). Based on the feedback from D terminal, sub-band Channel Quality Indicator (CQI) is obtained (or equivalently, channel estimations on  $T_R$  DL channels), which is used for subcarrier selection on DL. In our analysis we have considered two cases, i.e.  $T_R = 2$  and  $T_R = 3$ . CQI of the R-D link 1 to the R-D link  $T_R$  are fed into the block which performs comparison of corresponding subcarriers from  $T_R$

links and selects the subcarrier with the best channel transfer function on each subcarrier position. After symbol mapping block and OFDM modulation, realized through IFFT (Inverse Fast Fourier Transformations), there is a block for subcarriers scheduling to corresponding transmit antenna. This block also uses the information on the set of the selected subcarriers from the block for subcarrier comparison and selection.

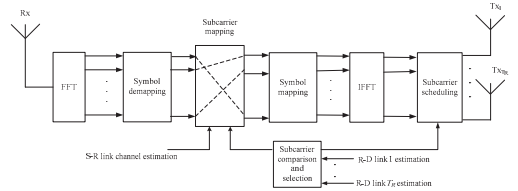


Fig. 2. Block scheme of OFDM DF relay station with TAS & SCM

We assumed perfectly synchronized three communication terminals and that D knows SCM function performed at R station. In our analysis we take that additive white Gaussian noises (AWGN) at R and D have the same variances. We considered the most realistic scenario where direct line of sight (LoS) between base station (S) and relay station (R) exists, because the mobile operator will choose the best possible position of R station. For the R-D link we assumed non-line of sight (NLoS) propagation, with Extended Pedestrian A (EPA) fading model, [2]. In case where end-user (D) is placed indoor, we take that there is only one wall between R and D.

## III. UNDERTAKEN APPROACH AND SYSTEM PARAMETERS

In our analysis of energy efficiency of OFDM DF relay systems, we compare the level of total transmitted powers (including both S and R stations) in the baseline OFDM DF relay system and in the proposed solutions, while all these systems attain the same end-user capacities. However, this comparison is not obtained directly through simulation, but it is derived in two steps. Namely, we first compare achievable mean data rates of the analyzed systems, and based on that, we derive data on achievable transmit power reduction in the proposed solutions, when the capacity is preserved on the level of baseline OFDM DF system. In the following part, details on the conducted analysis, as well as implemented simulation system parameters are presented.

### A. Undertaken Approach

Through the Monte Carlo simulations we have calculated the mean data rate per subcarrier (normalized to unit of bandwidth) at D for each of the analyzed systems:

$$\bar{C} = \frac{1}{M} \sum_{k=1}^M C_k, \quad (1)$$

where  $M$  represents total number of subcarriers of OFDM system, and  $C_k$  is the data rate of  $k$ -th subcarrier at D, normalized to unit of bandwidth. Due to the nature of DF relaying, data rate of the  $k$ -th subcarrier at D ( $C_k$ ) is limited by

the worse (having lower SNR) of the two subcarriers on S-R and R-D links, that transmit the information from S to D:

$$C_k = 0.5 \cdot \log_2 \left( 1 + \min \{ SNR_{k\_SR}, SNR_{k\_RD} \} \right). \quad (2)$$

Factor 0.5 comes due to the half-duplex nature of relaying process, while  $SNR_{k\_SR}$  and  $SNR_{k\_RD}$  denote SNR values on the  $k$ -th subcarrier on S-R link and on R-D link, respectively.

Using the results on achievable mean data rates per subcarrier, we introduced “equivalent SNR” term, which corresponds to the SNR value in AWGN channel necessary for the analyzed relay system to attain the same mean data rate per unit of bandwidth like in the assumed multipath fading scenario. For each of the considered systems “equivalent SNR”,  $SNR_{eq\_i}$ , is obtained as:

$$SNR_{eq\_i} = 2^{(2\bar{C}_i)} - 1, \quad (3)$$

where the index  $i$  denotes the  $i$ -th considered relay system, and  $\bar{C}_i$  is mean data rate per subcarrier of that system. If we want to keep QoS at the same level, then all the proposed solutions should have the same equivalent SNRs as the baseline OFDM DF relay system. Comparison of equivalent SNR of the considered proposed solution proposed solution, with the equivalent SNR of baseline OFDM DF relay system, gives the value of signal power level decrease at end-user (D) for the corresponding system. Thus, the expression:

$$P_{red\_i} [\%] = \left( 1 - \frac{SNR_{eq\_DF}}{SNR_{eq\_i}} \right) \cdot 100, \quad (4)$$

actually provides information on the percentage of the equivalent SNR reduction at D in the system implementing one of the proposed solutions, necessary to attain the same end-user capacity as in the OFDM DF relay system. This reduction of the signal level at D site is attained through reduction of the power transmitted from R station and from the base station (S), due to the described nature of DF relaying process (see (2)).

### B. System parameters

The proposed solution for improving energy efficiency in OFDM based DF relay system can be used irrespective of the type of environment where R station is deployed, but in this paper we put focus on the rural scenario. Using the reference scenario parameters corresponding to relay implementation for the coverage of rural area in LTE-Advanced system and appropriate system parameters given in [2], we have developed a simulation model for evaluating mean data rate per subcarrier of OFDM DF relay system, then for the same system implementing only SCM, for the system implementing only TAS, and for the system implementing TAS and SCM. R station with  $T_{R=2}$  and  $T_{R=3}$  transmit antennas are considered in systems implementing TAS or TAS jointly with SCM. Through this simulation model we manage to extract information on equivalent SNRs attained in the considered systems for the given mean data rate, and based on that, the level of transmit power reduction attained in the analyzed systems, when the goal is to attain the same end-user capacity like in baseline OFDM DF relay system.

In the following part we provide details of the simulation model used.

We have modelled LTE-Advanced relay system having 5MHz bandwidth, which assumes 25 resource blocks with 12 subcarriers each, where subcarrier spacing is equal 15kHz. For the system operating at carrier frequency of 2GHz, we have chosen guard interval of 17 $\mu$ s. S (base station) transmit power is equal to 43dBm and R station transmit power is 30dBm. We take that feeder loss at S is 3dB, and feeder loss at R is 1dB. Antenna gains of S and R stations are equal to 17dBi and 5dBi, respectively, where each R station antenna has the same gain. End-user (D) antenna has the gain of 0dBi. Additive white Gaussian Noise (AWGN) is assumed at both R and D stations, with noise power  $10\log_{10}(kTB) = -137.45$ dB. Noise figures of D and R stations are 7 dB and interference margin is 4dB.

We consider LoS scenario for the S-R link, with the medium dominant component. Path loss models, defined as the reference ones for LTE-Advanced relay systems, are taken from [2]. We assumed outdoor R station, which is placed with site planning, [2]. For D terminal we consider both possible scenarios, i.e. where D is in indoor environment, which in real scenarios happens in almost 80% of the cases, and the other one, where D is in outdoor environment. Through-the-wall penetration loss is equal to 17dB, [2], when the scenario with indoor user is considered.

Multipath propagation fading models for both S-R and R-D links are also taken from [2]. The scenario taken for describing the multipath propagation fading on R-D link is Extended Pedestrian A (EPA) model, given in [2]. The Doppler shift of 10Hz for EPA model is chosen for outdoor environment, which corresponds to user mobility of ~5km/h, while Doppler shift of 0Hz is taken for the indoor user. Doppler shift for the S-R link is equal to 0Hz for each scenario considered.

In the same described scenario conditions we have modeled all the considered OFDM based DF relay systems, in order to assess the level of mean data rates achievable with the same transmitted power in each of the systems. Using this, we obtained information on equivalent SNR for each analyzed systems. Based on comparison of equivalent SNRs of the propose OFDM DF relay systems, with equivalent SNR of the baseline OFDM DF relay system, the level of transmit power reduction is calculated for the specific solution, while the mean data rate is kept on the same level as in the baseline system.

## IV. RESULTS

In the following, we first present the simulation obtained results on achievable mean data rates per subcarrier for all the considered OFDM based DF relay systems in rural scenario, for D placed in both indoor and outdoor environments. Fig. 3 relates to scenario with end-user (D) placed indoor, while Fig. 4 relates to scenario with outdoor user. Both figures assume that S-R distance is equal to 8km. Achievable mean data rates per subcarrier are shown as the function of distance between R and D, with the range 0.2km – 1km for the scenario with indoor user, and 0.2km - 2km for the scenario with outdoor user. OFDM DF baseline configuration is denoted in figures’

legends as “DF”, OFDM DF relay system with SCM implemented is denoted as “SCM”, while OFDM DF relay system with 2 and 3 TAS implemented are denoted as “2 TAS” and “3 TAS”, respectively. In accordance to these markings, the solutions combining SCM and TAS are denoted as “2 TAS & SCM” and “3 TAS & SCM”, for the cases where R has 2 and 3 transmit antennas, respectively.

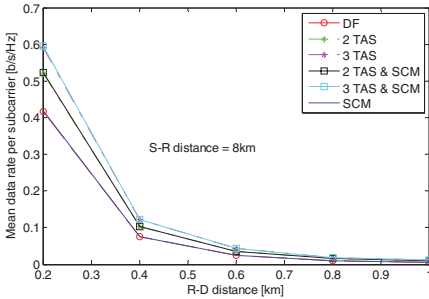


Fig. 3. Mean data rates per subcarrier for indoor user

From the given plots it can be seen that implementation of TAS solution significantly improves achievable data rates compared to OFDM DF relay system. Thus for example, for indoor scenario, comparison of obtained results has shown that 2 TAS solution can bring more than 44% end-user capacity improvement, while 3 TAS solution improves end-user capacity up to 75.7% (see Table I) in the considered cases. This in turn enables transmit power reduction up to 30.9% with 2 TAS solution and up to 43.2% with 3 TAS solution, when compared to baseline system (see Table II), and in the case of the same data rate achieved.

TABLE I. CAPACITY IMPROVEMENT COMPARED TO OFDM DF SYSTEM (INDOOR USER)

R-D distance [km]	0.2	0.4	0.6	0.8	1
SCM [%]	0	0	0	0	0
2 TAS [%]	24.84	36.03	41.38	43.56	44.54
3 TAS [%]	41.44	61.12	70.42	74.32	75.72
2 TAS & SCM [%]	25.32	36.04	41.39	43.56	44.56
3 TAS & SCM [%]	42.28	61.18	70.48	74.32	75.84

On the other side, implementation of only SCM technique at R station brings negligible end-user capacity improvement, when compared with baseline configuration of OFDM DF relay system. Even when combined with TAS solution, SCM can bring only minor additional capacity enhancement to TAS system, and the same holds for the power efficiency this solution enables (Table II). Thus, for indoor users, in case where data rate would be the only performance of interest, then implementation of SCM solution would not be reasonable solution.

TABLE II. TRANSMIT POWER REDUCTION COMPARED TO OFDM DF SYSTEM (INDOOR USER)

R-D distance [km]	0.2	0.4	0.6	0.8	1
SCM [%]	0	0	0	0	0
2 TAS [%]	26.02	27.90	29.78	30.56	30.94
3 TAS [%]	38.15	39.95	42.04	42.94	43.27
2 TAS & SCM [%]	26.35	27.90	29.78	30.56	30.95
3 TAS & SCM [%]	38.38	39.97	42.06	42.94	43.31

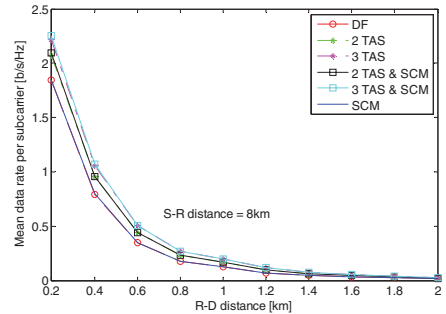


Fig. 4. Mean data rates per subcarrier for outdoor user

The results on mean data rates per subcarrier for outdoor user presented at Fig. 4, show that, in terms of performance, the order among the analyzed systems is the same like for the scenario with indoor user. If we compare values of achievable data rates with the data rate of baseline OFDM DF relay system, we can see that in percentage (Table III), performance improvement goes up to 43.4% for the system implementing 2 TAS, and up to 74% for the system implementing 3 TAS. Like for the indoor scenario, SCM solution in this scenario also brings negligible performance improvement in terms of achievable mean data rates.

TABLE III. CAPACITY IMPROVEMENT COMPARED TO OFDM DF SYSTEM (OUTDOOR USER)

R-D distance [km]	0.2	0.6	1	1.4	2
SCM [%]	0.3	0.13	0.1	0	0
2 TAS [%]	12.65	26.81	32.58	36.43	43.49
3 TAS [%]	20.29	44.95	55.03	61.67	74.18
2 TAS & SCM [%]	13.7	27.08	32.66	36.91	43.49
3 TAS & SCM [%]	22.22	45.61	55.35	62.36	74.23

Extracting equivalent SNR values for the considered systems from Fig. 4, and comparing them with the equivalent SNR of baseline OFDM DF relay system, gave the results on achievable transmit power reduction for each analyzed system (see Table IV). Power efficiency analysis of TAS solutions shows that transmit power can be reduced up to 30.6% and up to 43%, with 2 TAS and 3 TAS solutions, respectively, compared to baseline system. In the outdoor scenario SCM solution, when combined with TAS can bring additional few

per cents of power reduction to the system implementing only TAS, in cases where D is placed close to R station.

TABLE IV. TRANSMIT POWER REDUCTION COMPARED TO OFDM DF SYSTEM (OUTDOOR USER)

R-D distance [km]	0.2	0.6	1	1.4	2
SCM [%]	0.72	0.21	0.06	0	0
2 TAS [%]	29.26	26.52	26.75	27.55	30.62
3 TAS [%]	42.41	38.76	38.63	39.36	43.03
2 TAS & SCM [%]	31.24	26.72	26.81	27.82	30.62
3 TAS & SCM [%]	45.30	39.15	38.77	39.63	43.04

When analyzing pros and cons of SCM implementation, one should bear in mind that modified SCM in combination with TAS significantly improves BER performance of OFDM DF relay system, [9]. Thus, if BER is performance of interest, then OFDM DF relay system implementing TAS and modified SCM can achieve great power saving compared to baseline system, while attaining the same BER performance. Besides, it is well known that SCM brings highest capacity enhancement if the average SNRs on hops do not differ too much, [8]. In the analyzed scenarios it was not the case, which means that in some other cases SCM could bring notable capacity enhancement, and through this, could enable power reduction if the system capacity is preserved on the same level.

## V. CONCLUSIONS

In this paper we analyzed solutions for increasing power efficiency of OFDM DF relay system, while keeping the system capacity on the same level. The system with implemented ordered subcarrier mapping (SCM) at R station is considered, then the system implementing transmit antenna selection (TAS) on subcarrier basis at R station, and the system combining these two solutions. Using the originally developed simulation model, with parameters for the reference LTE-Advanced relay system in rural scenario, we examined the achievable data rates in the considered systems. Based on this, the level of transmit power reduction from both R and base station (S) are extracted.

The conducted analysis has shown that implementation of TAS at R station can bring significant power saving (e.g. more 43% in both indoor and outdoor environments, when R station has 3 antennas), while with employment of only SCM technique negligible transmit power reduction is attained. Joint implementation of SCM with TAS brings additional 2% of transmit power saving to 3 TAS solution, for the considered scenario with outdoor user.

It should be noted that the proposed antenna selection at R may also be implemented for improving energy efficiency for uplink (UL) communications in LTE-Advanced networks. This is particularly interesting for the considered rural scenario, where R station may serve for collecting data from different

types of sensor nodes implemented at homes, in agriculture and for forest ecosystem monitoring. The energy efficiency of the proposed solution would enable long-range communications with low transmitted power from sensor nodes, thus extending the life of their battery supplies.

## ACKNOWLEDGMENT

This work has been supported by the Ministry of Science of Montenegro and the HERIC project through the BIO-ICT Centre of Excellence (Contract No. 01-1001). The paper is also supported by the research project "Advanced solutions for performance improvement of wireless cooperative e-service infrastructures" funded by Ministry of Science of Montenegro.

## REFERENCES

- [1] G. E. M., Zhioua, H. Labiod, N. Tabbane, S. Tabbane, "LTE Advanced relaying standard: A Survey, *Wireless Personal Communications*," Vol. 72, Issue 4, pp.2445-246, 2013.
- [2] 3GPP TR 36.826 v11.3.0 (2013-07) Relay radio transmission and reception (Release 11)
- [3] A. Hottinen and T. Heikkinen, "Optimal subchannel assignment in a two-hop OFDM relay," in Proc. of the 8th IEEE Workshop on SPAWC, pp. 1 – 5, Helsinki, Finland, 2007.
- [4] Y. Li, W. Wan, J. Kong, M. Peng, "Subcarrier pairing for amplify-and-forward and decode-and-forward OFDM relay links," *IEEE Comm. Letters*, vol. 13, no. 4, pp. 209-211, 2009.
- [5] E. Kocan, M. Pejanovic-Djurisic, D. S. Michalopoulos, G. K. Karagiannidis, "BER performance of OFDM Amplify-and-Forward Relay System with Subcarrier Permutation," in Proc. of IEEE Wireless VITAE 2009 Conference, pp. 252-256, Aalborg, Denmark, May 2009.
- [6] E. Kocan, M. Pejanovic-Djurisic, G. K. Karagiannidis, "New Solution for BER Performance Improvement of OFDM AF Relay Systems," in Proc. of IEEE conf. TELFOR 2012, pp. 412-415, Belgrade, Nov. 2012.
- [7] E. Kocan, M. Pejanovic-Djurisic, C. Meckenbraüker, "On the Ergodic Capacity of Dual-Hop OFDM based DF Relay System with Subcarrier Mapping," in Proc. of IEEE conf. ICCSPA'13, pp. 1-5, Sharjah, UAE, Feb. 2013.
- [8] E. Kocan, M. Pejanovic-Djurisic, Performance Improvement of Dual-Hop OFDM Decode-and-Forward Relay System, *Wireless Personal Communications*, DOI 10.1007/s11277-014-2227-8.
- [9] E. Kocan, M. Pejanovic-Djurisic, "A Novel Solution for OFDM Based Relay Systems," *Wireless Personal Communications*, vol 87, no. 3, pp. 679-691, April 2016.
- [10] C. Y. Ho and C.-Y. Huang, "Energy efficient subcarrier-power allocation and relay selection scheme for OFDMA-based cooperative relay networks," in Proc. of IEEE ICC 2011, pp. 1-6.
- [11] C. Sun, Y. Cen, C. Yang, "Energy efficient OFDM relay systems," *IEEE Trans. on Communications*, vol. 61, no.5, May 2013.
- [12] C. Sun and C. Yang, "Energy efficiency analysis of one-way and two-way relay systems," *EURASIP J. Wireless Commun. and Networking*, vol. 2012, no. 46, pp. 1–18, Feb. 2012.
- [13] M. Pejanovic-Djurisic, E. Kocan, M. Ilic-Delibasic, "Energy Efficient Wireless Communications through Cooperative Relaying," *Journal of Green Engineering*, vol. 3, no. 1, pp. 71-90, Oct. 2012.
- [14] H. Holma, A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, 2nd edition, Wiley, March 2011

# Identifying and Analyzing Hotspots in Urban Cities from Cellular Network Data

Sudong Zhang, Yuanyuan Qiao, and Jie Yang

Beijing Laboratory of Advanced Information Networks of BUPT, Beijing 100876, China.

Beijing Key Laboratory of Network System Architecture and Convergence of BUPT, Beijing 100876, China.

Beijing University of Posts and Telecommunications, Beijing 100876, China.

Email: zhangsudong@bupt.edu.cn, yuqiao@bupt.edu.cn

**Abstract**—Massive geo-located cellular data collected from ubiquitous mobile devices creates an unprecedented opportunity to study the overall and regional characteristics of urban cities. As the indispensable part of the urban regions, hotspots, the most popular places in the city, play a vital role in people's daily life. In this paper, we perform an empirical study on hotspots in six major cities of one great province in northeast China using one-month cellular dataset with more than 6 million users and more than 80 thousand base stations. Our analysis examines the aggregate amount, stability and importance of population hotspots and traffic hotspots for six cities. The similarities and differences of hotspots in different cities are concerned. In addition, we pay close attention to the "constant hotspots" which are the most important areas in cities. Furthermore, we also find that strong correlation exists between spatial distribution of constant hotspots and the organization of city structure. Our analyses lead to several key findings and provide an in-depth understanding of urban hotspots.

## I. INTRODUCTION

With the rapid development of communication technology and the popularity of mobile devices, such as smartphones and tablets, users of cellular networks generate a large volume of mobile Internet traffic data. Mobile devices, especially for smartphones, play an indispensable role in people's daily life. The cellular networks need to know where each phone is in order to provide the phones with services. That allows us to know the location of users when they connect to cellular networks. Massive geo-located data brings us new opportunities to study the mobility behavior of mobile phone users, which can help capture the mobility behavior of large scale population. More importantly, cellular data contains the spatial information about individuals and how it evolves during the day, which gives us the chance to investigate the spatial characteristics of urban cities, such as, hotspots in the city.

Urban hotspots are several areas that display a parameter density much larger than others in the city. Different parameters lead to different types of hotspots. For example, population hotspots which use the number of individuals as parameter have higher population density than other areas. As the most important part of urban space, hotspots have attracted lots of attention from city managers, researchers, businessmen, and so on. The results of hotspots detection has multiple applications in real life. In city management, more measures should be taken to solve the noise or pollution problems in population hotspots; in marketing, marketer can ensure the effectiveness

of an advertisement by choosing population hotspots to advertise; as for cellular network operators, reasonable network performance should be provided for mobile Internet traffic hotspots.

In this paper, two kinds of hotspots are considered. The one is population hotspot, which is the most crowded areas in the city. Another is traffic hotspot, which exhibits very large traffic volume. We perform a detailed study on hotspots in urban cities using a large-scale mobile Internet traffic data from a major Chinese cellular network provider. Our dataset is well suited for studying hotspots characteristics because it's collected from real cellular network that serves more than 6 million users and each record in the dataset contains the location information of users. The large and detailed dataset enables us to perform an in-depth study on hotspots identification and high-level analysis of hotspots. We detect population hotspots and traffic hotspots of cities, seek to understand some key characteristics of them and the potential correlations between them. In addition, in order to investigate the most important locations in the cities, we also focus on the most stable hotspots. They give us the opportunity to study the spatial organization of cities from the view of cellular data.

The remainder of paper is organized as follows. We provide a survey of the recent related work in Section II, which is about hotspots identification and the application of geo-located data. Section III describes our dataset and our methodology for detecting the hotspots in urban cities. The experiment results are presented in Section IV, including global characteristics and hotspots characteristics in different cities. Finally, conclusions are drawn in Section V.

## II. RELATED WORK

Given the pivotal role in city management, commercial marketing and mobile network optimization, urban hotspots have gained much more attention from researchers lately. Depend on different purpose of hotspots detection, methods for identifying hotspots may diverse. There are at least four parameters have been considered to identify hotspots: the number of users [1], [2], the mobile Internet traffic volume [3], [4], the visited times by users [5], [6] and the stay time of users [7], [8]. There are also more than one way to choose an appropriate threshold to determine whether one region is a hotspot. In paper [3], the author chooses five or ten times of

average traffic volume as the threshold. If the traffic volume of one region is higher than the threshold, the region become a hotspot. In paper [4] the author chooses the top of regions that the total traffic volume of them is more than 50 percent of all traffic volume as hotspots.

Recently, geo-located data generated by individuals have been widely used to study human behaviors [9]–[13] and urban social networks [14], [15]. Various data has been used to investigate user mobility behavior, such as Global Positioning System (GPS) data [16], location based online social network data [17], [18] and cellular network data [19]–[21]. Researchers found that data traffic from 2G/3G/4G data networks is extremely useful for studying human dynamics [22]. Passively collecting human movement trajectories while he/she is accessing to mobile Internet has lots of advantages: high cost efficiency, low energy consumption, covering a wide range and a large number of people, and with fine time granularity (people tend to surf mobile Internet frequently while moving, and many apps may send or receive network traffic packets periodically when running in the background).

Given these previous studies, the contributions of our research are as follows:

- We detect both population hotspots and traffic hotspots from six major cities in the province. We study the similarities and differences of hotspots characteristics in different cities. Based on the stability of hotspots, we classify the hotspots and find the most stable hotspots, which are called “constant hotspots”. We also find how people dependence on constant hotspots in different cities by a metric which is defined to measure the importance of constant hotspots.
- We investigate the spatial organization of cities from the view of hotspots and find some interesting results. Our research shows that it’s feasible to learn the spatial structure of cities using cellular network data.
- Our work focuses on the massive mobile Internet traffic data collected from cellular network. The real data used in our experiments was collected from a big province in northeast China over one month, covering more than 6 million users and more than 80 thousand base stations.

### III. METHODOLOGY

In this section, we introduce the procedure of cellular traffic data collection and then present a description of the dataset. Besides, we introduce our methodology for identifying and analyzing the hotspots in urban cities.

#### A. Data Collection

The cellular traffic data used in our experiments was collected from a major Chinese cellular network operator using our self-developed Traffic Monitoring System (TMS). TMS is a hardware and software platform used for network traffic monitoring. The high level view of the deployment of TMS in cellular network is shown in Fig.1.

There are three major components in the cellular network, including mobile devices, access network and core network.

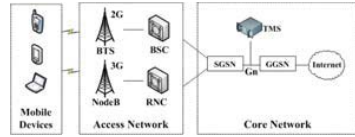


Fig. 1: Deployment of the Traffic Monitoring System in cellular network.

TABLE I: Basic Statistics of Six Cities

City	Users	Base Stations	Area( $km^2$ )
A	2347132	11849	1310
B	431551	4584	435
C	247942	2803	224
D	394371	4305	366
E	230679	2858	199
F	218113	3038	223

The TMS is deployed at the Gn interface between Serving GPRS Support Node (SGSN) and GPRS Gateway Support Node (GGSN). The captured traffic data is stored in the format of flow logs, each line of which comprises a sequence of time-stamped records. The records contain mobile device ID, timestamp, uplink traffic volume, downlink traffic volume, base station ID, and so forth.

#### B. Data Description

The data collection spans 31 days in October 2015, and includes data at more than 80,000 base stations that cover the entire province. We mainly focus on demographic and regional characteristics of urban areas, so we choose the urban areas of six major cities in the province to apply our analysis. Some basic statistics of these cities are shown in table I.

The six cities are represented by A, B, C, D, E and F. These urban cities are diverse in terms of number of users, number of base stations and area. City A is the capital of the province and has a high development level. It has much more users, base stations, and larger area than others cities. City B and D, with a moderate amount of users and base stations, have a medium development level. The development level in city C, E and F is relatively low. They have less users and base stations than city A, B and D.

In our experiment, we divide the urban areas to grid composed of  $1 km^2$  square cells on which we aggregated the density of unique users or mobile Internet traffic associated to each station.

#### C. Hotspots Identification

The key problem for identifying hotspots is parameter selection and threshold determination. Different parameter means different types of hotspots. In this paper, we use the density of population and generated traffic to identify two kinds of hotspots respectively. For threshold determination, a simple method is to choose a fixed threshold. If the population

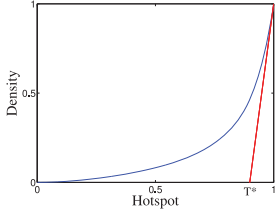


Fig. 2: Illustration of the criteria selection on the Lorenz curve. The threshold corresponds to the value of  $(1-T^*)$ .

or traffic density of one cell is larger than the threshold at time  $t$ , the cell is a hotspot at time  $t$ . This method has some arbitrariness in the choice of threshold: The density of each cell has significant variations through the day, a fixed threshold is not suitable for all 24 hours. So, we need an adaptive method. In our experiment, we employ a parameter free method proposed by Thomas Louail [1] to select the threshold. The method is based on the Lorenz curve. For a given hour, we sort the densities of each cell in increasing rank, and name them by  $(1, t) < (2, t) < \dots < (n, t)$  where  $n$  is the number of cells. The Lorenz curve is constructed by plotting on the x-axis the proportion of cells and on the y-axis the corresponding proportion of density with:

$$L(i, t) = \frac{\sum_{j=1}^i \sigma(j, t)}{\sum_{j=1}^n \sigma(j, t)} \quad (1)$$

where  $n$  is the number of cells and  $j \leq n$ . If all the densities have the same value, the Lorenz curve would be the diagonal from  $(0,0)$  to  $(1,1)$ . To determine a proper threshold, we take the intersection point  $T^*$  between the tangent of  $L(T)$  at point  $T = 1$  and the horizontal axis  $L = 0$  (see Fig.2). Then  $(1 - T^*) * n$  would be the number of hotspots.

#### IV. RESULTS ANALYSIS

We begin our analysis by describing the global characteristics of six cities. Then, we present the result of hotspots detection and the key characteristics of hotspots. In addition, we do some analyses of constant hotspots in the cities.

##### A. Global Characteristics

We first focus on the general features of the data. We plot the number of unique users and the traffic volume change with time in a day to see if they follow the same pattern in every city. As shown in Fig.3, from 10pm to 6am, the number of users is relatively low. It reaches a minimum at about 4am. The number of users is stable from 8am to 5pm, with a relatively high value. The varying pattern of traffic volume is a little different from the number of users. For all cities, we observe one peak at 8pm. We can learn that users of this province are more active during the daytime and they generate more traffic in the early night.

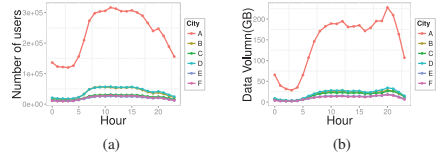


Fig. 3: the number of unique users (a) and the traffic volume (b) change with time in an average day for six cities.

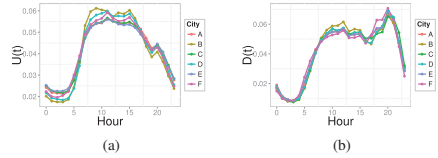


Fig. 4: the normalized number of unique users (a) and the normalized traffic volume (b) change with time in an average day for six cities.

In order to compare the values obtained from different cities, we normalize the values by the total number. For the number of users, value  $U(t)$  is equal to the number of unique users at time  $t$  divided by the number of unique users during the entire day:

$$U_i(t) = \frac{N_i(t)}{\sum_{t=1}^{24} N_i(t)} \quad (2)$$

We calculate the value of normalized traffic volume by the same way. The results are shown in Fig.4. The variation of normalized number of users follows the same pattern in all six cities, so does the variation of normalized traffic volume. As a result, we can conclude that, users living in different cities exhibit similar usage behavior of mobile Internet.

##### B. The Duration of Hotspots

We employ the method introduced in Section 3 to detect population hotspots and traffic hotspots in six cities at each hour of the average day.

Firstly, we study on the duration of each hotspot along the day. Urban cities usually have lots of functional regions that have chances to become hotspots in different hours. A workplace may become a hotspot during office hours, while a residential area may become a hotspot at night. Besides, some specific events can make one or more locations in the city become hotspots in particular hours. We count the duration of each hotspot and draw the duration distribution of hotspots. The results of city A are shown in Fig.5(a). The duration of most hotspots is more than 20 hours. Constant hotspots take up a great proportion of all hotspots. Besides, the amount of traffic hotspots is obviously lower than population hotspots.

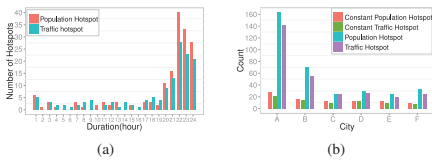


Fig. 5: The number of population hotspots and traffic hotspots in each duration for city A (a) and average number of population hotspots, traffic hotspots, constant population hotspots and constant traffic hotspots in one day for six cities (b).

That means the distribution of traffic is more unequal than the distribution of population.

Then, in order to calculate the proportion of constant hotspots for all hotspots, we count the number of unique cells that have ever become hotspots at any hour as the number of hotspots in one day. The results of six cities are shown in Fig.5(b). We can see that the proportion of constant hotspots for all hotspots is extremely high in city C, E and F, whose development level is relatively low. And in city A, whose development level is highest in all cities, a place have more chance to become occasionally generated hotspot.

### C. Measure the Importance of Constant Hotspots

Constant hotspots play a vital role in the cities. In order to study the role of constant hotspots, measuring the importance of constant hotspots becomes crucial. For this, we define an indicator called “gathering coefficient”. For population, the value of gathering coefficient is equal to the number of users who have ever appeared in any constant hotspots divided by the number of total users. And for traffic, it is equal to the traffic volume generated in constant hotspots divided by total traffic volume.

The values of gathering coefficient for six cities are shown in Fig.6. We can find that gathering coefficient of traffic is a little higher than gathering coefficient of population. Besides, the gathering coefficient in city A is the lowest among six cities both in population and traffic. It implies that the importance of constant hotspots in city A is relatively low comparing with other cities. We can also learn that the importance of constant hotspots is related with the development level of the city. The gathering coefficient is higher in cities whose development level is relatively low. In a small city, people depend more on those constant hotspots, which play a crucial role in their life. However, in high developing cities, people have less dependence on constant hotspots. The development of cities leads to the emergence of some dispersed functional areas. Although those functional areas haven’t become a constant hotspots yet, they can meet lots of people’s daily needs.

### D. Organization of City

Due to the different states of development in each city, the organization of these cities may diverse. One interesting

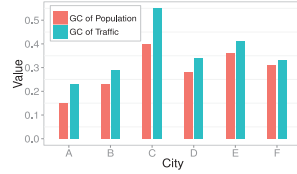


Fig. 6: Values of gathering coefficient of population and traffic for six cities.

question is whether we can learn the spatial structure of cities using cellular data. In order to answer the question, we study the spatial distribution of hotspots and individuals in the cities.

We use  $S_p(I)$  and  $S_t(I)$  to represent the average distance between constant population hotspots and the average distance between constant traffic hotspots in city  $I$ . Monocentric cities whose important places gathered in the central city should have a low value of the  $S_p(I)$  and  $S_t(I)$ . Polycentric cities, in the opposite, may have a relatively high value. Besides, the variation of average distance between individuals in a day also provides an opportunity to study the organization of the city. The weighted average distance between individuals  $D_V(t)$  is defined as:

$$D_V(t) = \frac{\sum_{i < j} s_i(t) s_j(t)}{\sum_{i < j} s_i(t) s_j(t) d_{ij}} \quad (3)$$

where  $d_{ij}$  is the distance between base station  $i$  and base station  $j$ , and  $s_i(t) = n_i(t)/N(t)$ , where  $n_i(t)$  is the number of users in base station  $i$  at time  $t$ ,  $N(t)$  represents the total number of users at time  $t$ .

We consider two typical cases. In monocentric cities with predominant central areas, the majority of workplaces and entertainment venues located in a few areas in the city center and most of residents live in the suburbs. The city collapses in the morning when people living in the suburbs gathering to the city center for work or entertainment, and expands in the evening when they get back home. As for polycentric cities, where workplaces, entertainment venues and residential areas are spatially less separated, have little variation in  $D_V$ .

We use an indicator called “dilatation coefficient” [1] to measure the variation of the average distance between individuals, which is defined as:

$$u = \frac{\max_t(D_V(t))}{\min_t(D_V(t))} \quad (4)$$

Here,  $D_V(t)$  represents the weighted average distance between individuals at time  $t$ .

In this paper, we use both the average distance between constant hotspots and dilatation coefficient to investigate the organization of cities. The results for six cities are shown in Fig.7. We can find that the dilatation coefficient has negative correlation with the average distance between constant





Fig. 7: Average distance between constant hotspots (km) and the dilatation coefficient for six cities.

hotspots. From the view of  $S_p$  and  $S_t$ , city A and B are more like polycentric cities and city D, E, C, F are more like monocentric cities. Besides, the dilatation coefficient of city A, B and D are lower than 1.5, which shows the feature of polycentric cities. And the dilatation coefficient of city C and F are more than 2, which implies that they are more like monocentric cities. For those two reasons, we suggest that city A and B are polycentric cities and city C and F are monocentric cities. As a result, we can conclude that the spatial organization of cities are related to the cities' state of development. A highly developed city is more likely to have more constant hotspots and become a polycentric city.

V. CONCLUSION

In the paper, we study hotspots in six major cities in one great province of northeast China using a one-month province-wide dataset collected from a major Chinese cellular operator. We show that the number of users and traffic volume change with time in a day follow similar pattern in all six cities. Then, we deploy an adaptive method to detect population hotspots and traffic hotspots in those cities. Based on the results of hotspots detection, our analysis led to several key findings. It is concluded that both population hotspots and traffic hotspots have high stability in urban cities. Besides, based on analysis of constant hotspots, we discover that the gathering degree of individuals and traffic are both smaller in underdeveloped cities, which means people are more dependent on constant hotspots in small cities. Finally, the relation between the spatial structure and the state of development in cities are studied by investigating the spatial distribution of constant hotspots and individuals. The results show that it's feasible to learn the spatial structure of cities using cellular data.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China (61671078), the Fundamental Research Funds for the Central Universities (2015RC11) , Director Foundation Project (2015BKL-NSAC-ZJ-01), 111 Project of China (B08004), and EU FP7 IRSES MobileCloud Project (Grant No. 612212).

REFERENCES

[1] T. Louail, M. Lenormand, O. G. Cantu Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy, "From mobile phone data to the spatial structure of cities." *Physics*, vol. 4, no. 2973, pp. 61–61, 2014.

[2] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces." *Proceedings - IEEE INFOCOM*, vol. 6, pp. 1–13, 2006.

[3] H. Klessig, V. Suryaprakash, O. Blume, A. Fehske, and G. Fettweis, "A framework enabling spatial analysis of mobile traffic hot spots." *IEEE Wireless Communication Letters*, vol. 3, no. 5, pp. 537–540, 2014.

[4] A. Nika, A. Ismail, B. Y. Zhao, and S. Gaito, "Understanding data hotspots in cellular networks," in *International Conference on Heterogeneous NETWORKING for Quality, Reliability, Security and Robustness*, 2014, pp. 70 – 76.

[5] M. Fort, J. A. Sellars, and N. Valladares, "Computing and visualizing popular places," *Knowledge & Information Systems*, vol. 40, no. 2, pp. 411–437, 2013.

[6] S. Hoteit, S. Secci, G. Pujolle, S. Sobolevsky, and C. Ratti, "Estimating human trajectories and hotspots through mobile phone data." *Computer Networks the International Journal of Computer & Telecommunications Networking*, vol. 64, no. 4, pp. 296–307, 2014.

[7] D. Ashbrook and T. Starmer, "Using gps to learn significant locations and predict movement across multiple users." *Personal & Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.

[8] T. M. T. Do and D. Gatica-Perez, "Contextual conditional models for smartphone-based human mobility prediction," in *ACM Conference on Ubiquitous Computing*, 2012, pp. 163–172.

[9] K. Zhao, M. Musolesi, P. Hui, W. Rao, and S. Tarkoma, "Explaining the power-law distribution of human mobility through transportation modality decomposition." *Scientific Reports*, vol. 5, 2014.

[10] S. Lee, J. Lim, J. Park, and K. Kim, "Next place prediction based on spatiotemporal pattern mining of mobile device logs," *Sensors*, vol. 16, no. 2, 2016.

[11] M. Zignani, M. Papandrea, S. Gaito, and S. Giordano, "On the key features in human mobility: Relevance, time and distance," in *IEEE International Conference on Pervasive Computing and Communication Workshops*, 2014, pp. 46–58.

[12] P. Pirozmand, G. Wu, B. Jedari, and F. Xia, "Human mobility in opportunistic networks: Characteristics, models and prediction methods," *Journal of Network & Computer Applications*, vol. 42, no. 3, pp. 45–58, 2014.

[13] M. Dash, K. K. Koo, S. P. Krishnaswamy, and Y. Jin, "Visualize people's mobility - both individually and collectively - using mobile phone cellular data," in *IEEE International Conference on Mobile Data Management*, 2016.

[14] C. Herrerayage, C. M. Schneider, T. Couronn, Z. Smoreda, R. M. Benito, P. J. Zufria, and M. C. Gonzalez, "The anatomy of urban social networks and its implications in the searchability problem," *Scientific Reports*, vol. 5, 2015.

[15] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, "Distance matters: geo-social metrics for online social networks," in *Wanference on Online Social Networks*, 2010, pp. 8–8.

[16] R. Lambiotte, V. D. Blondel, C. D. Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren, "Geographical dispersal of mobile communication networks," *Physica A Statistical Mechanics & Its Applications*, vol. 387, no. 21, pp. 5317–5325, 2008.

[17] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," in *International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July*, 2011.

[18] N. Li and G. Chen, "Sharing location in online social networks," *IEEE Network Magazine of Global Internetworking*, vol. 24, no. 5, pp. 20–25, 2010.

[19] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July*, 2011.

[20] S. Isaacman, R. Becker, R. Caceres, and S. Kobourov, "Ranges of human mobility in los angeles and new york," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2011, pp. 88–93.

[21] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, and B. Lepri, "The death and life of great italian cities: A mobile phone data perspective," in *International ACM Conference on World Wide Web*, 2016.

[22] Y. Zhang, "User mobility from the view of cellular data networks," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, 2014, pp. 1348–1356.

# Resource Scheduling Approach for LTE-A based Network Incorporating a Moving Relay Node System Equipped Train

Ayotunde O. Laiyemo, Harri Pannanen, Pekka Pirinen, and Matti Latva-aho  
 Centre for Wireless Communications (CWC), University of Oulu, Finland  
 P.O. Box 4500, FI-90014 University of Oulu  
 email: forename.surname@oulu.fi

**Abstract**—Recent trends show that mobile users accessing the wireless communication network from public transportation such as the high speed train are on the increase. To serve onboard users effectively, the use of moving relay nodes in a two-hop communication link is a promising solution. However, the direct use of known resource scheduling methods will not be appropriate to efficiently and fairly share resources between ground macro users (GMUs) and moving relay nodes (MRNs) used on the high speed trains. In order to address this challenge, we examine two hybrid resource scheduling methods based on system level simulations and analyze joint and disjoint scheduling approaches for efficient and fair resource sharing between GMUs and MRNs on the downlink of a long term evolution advanced (LTE-A) cellular communication system.

## I. INTRODUCTION

Demand for cellular broadband wireless communications by high mobility users has tremendously increased due to the rapid deployment of large public transport vehicles such as the high speed train (HST) in many parts of the world. Providing high data rates and good quality of service (QoS) required by onboard users in an HST is a challenging task in the presence of rapidly varying channel conditions. An effective way to improve data rate is to take advantage of the spatial dimension of the high speed train using a two-hop system architecture and applying advanced multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) techniques [1]. One advantage of the OFDM technique is the ability to split the system bandwidth into several orthogonal channels, thereby providing the possibility of sharing radio resources to multiple users within the same time slot. Maximizing the downlink throughput of moving relay nodes installed on a HST within LTE-A cellular communication system was studied in [2]–[5]. However, the issue of fairness and the impact on the throughput of macro users scheduled along with the MRNs were not addressed. Various scheduling strategies have been proposed for traditional OFDM systems. For example, studies in [6] demonstrate that enhanced system throughput and fairness among users can be achieved with maximum carrier to interference ratio scheduling. Proportional fair scheduling was proposed for OFDM systems in [7]–[9] where fairness among users was addressed while maximizing system throughput. With the introduction of relaying concept into the cellular network, some studies have addressed the issue of scheduling with the presence of relays such as in [10]–[12]. To the

authors' knowledge, no contribution has been done on the investigation and impact of scheduling macro users along with MRNs installed on a HST to serve passengers. The notion of the employing MRN systems for HST was presented in [13]. The idea is to deploy interconnecting individual moving relay node in each of the carriages, which connects to the donor cellular network(s) with multiple wireless backhaul links in a coordinated and controlled manner. This concept ensures that the access link is an integrated extension of the cellular network and sufficient high data rates are achieved due to the multiple backhaul links established.

In this study, we develop a system model for the downlink cellular network with a train equipped with an MRN on each carriage of the train. The radio resources available in each cell, the train is present are shared fairly with other GMUs with the assumption that onboard users are only served by the MRNs on the access link. We analyze the effect of joint and disjoint scheduling approaches. We also derive two hybrid scheduling algorithms and show that disjoint scheduling approach leads to a fair balance in throughput performance for both the GMUs and the MRNs. The LTE framework is adopted and simulations are carried out on an LTE-A compliant system level simulator.

The rest of the paper is organized as follows. In Section II, the system model is introduced. The scheduling problem within the LTE framework is presented in Section III. Section IV presents the two hybrid scheduling algorithms and scheduling approaches. Section V describes the simulator and the results are shown in Section VI with Section VII concluding the paper.

## II. SYSTEM MODEL

We consider a train with a two-hop network system architecture created with the use of MRNs in a multi-cell system as shown in Fig. 1. The train consists of 8 carriages, each equipped with an MRN and having an external antenna array that is evenly spaced along the length of the train and interior antennas to serve onboard users. The multi-cell system is modelled as an hexagonal layout consisting of 19 trisector cells with replicas of the layout wrapped around the main layout as shown in Fig. 1. This is to ensure uniform interference levels across the main layout. Each sector operating at 2 GHz band with a 10 MHz bandwidth has  $N_t$  transmit antennas transmitting on  $L$  transmission layers with the MRN having

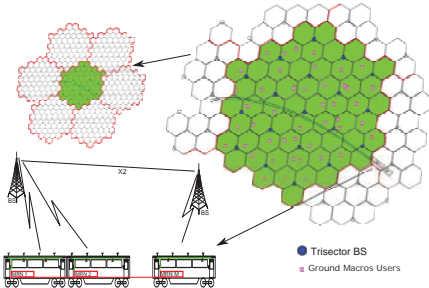


Fig. 1: High speed train in practical cellular system.

$N_{mrn}$  receive antennas. Signals from onboard users interfering with the MRN's external antenna array are omitted. We also model ground macro users (GMUs) with low mobility, each having  $N_r$  receive antennas. WINNER II channel model [14] is adopted as the fast fading channel model.

In this scenario, MIMO-OFDM based system is considered, operating in frequency division duplexing (FDD) mode and assuming that the MRN operates in half duplex mode. In the case of transmission to the GMUs, transmit diversity transmission scheme is implemented. Correspondingly, spatial multiplexing transmission scheme is used for the MRNs since we can take advantage of the spatial dimension of the HST. Each GMU can only communicate directly with the BS and users on the HST can only communicate with the BS via the MRNs. The BS is equipped with a resource scheduler, which allocates the available resources to the possible connections according to a scheduling method. Due to the half duplex operation of the MRNs, the MRNs are scheduled at every second time slot. The receive signal vector for the MIMO transmission  $\mathbf{y}_c \in \mathbb{C}^N$  for subcarrier  $c$  can be expressed as

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{s}_c + \sum_{k=1}^K \tilde{\mathbf{H}}_{k,c} \mathbf{s}_{k,c} + \mathbf{n}_c \quad (1)$$

where  $\mathbf{H}_c \in \mathbb{C}^{N \times N_t}$  is the channel matrix for subcarrier  $c$  of the desired signal with dimension  $N \in \{N_{mrn}, N_r\}$  and  $\mathbf{s}_c \in \mathbb{C}^{N_t}$  is the desired transmit signal vector per subcarrier.  $\tilde{\mathbf{H}}_{k,c}$  and  $\mathbf{s}_{k,c}$  are the channel matrix and the transmit signal vector from interfering sector  $k$  to the MRN on subcarrier  $c$ ,  $K$  is the total number of the interfering sectors, and  $\mathbf{n}_c$  is the additive white Gaussian noise vector. The transmitted signal vector is estimated as  $\hat{\mathbf{s}}_c = \mathbf{W}_c \mathbf{y}_c$ , where  $\mathbf{W}_c = (\hat{\mathbf{H}}_c \hat{\mathbf{H}}_c^H + \mathbf{R}_c)^{-1} \hat{\mathbf{H}}_c$  is the optimal linear minimum mean square error (MMSE) filter and  $\mathbf{R}_c = \sum_{k=1}^K \tilde{\mathbf{H}}_{k,c} \Sigma_{N_t} \tilde{\mathbf{H}}_{k,c}^H + N_0 \mathbf{I}_N$  is the interference plus noise covariance matrix.  $\Sigma_{N_t}$  is a  $N_t$ -dimensional diagonal matrix whose diagonal elements share equally the total transmission power  $p_{c,k}$  from the  $k^{\text{th}}$  interfering sector. The corresponding MRN's SINR per transmission stream for subcarrier  $c$  at the output of the MMSE receiver can be expressed as

$$\gamma_{c,l} = \frac{p_{c,l} |\mathbf{w}_{c,l}^H \hat{\mathbf{H}}_c|^2}{\sum_{j=1, j \neq l}^L p_{c,j} |\mathbf{w}_{c,j}^H \hat{\mathbf{H}}_c|^2 + \mathbf{w}_{c,l}^H \mathbf{R}_c \mathbf{w}_{c,l}} \quad (2)$$

where  $\hat{\mathbf{H}}_c$  is the estimate of  $\mathbf{H}_c$ ,  $p_{c,l}$  denotes the transmission power for the  $l^{\text{th}}$  stream at subcarrier  $c$  and  $\mathbf{w}_{c,l}$  denotes the  $l^{\text{th}}$  column of  $\mathbf{W}_c$ .

### III. SCHEDULING PROBLEM

GMUs and MRNs are associated with any of the 57 BSs in the central layout based on variation in the instantaneous channel conditions, shadow fading and distance-dependent path loss. Hence, with the possibility of having more than one GMU/MRN associated with a BS, scheduling is achieved based on the channel quality of each GMU/MRN. Scheduling ensures that the available radio resources are shared with as many users as possible, while still satisfying GMU/MRN's QoS requirements. The throughput of each GMU/MRN and the throughput of the entire network are affected by the type of scheduler used. The type of scheduling method used also influences the impact of the HST on the throughput of the GMUs. Hence, evaluating the efficiency of different scheduling methods as it relates to the GMUs and MRNs, is required.

We consider an LTE-A system with a radio frame structure, such that each radio frame is divided into 10 subframes. Resources are allocated to users on a subframe basis at every 1 ms transmission time interval (TTI), (i.e., the scheduler allocates resources to GMU/MRN in minimum portions of two consecutive resource blocks (RB)), due to signalling overhead limitations. One RB consists of 12 adjacent subcarriers and corresponds to 0.5 ms (i.e.,  $\frac{1}{2}$  of a subframe) in the time domain. In any given cell, GMUs are always allocated resources at every TTI. However, MRNs are allocated resources at every second TTI, since we assume half duplex functionality of the MRNs. Let us assume a time slot  $t$  equals 1 TTI (1 ms) and the overall system bandwidth is  $B$  MHz. With a scheduling block (SB) corresponding to two consecutive RBs, the bandwidth of each SB is  $\frac{B}{N_{SB}}$ , where  $N_{SB} = 2N_{RB}$  is the number of SBs within a specified bandwidth and  $N_{RB}$  corresponds to the number of RBs. In every TTI, scheduling is done across  $N_{SB}$  scheduling blocks and we assume uniform distribution of the total transmit power across all subcarriers. The capacity of the  $r^{\text{th}}$  scheduling block for the  $m^{\text{th}}$  GMU/MRN on the  $t^{\text{th}}$  TTI that establishes a connection link is given by

$$C_{m,r}(t) = \frac{B}{N_{SB}} \log_2 \left( 1 + \sum_{c=1}^{C^r} \sum_{l=1}^L \gamma_{c,l} \right) \quad (3)$$

where  $C^r$  is the number of subcarriers in the  $r^{\text{th}}$  scheduling block. Let  $\mathbb{1}_{t,\hat{m}}$  denote an indicator function that describes GMUs/MRN's that are eligible to be allocated resources such that

$$\mathbb{1}_{t,\hat{m}} = \begin{cases} 1 & \text{if } (t_{ind} \in [1, 2] \wedge \hat{m} \in \text{GMU}) \vee (t_{ind} \in [1] \wedge \hat{m} \in \text{MRN}) \\ 0 & \text{Otherwise} \end{cases}$$

where  $t_{ind} = (t \bmod 2) + 1$  is an index that ensures an MRN is not scheduled on every second TTI and  $\hat{m}$  is the GMU/MRN being allocated resources on the  $r^{\text{th}}$  scheduling block at the  $t^{\text{th}}$  TTI. The total achievable rate up to  $T$  TTIs in the  $b^{\text{th}}$  cell can be calculated as

$$R_{tot}(b) = \sum_{\hat{m} \in \mathcal{U}} \sum_{t=1}^T \sum_{r=1}^{N_{SB}} \mathbb{1}_{t,\hat{m}} C_{\hat{m},r}(t) \quad (4)$$

where  $\mathcal{U}$  is the set of GMU/MRNs associated with the same BS. The overall throughput across the network, i.e.,  $\sum_{b \in \mathcal{B}} R_{tot}(b)$  is maximized using a scheduling method, where  $\mathcal{B}$  is the set of BSs within the network. However,  $R_{tot}(b)$  is a combination of the throughputs for GMUs and MRNs, i.e.,  $R_{tot}(b) = R_{gmu}(b) + R_{mrn}(b)$ . Therefore, combined scheduling of both GMUs and MRNs may not necessarily ensure fairness between the two groups of users. To ensure adequate fairness between the two groups of users with minimal impact on the throughput, we analyze the performance of joint and separate scheduling of GMUs and MRNs using two hybrid scheduling methods.

#### IV. RESOURCE ALLOCATION METHODOLOGY

##### A. Hybrid Scheduling Algorithms

Here, we consider two hybrid scheduling algorithms based on existing algorithms with the aim of optimizing the scheduling performances.

1) *Round robin with max rate (RRMR)*: This algorithm is a combination of round robin scheduling algorithm and maximum rate scheduling algorithm to bring about a balance between fairness and performance. The RRMR scheduling brings a balance between fairness among users and scheduling users with the highest achievable rate. In one TTI, users are scheduled across  $N_{SB}$  scheduling blocks. The user with the best channel condition is first allocated resources on a corresponding pair of scheduling blocks, then the user with second best channel condition is allocated resources on the next available pair of scheduling blocks. This process continues for all users associated with the BS. After all the users have been allocated resources, the first user with the best channel condition is assigned two more scheduling blocks. This process continues until there are no more available scheduling blocks or until the target rate of each user is reached. More details are given in Fig. 2. The principle is to allocate a small number of SBs to all users cyclicly in a predetermined order based on maximum rate. Hence, the total SBs are shared as evenly as possible on a best performance basis, without unduly reducing the throughput.

2) *Modified Proportional fair (MPF)*: This algorithm is a slight modification of the proportional fair algorithm. The proportional fair scheduling algorithm provides a balance between optimized throughput and fairness among the users. The balance provided by the algorithm is as a result of taking into account the current channel conditions and transmission history of the users. In addition to the PF, the MPF also provides a balance between fairness and performance such that the selected scheduling block for a user is based on the user with the best available rate on the scheduling block and the minimum average available rate of all the users associated with the BS. Therefore, at each TTI, resources are allocated

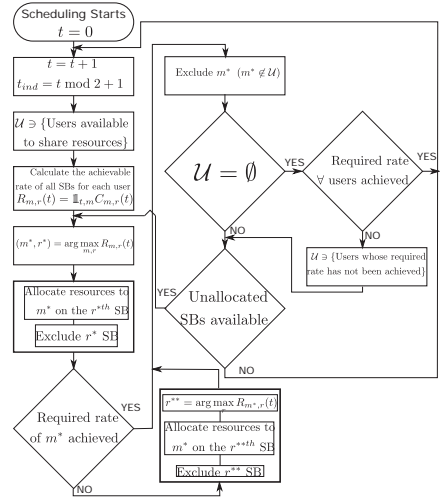


Fig. 2: RRMR flow chart.

to users with the relatively best channel condition such that a user is allocated resources on a scheduling block according to

$$\hat{m}^* = \arg \max_{\hat{m}} \frac{R_{\hat{m},r^*}(t)}{\hat{R}_{\hat{m},r^*}(t) + \epsilon} \quad (5)$$

where  $R_{\hat{m},r^*}(t) = \mathbb{1}_{t,\hat{m}} C_{\hat{m},r^*}(t)$  is the fed back data rate for user  $\hat{m}$ , with the selected scheduling block  $r^*$  described as

$$r^* = \arg \min_r \left( \sum_{\hat{m} \in \mathcal{U}} R_{\hat{m},r^*}(t) \right) / |\mathcal{U}|. \quad (6)$$

$\hat{R}_{\hat{m},r^*}(t)$  is the moving average throughput on the  $r^{\text{th}}$  scheduling block for user  $\hat{m}$  and  $\epsilon$  is a small positive constant to prevent the error of dividing by zero. The moving average throughput (transmission history) is calculated over a certain proportional window length  $P_w$  and updated as

$$\hat{R}_{\hat{m},r^*}(t+1) = \left( 1 - \frac{1}{P_w} \right) \hat{R}_{\hat{m},r^*}(t) + \frac{1}{P_w} R_{\hat{m},r^*}(t). \quad (7)$$

##### B. Scheduling Approaches

We consider the joint and disjoint scheduling approaches w.r.t. the two groups of users, i.e., GMUs and MRNs.

1) *Joint Scheduling*: In the joint scheduling approach, the GMUs and MRNs are scheduled together. All users associated with a BS are given equal priority. Hence, for both scheduling algorithms in Section IV-A, the set  $\mathcal{U}$  which consist of users available to share resources include both GMUs and MRNs.

2) *Disjoint Scheduling*: The disjoint scheduling approach identifies and separates the types of users before applying the scheduling algorithm. Hence, the GMUs are scheduled separately from the MRNs. However, scheduling priority has to be established. The idea of scheduling with priority is to either schedule the GMUs first or the MRNs first. The separate scheduling of the GMUs and MRNs gives the possibility of using different scheduling methods for

each group of users, thereby enhancing the flexibility of the achievable rate and fairness between the groups.

V. SIMULATOR DESCRIPTION

The LTE-A based system level simulator was configured to follow the guidelines established by the international telecommunications union radiocommunications sector (ITU-R) for IMT-A radio interface evaluation [15]. The simulation parameters were set to closely follow the LTE system. The central layout with inter-site distance of 1.3 km was modelled as shown in Fig. 1 and the antenna configurations for  $N_t$ ,  $N_{mrn}$ , and  $N_r$  were set to 4, 4 and 2, respectively. The simulation runs consist of 1000 drops, with 100 GMUs randomly distributed across the central layout at the start of each drop. A track with a radius of approximately 5 km is placed across the central layout such that the minimum distance between the track and any BS is 50 m. At the start of the first drop, the train is positioned on one end of the track and as the drops go on, the train moves along the track to the other end. At the start of each drop, each of the GMUs and the train equipped with MRNs are paired with 57 cells that provide the strongest received signal strength. The received signal strength is calculated based on path-loss distance and angular antenna gain.

Scheduling of mobile users (GMUs and MRNs) at the BS rely on channel state information (CSI) feedback, i.e., the channel quality indicator (CQI), which is derived from the mobile users and is made available at the BS after a delay. A target rate of 10 Mbps is set for the GMUs, and full buffer traffic model is considered for the MRN with the assumption that the number of users in the train can be large. A link to system (L2S) interface is employed such that the SINRs obtained are mapped to mutual information values using mutual information effective SINR metric (MIESM) link layer abstraction [16]. A modulation and coding scheme (MCS) is set for each mobile user and the MCS values determine the frame error probability (FEP) at the link to system level interface and the transport block size, i.e., the number of bits transmitted for throughput calculations. Successful transmissions/retransmissions are identified by hybrid automatic repeat request (HARQ) acknowledgements, which are determined in the system level interface and fed back to the BS after a delay. For mobile users with successful transmissions/retransmissions, the number of correctly received bits is calculated and used in throughput calculations.

VI. SIMULATION RESULTS

We provide simulation results according to our model for downlink transmission. Fig. 3 gives the cumulative distribution function (CDF) of the GMU/MRN throughput for different scheduling approaches using the RRRM scheduling algorithm. The following scheduling approaches are considered:

- RRRM algorithm giving MRN priority (RRRM MRN)
- RRRM algorithm giving GMU priority (RRRM GMU)
- RRRM algorithm with joint scheduling (RRRM Joint).

With the RRRM GMU approach, the total GMU throughput in Fig. 3(a) improves significantly compared to the joint and

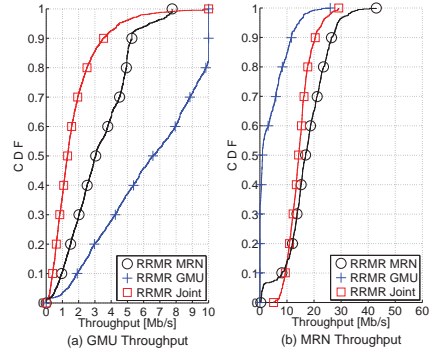


Fig. 3: CDF throughputs with RRRM scheduling method.

MRN priority approaches. However, in Fig. 3(b), the MRNs are in outage for a significant part of the journey. On the other hand, when the GMUs and MRNs are jointly scheduled (i.e., RRRM Joint), the total MRN throughput in Fig. 3(b) is significantly improved with no outage. But the total GMU throughput for RRRM Joint approach suffers a significant reduction, even when compared to RRRM MRN approach. This is because, in a given cell the MRNs are more likely to have better channel conditions than the GMUs and as a result, the GMUs are most often not allocated the best resources. The MRNs will most often have better channel conditions because the antenna heights are much higher than the GMUs' since the antennas are mounted on the top of the train. In a case where the GMUs are much closer to the BS than the MRNs, the GMUs can have better channel conditions. This explains what happens, when RRRM MRN and RRRM Joint approaches are compared at the cell edge in Fig. 3(b).

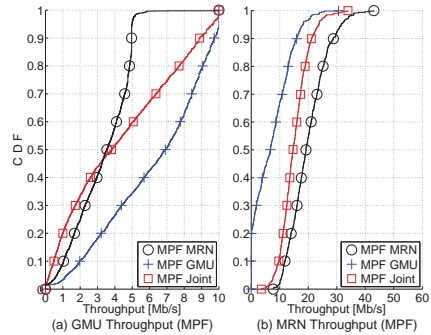


Fig. 4: CDF throughputs with MPF scheduling method.

The MPF scheduling method is applied using the scheduling approaches discussed with the CDF throughputs shown in Fig. 4. The following scheduling approaches are considered:

- MPF algorithm giving MRN priority (MPF MRN)
- MPF algorithm giving GMU priority (MPF GMU)
- MPF algorithm with joint scheduling (MPF Joint).

As expected, With the MPF GMU approach, the total GMU throughput (Fig. 4(a)) is significantly improved as compared to the other two approaches, but the total MRN throughput (Fig. 4(b)) is poor with outage. The total MRN throughput for the other two approaches show that giving MRN priority (i.e., MPF MRN approach) can achieve about 5 Mbps improvement as compared with joint scheduling (MPF Joint). However, considering Fig. 4(a), the total GMU throughput for MPF Joint approach gets a large improvement in throughput compared to MPF MRN approach from the 50% point of the CDF, but at the cell edge, the MPF MRN approach performed better than MPF Joint approach. This is because in the MPF Joint approach, the MRN that is jointly scheduled brings an increase in interference to the cell edge GMUs. From Fig. 3 and 4, it can be seen that scheduling MRNs first, brings a fair balance in the GMU and MRN throughputs. Due to the ability to separate the scheduling of each group of users, we can use different scheduling algorithms for each group of users with the aim to improve performance on both throughputs.

Fig. 5 shows the throughput CDF performances of GMUs and MRNs, where MPF scheduling algorithm is applied for the GMUs and maximum rate (MR) scheduling algorithm is applied for the MRNs. We call it the MPF-MR scheduling algorithm and compare scheduling GMU first (MPF-MR GMU) with scheduling MRN first (MPF-MR MRN).

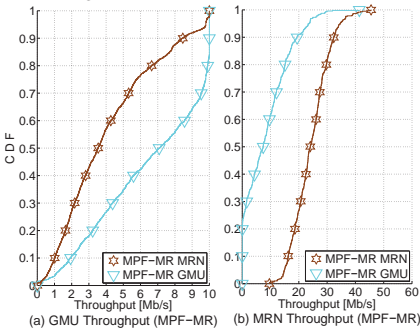


Fig. 5: CDF throughput with MPF scheduling for GMU and MR scheduling for MRN.

The results show that to avoid outages on the train, it is better to schedule MRNs first. The total MRN throughput as seen in Fig. 5(b) shows that giving MRN priority with MR algorithm can further improve throughput performance by about 5 Mbps compared to MPF MRN scheduling approach in Fig. 4(b) with minimal impact on the total GMU throughput performance (Fig. 5(a)) when compared with MPF Joint scheduling approach in Fig. 4(a). The gain of using MR scheduling algorithm on the MRNs stems from the fact that the train with multiple carriages equipped with an MRN in each carriage can be made to cooperate in a coordinated fashion.

VII. CONCLUSION

In this paper, we have provided a system model to evaluate the throughput performance of both GMUs and

MRNs mounted on a train in a communication network under different resource scheduling approaches. We also proposed two hybrid scheduling algorithms which were used to implement the scheduling approaches. The simulations were carried out on an LTE-A compliant system level simulation platform. Results show that Joint scheduling does not provide the best overall performance and there is a need to schedule each group of users separately. Specifically, scheduling MRNs first gives a fair balance in throughput performance and improved performance can be achieved by taking advantage of the MRN cooperation that can be achieved on the train.

ACKNOWLEDGMENT

This research was supported by the Finnish Funding Agency for Technology and Innovation (TEKES), Nokia, Keysight Technologies, and Huawei Technologies.

REFERENCES

- [1] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.
- [2] S. Scott, J. Leinonen, P. Pirinen, J. Vihriälä, V. Van Phan, and M. Latva-aho, "A cooperative moving relay node system deployment in a high speed train," in *Proc. IEEE Veh. Technol. Conf.*, Jun. 2013, pp. 1–5.
- [3] A. O. Laiyemo, P. Pirinen, and M. Latva-aho, "Alternative to dynamic rank transmission for LTE mobile relay node system," in *Proc. European Wireless Conf.*, May 2014, pp. 1–6.
- [4] A. O. Laiyemo, P. Pirinen, M. Latva-aho, J. Vihriälä, and V. Van Phan, "Impact of precoding for fixed and adaptive rank transmission in moving relay node system," in *Proc. IEEE ITS Telecommunications (ITST)*, IEEE, 2013, pp. 250–254.
- [5] A. Laiyemo, H. Pannanen, P. Pirinen, and M. Latva-aho, "Transmission strategies for throughput maximization in high speed train communications: From theoretical study to practical algorithms," *IEEE Trans. Veh. Technol.*, in press.
- [6] L.-C. Wang and W.-J. Lin, "Throughput and fairness enhancement for OFDMA broadband wireless access systems using the maximum C/I scheduling," in *Proc. IEEE Veh. Technol. Conf.*, vol. 7, IEEE, 2004, pp. 4696–4700.
- [7] H. J. Zhu and R. H. Hafez, "Scheduling schemes for multimedia service in wireless OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 99–105, 2007.
- [8] N. Ruangchaiatupon and Y. Ji, "Simple proportional fairness scheduling for OFDMA frame-based wireless systems," in *Proc. IEEE Wireless Commun. and Networking Conf.*, IEEE, 2008, pp. 1593–1597.
- [9] X. Liang, Z. Shidong, and Y. Yan, "Dynamic resource management in the fourth generation wireless systems," in *Proc. Commun. Technol.*, vol. 2, IEEE, 2003, pp. 1095–1098.
- [10] N. Challal and H. Cam, "Cost-aware downlink scheduling of shared channels for cellular networks with relays," in *Proc. IEEE Performance, Computing, and Commun.*, IEEE, 2004, pp. 793–798.
- [11] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2328, 2005.
- [12] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource scheduling for OFDMA/TDD based relay enhanced cellular networks." *Information Society Technologies*, Feb. 2009.
- [13] V. Van Phan, K. Horneman, L. Yu, and J. Vihriälä, "Providing enhanced cellular coverage in public transportation with smart relay systems," in *Proc. IEEE Veh. Networking Conf.*, Dec. 2010, pp. 301–308.
- [14] J. Meinilä, P. Kyösti, T. Jämsä, and L. Hentilä, "WINNER II channel models," *Information Society Technologies*, Feb. 2009.
- [15] M. Series, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," ITU, Tech. Rep., 2009.
- [16] X. He, K. Niu, Z. He, and J. Lin, "Link layer abstraction in MIMO-OFDM system," in *Proc. IEEE Int. Workshop on Cross Layer Design (IWCLD)*, Sep. 2007, pp. 41–44.

# Motion Estimation in Video Sequences by Using Region Based and Optical Flow Algorithm

Narasak Boonthep and Kosin Chamnongthai

Department of Electronic and Telecommunication Engineering

King Mongkut's University of Technology Thonburi

Bangkok, Thailand

narasakb@gmail.com and kosin.cha@kmutt.ac.th

**Abstract**—This paper proposes an algorithm of motion estimation in video sequences using region based and optical flow. Scale-invariant feature transform (SIFT) is basically applied to detect keypoints to using the search starting point in the region selection process, which are utilized as landmarks for motion tracking. Optical flow algorithm is used to track the object pixel. In optical flow, iterated two-step optimization framework and model parameter are subsequently employed for efficient prediction of appropriate area. Experimental results show the proposed algorithm is efficient and effective with satisfied accuracy. It can improve 0.2-0.3 dB.

**Keywords**— *Motion estimation; Optical flow; Optimization; SIFT*

## I. INTRODUCTION

Motion estimation (ME) is an important part of data compression methods adopted by all existing video coding standards, such as H.263, MPEG-4, etc. The idea behind video compression based on motion estimation is to save number of bits required for encoding the video. The purpose for doing motion estimation is to reduce the energy and bandwidth requirement for transmission of videos over wireless medium. Therefore, how to reduce the ME computational complexity while still maintaining almost the same coding efficiency has become an important issue.

Video sequences contain huge amount of data which has a high degree of the redundant data. It should be eliminated by using ME. ME has been used to reduce temporal redundancy, and disparity estimation (DE) has been employed to reduce inter-view redundancy. The ideal of ME is to track every pixel from frame to frame. However, it requires very expensive computational cost. Since this system has serious limitations on data distribution applications, such as broadcasting, multimedia streaming services, and other commercial applications, we need to compress the multi-view sequence efficiently without sacrificing its visual quality significantly [1-4]. The ME process usually takes about 60% of the computational load for the case of the single reference frame, and 80% for the case of the multiple reference frames. Many algorithms have been proposed for ME for video compression so far. All the methods are proposed keeping any one or more of the three directions aimed that 1. reducing computational

complexity 2. representing true motion (proving good quality) 3. reducing bit rate (high compression ratio). In videos, changes between the frames are movement. The encoder is estimate the motion that occurring between the reference frames and current frame, this process is called motion estimation.

The well-known full search (FS) [5] is the simplest, FS is the best one in terms of the quality of the predicted image and its resolution of the motion vector it is very computationally intensive. By searching every candidate position within the search range but it requires very expensive computational cost. For solving the complexity problem of motion estimation, many fast motion estimation algorithms were proposed. Due to the huge demand of the computational requirement several fast search algorithms have been developed and introduced in recent years including the Affine based method. These algorithms reduce their computation by using techniques such as sub-sampling search area and by restricting the search window. Due to their uniform search strategy without differentiating the types of motion, these algorithms perform well only for certain video sequences. These algorithms increased the searching speed significantly, but they still maintained the prediction quality comparable with the Full Search. These algorithms use uniformly allocated searching points in their step which becomes inefficient for estimation of small motions since it gets trapped into local minimum. The pattern block based fast search algorithm was proposed such as three step search (TSS)[6], new three step search (NTSS)[7], 4step Search (4ss)[8] but there is limit on the number of steps that the algorithm. Block Motion Compensation using fixed sized blocks has disadvantages that large blocks may fail to match the actual motion in a sequence, particularly along the moving edges, while small blocks require more overhead information. Most of the time, the objects in motion have different size and shape. Based on this characteristic of motion, in order to overcome the limitation of the traditional fixed-size block motion compensation (FSBMC)[9]. Adaptive Motion Estimation Method (ARPS) [10] algorithm makes use of the fact that the general motion in a frame is usually coherent, i.e. if the macro blocks around the current macro block moved in a particular direction then there is a high probability that the current macro block will also have a similar motion vector. This algorithm uses the motion vector of the macro block to its immediate left to predict its own motion vector.

Lili Hsieh et al. [11] have presented two efficient fast motion estimation algorithms with a two-stage predictive search based on joint spatial temporal correlations are proposed to reduce the search complexity. In the first stage, a rough search from the given motion vectors associated with six spatially and temporally correlated blocks attempted to find a starting point of the adequate search range that is closer to the global optimum. In the second stage, block-based gradient descent search and predictive partial search algorithms were used for fine search to elaborately search the adequate range from the starting point for the best motion vector. Simulation results have shown that the algorithms effectively reduce the average number of checked points to only 1.55% and 0.78% as compared to the full search method and yield a great performance improvement in terms of computational complexity, PSNR and bit rates as compared to full search and some well-known fast search methods.

Yasser Ismail et al. [12] recently proposed Fast Motion Estimation System Using Dynamic Models for H.264/AVC Video Coding which finds the initial search center prediction by using the median predictor the spatial correlation between the previous immediate coded blocks surrounding the current block which is the left, top, and top-right blocks. This is implemented by checking whether the SAD of the first searching point ( $SAD_{ini}$ ) is less than  $SAD_{AvgSta}$ . It simply works well, but search range is fixed in term of shape and it needs several threshold values for decision.

The authors try to efficiently fix the search range in flexible shapes, and find a way to accurately detect motion in pixel level. In this paper, we employ Scale Invariant Feature Transform (SIFT) to determine a search-starting point, and fix the search range in flexible shapes. The pixel-based motions are then detected by using Optical Flow, which is widely accepted as excellent tool.

The rest of this paper is organized as follows. Our proposed Method is presented in Section 2, and the experimental results are shown in Section 3. Finally, the conclusion is drawn in Section 4.

II. PROPOSED METHOD OF MOTION ESTIMATION

Motion, which estimated is basically movement of pixels. Pixel movement is necessary to find in an image. It is basically considered that searching moving pixels in entire image is not an efficient way. It is better to sense movement of pixels in previous frames, and efficiently works on only moving pixels. To know moving-pixel range, our basic concept is to firstly determine a search starting point in the first frame of a stream, determine a search range by using a previous frame, and calculate pixel-based motion in the determined search range as shown in the flowchart in Fig.1. Since search range should be determined in any shapes in order to efficiently cover all moving pixels, Scale Invariant Feature Transform (SIFT) algorithm is recommended in this paper as an appropriate tool for the process of search range determination. In pixel-based motion detection as the next process, Optical Flow (OF) algorithm which is excellent in moving detection is used in the search range. Although OF is

regarded as high-complexity algorithm, it works reasonably in this case due to limitation of search range. How to apply those tools is explained in the following sub-sections.

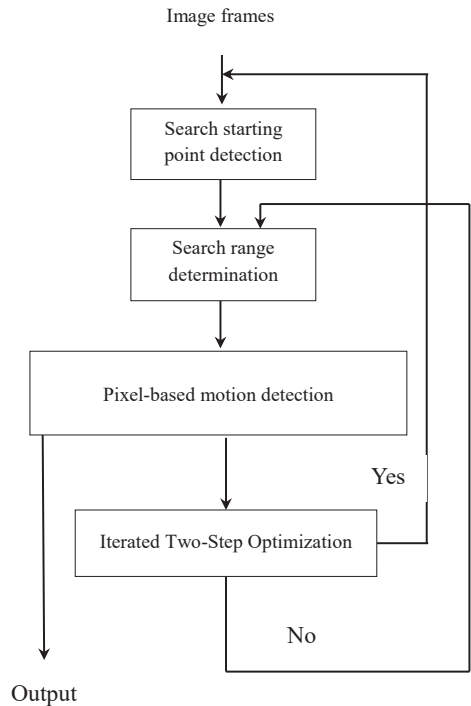


Fig.1. Flowchart of proposed method

A. Region Based Algorithm(SIFT)

The SIFT algorithm generates a stack of filtered images at different scales in order to determine features over the images that are invariant to rotation, scaling, and translation at different scales. SIFT implementation creates the family of derived signals by progressively convolving a series of 2-D Gaussian kernels with the original image in order to create the scale space. The scale parameter for this implementation is the  $\sigma$  value of the Gaussian function. The scale space is divided in octaves, which can be defined as a group of filtered images resulted from the convolution of the base image with 2-D Gaussian filters at different scales. One octave is created once the  $\sigma$  value is doubled. For this specific implementation, the initial value of  $\sigma$  is 0.6 and five images are produced for every octave. Finally, a pyramidal representation is determined by computing the difference of Gaussian functions (DoG) of the images. The candidate landmarks are calculated as the scale-space extrema in a  $3 \times 3 \times 3$  neighborhood for each pixel in every layer of pyramidal representation.



### B. Optical Flow Algorithm

As an alternative to scale-space approaches, optical flow (OF) algorithms provide motion information for every pixel of the images instead of information for a limited number of landmarks; this characteristic and the fact that the result is a smooth vector field makes OF a desirable approach. We aim to determine the feasibility of this type of algorithms to calculate accurate velocity gradients with the characteristics of this type of images.

OF differential methods are common techniques and can be classified in two types. Local methods define a local energy function to be minimized, which require the motion model to be defined for windows of specific sizes since large motions can only be detected with larger windows. Global methods minimize a global energy function; this allows evaluating motion information for the complete image. The flow field is smooth across the image which is desirable for our application since the motion of the heart is smooth among frames. HS is developed on the basis of a smoothing constraint of the velocity field, which means that neighboring pixels will present similar velocities in the flow field that has to be maintained for the complete image. A constraint on the brightness gradient is also introduced for the calculation of the OF velocities in the horizontal, vertical, and time directions.

The OF of a pixel cannot be calculated independently of the neighbor of such pixel, hence constraints have to be added. The aperture problem includes two of the main assumptions of OF methods: 1) the brightness or data conservation constraint, where it is assumed that the brightness of a region is maintained while the location may change; and 2) the coherence constraint, where it is assumed that the flow in a region should change smoothly; however, for a large region, it is more likely to find multiple motions that would violate those constraints.

The OF objective function for a classical model is composed of two terms from the work

$$E_{\text{Global}} = E_{\text{Data}} + \lambda E_{\text{Regularization}} \quad (1)$$

$E_{\text{Data}}$  measures the consistency of the calculated OF field with the movement of individual features, while the regularization term adds a coherence constraint, which restricts to the possible solutions to the optimization problem giving a preference for smooth solutions.

The implementation chosen utilizes the classical formulation of energy with the utilization of median filters for the creation of a pyramidal evaluation. The energy function is

$$E(u, v) = \sum \{ \rho_D (I_1(i, j) - I_1(i + u_{ij}, j + v_{ij})) + \lambda \cdot \rho_s [ (u_{ij} - u_{i+1,j}) + (u_{ij} - u_{i,j+1}) + (v_{ij} - v_{i+1,j}) + (v_{ij} - v_{i,j+1}) ] \} \quad (2)$$

where  $u$  and  $v$  correspond to the horizontal and vertical velocities that form the flow field,  $\lambda$  is a regularization parameter, and  $\rho$  are penalty functions calculated as  $L^2$  norms.

A constraint on the brightness gradient as well as the coherence constraint are introduced for the calculation of the

OF velocities in the horizontal, vertical, and time directions in the chosen implementation.

### C. Iterated Two-Step Optimization

This section describes the proposed iterated two-step optimization framework for optical flow and model parameter determination by alternating between local and global optimization. The local optimization (step 1), given small initial overlapping/sliding image window size  $|\Omega|$ , i.e.,  $5 \times 5$  and number of terms  $M$ , the nine model parameters are locally optimized. However, it can be assumed that the remaining two model parameters,  $|\Omega|$  and  $M$ , are needed to identify the best combinations, which will further enhance the estimation accuracy of model parameters, i.e., optical flow. That is, in step 1 optimization, these may impact the estimation accuracy as follows: Different numbers of  $M$  will locally cope with different image features such as orientation and frequency of DT. On the other hand, when  $|\Omega|$  is taken from a larger image region, optical flow will become smoother. For discontinuous local motion, a smaller  $|\Omega|$  must be set beforehand. Again, step 1 aims to independently achieve optimization in each small, i.e., local, window. Moreover, it is assumed that global optimization for the entire frame may enhance the consistency of the local optical flow solutions. Since it is difficult to apply a deterministic rule, statistical characteristic properties of moving objects in images are used. As, from step 1, different optical flow results will be obtained due to the range of given different combinations of  $|\Omega|$  and  $M$ , so different energy spectrum distributions can be obtained. Subsequently, when given a pair of  $|\Omega|$  and  $M$  in step 2, the match between the theoretical and estimated profiles is analyzed by the standard zero-mean normalized cross-correlation function.

## III. EXPERIMENTAL RESULTS

In order to evaluate performance of the proposed method, experiments done by 4 dataset with 100 samples each as shown in some samples in Fig.2. The specification of experiments is shown in Table 1. Evaluation approach and experimental results are explained in the following subsection.



Fig.2. Data set for Experiments

TABLE I. DATA SET

Data set	Resolution	Frame rate
Akiyo	352x288	15 fps
Akko&Kayo	640 x 480	30 fps
Breakdancers	1,024 x 480	30 fps
Foreman	176 x 144	30 fps

A. Evaluation

Evaluation of video coding algorithms can be done using objective and subjective measures and processing time all the process. The most widely used objective measure is the peak signal-to-noise-ratio (PSNR) of the signal which is given as

$$PSNR = 10 \log \frac{255^2}{MSE} \tag{3}$$

with being the mean squared error (MSE) between the original and decoded video samples. Typically, PSNR values are plotted over bit rate and allow then comparison of the compression efficiency of different algorithms. This can be done in the same way for MVC. However, PSNR values do not always capture video quality as perceived by humans. Some types of distortions that result in low PSNR values do not affect the human perception in the same way. In this subjective test, subjects are being shown the decoded video signal from a candidate codec. The subjects judge the quality

B. Experimental Results

The result shows candidate pixel from SIFT process as shown in Fig. 3. It calculates to characterize the surrounding neighboring of landmark and then the matching process is compared the descriptors of every landmark in both the first frame.

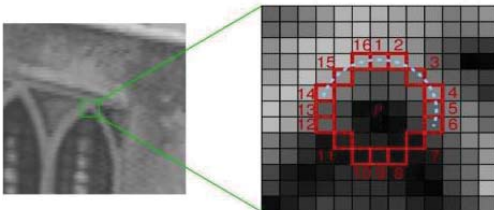


Fig.3. SIFT process result

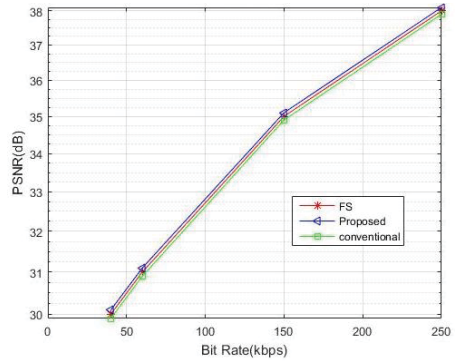


Fig.4. compares the rate-distortion performance of Akiyo

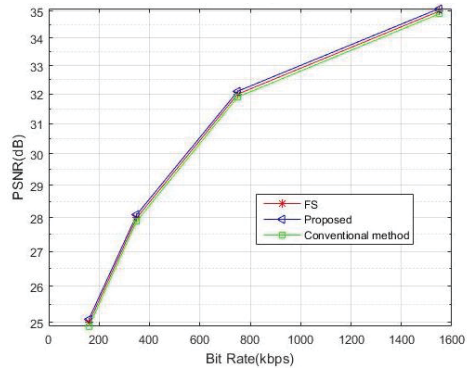


Fig.5. compares the rate-distortion performance of average every data set

TABLE II. COMPUTATIONAL TIME

Data set	FS	Conventional	Proposed
Akiyo	22,367	18,452	19,929
Akko&Kayo	27,324	19,444	21,766
Breakdancers	30,032	25,307	25,609
Foreman	21,070	15,370	15,560

The performance of the OF algorithm with region based was evaluated by PSNR and computational time. The PSNR were shown in Fig. 4 and 5. The proposed method show good performance by leading the conventional method and full search algorithm. the computational time of all process, the proposed method still improve than FS but nearly equal to conventional method as shown in table II.

IV. CONCLUSION

This paper proposed motion estimation to reduce processing cost and complexity when compare with only optical flow. It can solve the redundant problem by

using region selection. The region selection process use scale-invariant feature transform (SIFT) detected keypoints, to be used as landmarks for motion tracking and use optical flow algorithm tracked the object pixel after that apply iterated two-step optimization framework for optical flow and model parameter determination by alternating between local and global optimization for improving the new search range. Experimental results show that the proposed algorithm has good image equality measured in PSNR term. It can improve 0.2-0.3 dB. and can reduce time when compare with FS.

## REFERENCES

- [1] A. Smolic and P. Kauff, "Interactive 3-D Video Representation and Coding Technologies," *Proceedings of IEEE*, vol.93, No. 1, pp. 98-110.
- [2] T.Wiegand, J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264 / AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, 2003.
- [3] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions," *SPIE Conference on Applications of Digital Image Processing XXVII*, Colorado, USA., 2004.
- [4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable H.264/MPEG4-AVC Extension," *Proceedings of the IEEE International Conference on Image Processing*, Georgia, USA., 2006.
- [5] J. Lu, and M. L. Liou, "A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation", *IEEE Trans. Circuits And Systems For Video Technology*, vol 7, no. 2, pp. 429-433, April 1997
- [6] Renxiang Li, Bing Zeng, and Ming L. Liou, "A New Three-Step Search Algorithm for Block Motion Estimation", *IEEE Trans. Circuits And Systems For Video Technology*, vol 4., no. 4, pp. 438-442, August 1994.
- [7] Lai-Man Po, and Wing-Chung Ma, "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", *IEEE Transactions on Circuits and Systems for Video Technology*, vol 6, no. 3, pp. 313-317, June 1996.
- [8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, July 2003.
- [9] Avijit Kundu, "Modified Block Matching Algorithm for Fast Block Motion Estimation", *IEEE- International Conference on Signal and Image Processing*, pp. 260-264, 2010.
- [10] Yao Nie, and Kai-Kuang Ma, "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", *IEEE Trans. Image Processing*, vol 11, no. 12, pp. 1442-1448, December 2002.
- [11] Lili Hsieh, Wen Shiung Chenb and Chuan Hsi Liub, "Motion estimation using two-stage predictive search algorithms based on joint spatio-temporal correlation information", *Journal on Expert Systems with Applications*, Vol. 38, No. 9, pp. 11608–11623, 2011
- [12] Yasser Ismail, Jason B. McNeely, Mohsen Shaaban, Hanan Mahmoud, and Magdy A. Bayoumi, "Fast Motion Estimation System Using Dynamic Models for H.264/AVC Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, NO. 1, January 2012.

# A Novel Naturalistic Concept Proposal “ATMAN” towards Human Evolution for Personal Bliss, Harmony and Global Peace

Shobha Lala

Center for TeleInfrastruktur Global Capsule (CGC),  
Aarhus University,  
Aarhus, Denmark,  
lala\_shobha@yahoo.com

Prof. (Dr.) Ramjee Prasad

Center for TeleInfrastruktur Global Capsule (CGC),  
Aarhus University,  
Aarhus, Denmark,  
ramjee@btech.au.dk

**Abstract**— There has been constant evolution in the human species since the time of its inception on the Earth planet. The rapid changes in life are evident in the technological advancements. The altered way of life has a great impact on the behavior of individuals in the form of stress, anger, jealousy, impatience, frustration, etc. which leads to diseases like mental stress, sleepless nights, hypertension, cardiac arrests, and many more. This paper proposes a natural cure that is a combination of natural processes like Sun-gazing, Physical Exercise, Breathing, Relaxation, Laughter and Chanting to be practiced under a process Called “ATMAN”. It is said that “Healthy mind lives in a healthy body”, however with the practice of ATMAN we can say that “Healthy Mind leads to healthy body”. ATMAN is an endeavor on the path of human evolution for the utmost growth of Concentration, Confidence, Creativity, Calmness and Contentment (C5) and unveil the positive traits like compassion, consideration, patience, harmony and so on. The improved individuals can become path leaders and torch bearers towards personal bliss, harmony and Global peace.

**Keywords** □ □man Mind and □ody; ATMA□; □□; □lobal □armony; □eace

## I. BACKGROUND

Recent technological advancements have brought a full swing change in our life and in a much faster pace than we can imagine. This has transformed our life considerably in terms of comfort and accessibility, however, it is too short period for a body that took a long time of evolution by natural process. The natural resources which are present on the planet in abundance have a significant impact on our lives. The energies of the natural resources like sun, water, air and soil are the fundamentals to our living. The natural heat and light of the sun has been used to awaken the hidden qualities of a human being. The curative rays are known to have healing effects on human body and mind. The breathing pattern has a tremendous effect on the human body [1-11].

The human body can gain immense power and energy from the sun in one’s day to day life. It is because of this energy that we are more active during the day as compared to the evening or night. Our breath (Life force) is also dependent on the sun energy. The human life is all about breathing – no

breath means no life. The breathing pattern has a direct connection with the state of mind of an individual. The pattern changes according to the moods and thoughts; when one is angry, happy, stressed or at bliss. The reverse is also possible that when the breathing pattern is changed the state of mind will also change. A human being can also control and change the pattern of breathing.

This paper presents the impact of natural methods on human mind and body through a process called ATMAN. ATMAN is a process consisting of sun gazing, physical activity, relaxation, laughter, breathing exercise and chanting. The process will be practiced in daily life in context with Concentration, Confidence, Creativity, Calmness and Contentment (C5). ATMAN practiced with C5 will bring a significant change in the individuals on physical, mental and emotional levels. It will help people to unveil their hidden powers and talents and attain a state of contentment, happiness and bliss that will further help in the growth at society level and will go up to the Global level. ATMAN is a step forward in the direction of human evolution and will fulfill the innate desire of a human being for attaining higher levels of consciousness, calmness, contentment and eternal bliss. It will help the individual to have a control on the way of life and to win the hearts of other fellow beings.

In Meta Physics, the end part of the matter is the molecule- the Atom (neutrons and protons) which is a form of energy. The same energy is present in the human body in the form of life-force. This energy is acquired by the human body by regular breathing. The energy is present in abundance in the universe and - in the cosmos and this vast form of energy is BRAHMAN (The Whole) and a small part of this energy which is present in human body is ATMAN (The Self). Both the energies essentially have the same characteristics. ATMAN is a journey to unveil the hidden potentials present in the human body which have long been discarded due to the mechanical life styles. It involves a few changes in the life pattern of individual involving a method of sun gazing, physical exercise, relaxation and breathing exercise. It will help in inculcating the positive traits and help the individuals

become more aware, conscious and compassionate lot who pay heed to the well-being of the fellow human beings and care for this beautiful Earth Planet and make it a place where peace and harmony are prevalent all around.

The paper is organized as follows. Section 2 presents C5 concept. ATMAN process has been very well discussed in Section 3. Section 4 describes observation and practice of ATMAN and finally Section 5 concludes the paper.

II. C5 CONCEPT

The C5 concept has a deep impact on an individual’s personal growth leading to a stress free and healthy life. Reference [1] discusses methods to help people to lead a stress free and tension free life and have a restful sleep. The five Cs are, Concentration, Calmness, Confidence, Contentment and Creativity. C5 concept has been illustrated in Figure 1.

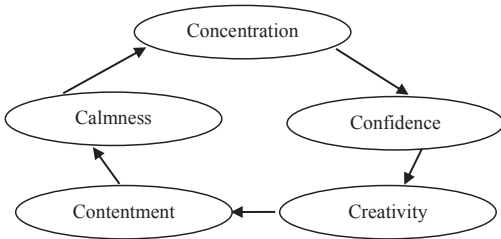


Figure 1. C5 Concept

A. Concentration

Concentration is a focused behavior to anything in life be it a problem, a situation, or a thought. A focused contemplation of a problem or a difficult situation can yield positive outcome and an improved way of life. An instant and thoughtless reaction to anything can have adverse effects on the outcomes. However, if we dedicate a few minutes on all the aspects of the problem or situation, we are sure to get positive results. Concentration does not come naturally to us, it has to be developed and achieved by regular practice. There are different ways of developing concentration. Meditation helps to a large extent to achieve concentration and improve the performance of an individual. .

Concentration is the primary requisite for the fruitful completion of any task. By dedicating a few minutes and contemplating the pros and cons of the task we are very well prepared to face any situation arising out of it.

B. Confidence

Concentration leads to confidence. When we concentrate on a thing, we become aware of everything involved with it, we gain knowledge. We are aware of all the aspects of the thing whether positive or negative and are well prepared for

the consequences. This gives a confidence to face any situation or outcome of the matter. Our confidence takes us a long way in the path of success. The confidence of an individual is reflected through the body language, way of communication and the way in which one handles a particular situation. Concentration brings knowledge and knowledge leads to clarity. Confidence is noticeable through one’s behavior in the society.

C. Creativity

After being armed with concentration and confidence, one can play with the task in hand. Since one is aware of all the aspects of the task in hand, one can fit pieces together and even experiment with the aspects to bring about a new form of the task. This is being creative. Creativity is not limited to artists, painters or poets. Everybody can be creative. A woman when she knows everything about cooking can experiment with the ingredients and her creativity would reflect in the food that she prepares and the way she presents. A gardener who knows everything about the plants in one’s garden can be creative in the way one grows the plants and the way one arranges them according to the season and environment and the creativity is evident in the way one maintains the garden. Creativity is not about looking for different opportunities or different fields of jobs, but it is about how one can excel in the tasks and resources available for utilization and how one can make the best out of it.

D. Contentment

A woman, when she is appreciated for the food she makes or the gardener when complimented for the beautiful garden, achieve a sort of inner happiness that is called contentment. When one is happy with oneself and knows that the efforts are yielding a positive response, one achieves contentment. The happy and contented state of mind leads one to other tasks, jobs and an encouragement to put in the best efforts and create master pieces. It gives direction to one’s life and helps one to decide the future course of action. Contentment generates acceptance of situation even if they are failures and leads to extra efforts to be put in for further success in every task.

E. Calmness

With a focused attitude, one gains clarity of vision and confidence and excellence in one’s tasks, and mental peace. One has a deep sense of calmness and control. The mind is tension free and stress free and the ability to handle difficult situations increases. With a calm mind one is at rest physically and mentally too. One can have a peaceful and sound sleep and be full of energy and enthusiasm, the next morning and be ready to face the challenges ahead.

A calm and peaceful mind has the best decision making power and an insight to be aware of the positive and negative aspects. A calm mind means no anger, lust or jealousy and a deep sense of satisfaction. It leads to a stress free life, a good perspective for the future and the maximum enjoyment of the present. The mantra is ‘Live for the Moment’. Calmness does not mean no activity, no sound or no hustle bustle of life but it

means a deep tranquility, a sense of inner peace even when one is surrounded by the crowd and noise and hustle bustle of life.

As depicted in Figure 1, C5 is a beautiful circle and once the C5 has been achieved, it helps in all the activities and different fields of life. After the achievement of a goal, a person can shift attention to the other fields and other activities as well, armed with C5 to move on the path of accomplishing success in whatever one is doing. It is rightly said that successful people do not do different things, they do things differently.

### III. THE PROCESS OF ATMAN

Since time immortal human kind has been growing, different aspects of growth have brought a radical change in the civilization. Physical changes, social changes and changes in technology are evident in modern life. Along with these changes: a man has been growing on intellectual levels as well as spiritual levels. There always has been a deep seated desire in human mind to experience higher levels of consciousness, calm, peace and bliss. The human mind seeks solace in different types of activities like music, dance, painting, meditation, adventure expeditions thrilling activities like sky diving and bungee jumping. The mind keeps looking for vents to give way to the emotions and have a feeling of contentment. ATMAN is a process that opens gates to our inner self and helps to discover our true identity through simple and naturalistic resources and techniques which are abundantly available, free of cost.

Figure 2 presents the ATMAN process, namely, Sun-Gazing, Physical exercise, Relaxation, Laughter, Breathing and Chanting. These six topics have been described in the following sub sections.

#### 1. Sun-Gazing

Natural resources like air, water and sun play a significant role in the development and growth of human body and mind. These natural resources were present even before the evolution of mankind. Without these natural resources: there would be no life on earth. The sun is the primary source of life on earth. No Sun means no Earth. The sun has a significant role on the life pattern of human beings. In [3], the author observed how people change their life style according to the presence of the sun. The effect of daylight on the day-to-day activity and its effects on the mood of individual have also been observed. People feel happy, positive and contented. The energy of the sun can have a tremendous effect on the human body. A major effect can be noticed during the day when we feel more active and energetic compared to the evening or night. The season with more sunlight like the spring, uplift the mood and generates feeling of love in human mind.

The energy of the sun can help in removing the negative traits like anger, jealousy, impatience and lust and is helpful in inculcating positive traits like consciousness, confidence,

concentration, calmness and contentment. The effect of the sun energy for individual growth is evident. The curative rays of the sun help developing Vitamin D in the human body and getting rid of the disease like rickets and are helpful in curing jaundice in new born babies.

#### 2. Physical Exercises

Life in our body is manifested in various regular physical activities like walking, working, doing the daily chores at the house or at work place. In recent times the physical activities have become mechanical; while some parts of the body are over stressed and others are hardly in movement. A balanced physical activity where most of the body parts are affected can help in getting physical fitness. Physical activities help in increasing power, strength and endurance in the body and are also useful in stress and anxiety management, and in having a better sleep. A healthy body houses a healthy mind.

With the practice of simple body postures, we can have a control over our body and limbs. A complete regime of physical activity essentially involves the types of physical activity by which all parts of the body are affected. The various parts of the body involved in physical activity should be the eyes, the neck, shoulders, abdomen, the back, thigh joints, knees, ankle joints and the small joints like toe joints and finger joints.

#### 3. Relaxation

Relaxation practice after physical activity helps to retain the effects of the physical activity in the body. There are, various method practiced by individuals for relaxation. Some people go for swimming, yoga, meditation or some just sit idle and relax. Focusing attention to different parts of the body helps in total relaxation of the body. Observing the breath and the path of the breath with every inhalation and exhalation is another method of relaxation. Relaxation helps to develop focused thinking, positivity and receptivity. ATMAN involves a method of total relaxation of the body and mind.

#### 4. Breathing Exercises

Human beings get energy by breathing which helps in smooth functioning of various important systems in the body like blood circulation, respiration, Etc. Breathing is the first natural activity that the living beings perform. The breath denotes the presence of life in a body. The pattern of breath keeps changing as the human beings grow with age. When we observe the breathing pattern of a baby, we see the movement of the abdomen - with every inhalation the naval rises and with each exhalation the naval falls. As one grows in age and gets engrossed in other activities of life the breathing pattern changes. When we observe the breathing pattern of an adult, we see that there is no movement of the abdomen during inhalation or exhalation. It shows that the breathing is shallow and the oxygen does not reach all parts of the body. This can be a cause of stress.

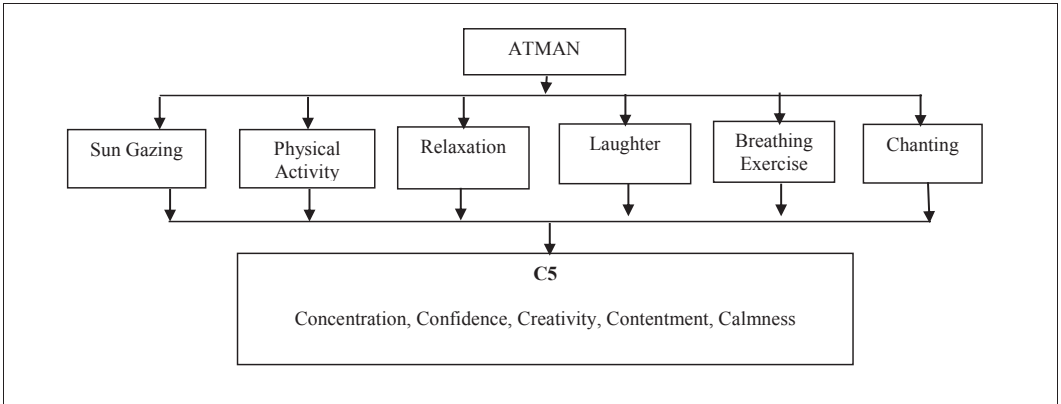


Figure 2. ATMAN Process

Breathing pattern has an association with the mental state of a person. It stimulates clear thinking and help in leading a stress free life and a good restful sleep. Breathing pattern changes with one's mood and the state of mind one is in. When one is angry, excited or exhausted, the breathing pattern is faster and when one is happy or at peace, the breathing pattern is different. By using simple breathing techniques, one can be free of fatigue, anxiety, stress and depression and can attain the state of happiness or bliss.

#### 5. *Laughter*

Laughter is the best medicine' is an old saying. A hearty and long laugh generates positivity in and around the body. It creates a happy and pleasant atmosphere and a beautiful interaction among individuals. A good laughing exercise can prepare individuals to deal with the stressful events of life in a proper manner. Laughter affects all elements of body like mind, soul and emotion and it can generate healing energy even a smile can win a million hearts. A good laughter will smoothen the creases of worry from one's face and uplift the spirits of individuals.

#### 6. *Chanting*

Natural sounds like the sound of flowing water, the humming of breeze, the chirping of the birds have a soothing effect on the body and mind. The process of Atman involves chanting of the Sound OM. The sound OM is the cosmic sound and is the mother of all sounds. Chanting of OM in a systematic controlled way generates clarity of mind, vision and enhances confidence and decision making power. It also brings positive thoughts, helps to release stress and makes the mind calmer. The C5 concept lays a great emphasis on the Om chanting.

### IV. OBSERVATION AND PRACTICE

The main observation of this investigation is to have a Global environment that is full of peace, free of stress and sleepless nights for everyone who practices the ATMAN process together with C5. If the natural techniques are followed by individuals, it will help them in living a healthy, happy and peaceful life by using just 60 minutes of their busy schedules.

Another observation is that the breathing pattern has a connection with the state of mind and moods of individuals. The breathing pattern changes when one is angry, excited, happy or at bliss. This research is to observe the changes in the state of mind of individuals in relation to the breathing pattern. For example when one is at bliss the breathing is deep and slow, so we can achieve a state of bliss by controlling the breaths and making them deep and slow. By practicing sun gazing technique an individual will develop concentration in the body and mind. A focused attitude in life towards family, friends and work can be the key to success and a good life. Concentration will generate awareness and knowledge. An awakened person will concentrate on the problems and difficult situations around oneself and will put efforts to find solutions to them. These people will work for the enhancement of the society Globally.

A regular regime of physical activity results in having total control over the body and its parts, in terms of movements and flexibility. Control over the body gives confidence and it is reflected in the body language. It is the confidence that makes one stand apart in a crowd of thousands of people. These confident individuals are the leaders who can lead the masses towards positive growth of the society and be the path leaders and torch bearers towards global peace and harmony.

Concentration, confidence and knowledge give birth to creativity. Innovative ways and means can be devised by confident people for enhancement of the society. Positive and effective solutions can be created for the problems being faced by the world at large. The inner positive traits of the human beings need to be awakened. A calm and positive mind generates an atmosphere of positive energy that observes the positive traits of others. Calm and confident people can identify the positive qualities of the individuals and can work for the utmost growth of hidden talents and qualities which have remained silent all through the years. An inner sense of satisfaction can prevail when one knows that the efforts are bearing a fruitful result. Satisfied with the positive results in one field, the person will focus in other fields and with new found qualities one can excel in every work on hand. This gives a multi-dimensional growth to humanity as a whole.

ATMAN is a journey on the vehicle of natural techniques for self-realization. ATMAN is a complete package for the positive growth of the body, mind and the soul. A person can gain control over the body with physical activities and the body obeys the commands of the individual. One is in charge of the movements in the body and can mold the body as per one's own wishes. A healthy body goes a long way in fulfilling the goals one has set. With the breathing techniques one gains control over the breaths. The breathing pattern changes according to one's wishes. One develops capacity to hold breath. One can control the number of breaths that one takes in one minute. By deep breathing techniques the number of breaths can be controlled. A normal human being takes 15-16 breaths in a minute. With the deep breathing technique the number can go down to 5-6 and with more practice it can reduce to 2-3 breaths in a minute.

The relaxation technique teaches the individual to control one's attention. In the total relaxation pose one diverts attention to different parts of the body and learns to focus attention and gain concentration. The relaxation technique helps to have a stress and tension free life. Laughing and chanting also help in leading a stress free life and an inner sense of peace, tranquility and calmness prevails.

ATMAN will help in the utmost growth of the C5; additionally, Control and Consciousness will also develop in the human being. Thus, a circle of positive traits and growth is formed by following the process of ATMAN together with C5.

ATMAN is a process of 60 minutes described as under:

- |                       |            |
|-----------------------|------------|
| 1. Sun Gazing         | 5 Minutes  |
| 2. Physical Exercises | 25 Minutes |
| 3. Relaxation         | 10 Minutes |
| 4. Laughter           | 2 Minutes  |
| 5. Breathing          | 13Minutes  |
| 6. Chanting           | 5 Minutes  |

V. CONCLUSION

ATMAN is a way of life towards creating humanity, positive aura around the individuals. The practice will grow the individuals into improved human beings. It is an endeavor to save humanity and to make the Earth Planet a beautiful and peaceful place to live for the future generations. ATMAN offers a complete package in the path of human evolution to lead a stress free, tension free, calm, contented and peaceful life having a sound sleep at night. ATMAN together with C5 helps individuals to be confident, conscious, considerate and contented. They have clarity of mind and a positive vision. These individuals will be the leaders and torch bearers on the path of human evolution. ATMAN is a cleanser for the human mind body and soul. ATMAN is a journey on the road of human evolution riding on the natural techniques towards positive self-realization leading to personal growth, bliss, harmony and global peace.

REFERENCES

- [1] Ramjee Prasad, 'Unlock Your Personalization', Aalborg University Press, 2012.
- [2] Apar Saoji., Sriloy Mohanty., & Suhas A.,Vinchurkar,' Effect of a Single Session of a Yogic MeditationTechnique on Cognitive Performance in Medical Students: A Randomized Crossover Trial', Springer Science Business Media,New York 2016.
- [3] Bettina Hauge, 'Lives under the Sun', The Senses and Society Journal ,Vol. 10, Iss. 1, 2015.
- [4] Maria Fernandez-del-Valle, Tyrel S. McCravens,' Resistance Training Guidelines for Active Females Throughout the Lifespan: Children, Adolescence, Adult Women, and the Aging Woman', Springer Science Business Media, New York 2016.
- [5] Jonathan C. Smith, Roosevelt University Alberto Amutio University of DeLlsto, Spain John P. Anderson and Leslie A. Aria Roosevelt University.' Relaxation: Mapping an Uncharted World Biofeedback and Self Regulation', Regulation of Mental States and Biofeedback, vol 21 no. 1, 1996.
- [6] Busch V., Mageri, W., Kern, U., Haas, J., Hajak, G., & Eichhammer, P., 'The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing –An experimental study', Pain Medicine, 13(2), 215-228, 2012.
- [7] Rickard, K. B., Dunn, D. J., & Brouch, V. M.,' Breathing techniques associated with improved health outcomes', Virginia Henderson Global Nursing eRepository (2015, June 29).
- [8] Tiwari, S., Tiwari, S.K, Gehlot, S., & Singh, G.,' Outcome of breathing exercise (Pranayam) on spirometric parameters in type 2 diabetic individuals: a clinical study'. Journal of Stress Physiology & Biochemistry, 8(4), 218-225. ISSN 1997-0838 .
- [9] Dhruva, A., Miaszkowski, C., Abrams, D., Acree, M., Cooper, B., Goodman, S., & Hecht, F. M.. 'Yoga breathing for cancer chemotherapy-associated symptoms and quality of life: Results of a pilot randomized controlled trial'. Journal of Alternative & Complementary Medicine, 18(5), 473-479. 2012.
- [10] Nicholas A. & Rod A., 'Laughter and Stress in Daily Life: Relation to Positive and Negative Affect Motivation and Emotion', Journal , Vol. 22, No. 2, 1998.
- [11] Bernie,W., Candace, H., & Sydney, H. ' A Healthy Dose of Laughter and the Arts: The Role(s) of Play,Laughter and the Arts in Promoting Good Health and Maintaining Quality of Life', Springer International Publishing Switzerland 2016.



## **Towards quality of life – A case study from the Netherlands and Denmark**

*Ramjee Prasad,*

*Aarhus University, Demark*

*ramjee@btech.au.dk*

### **Abstract**

The current case study examines the 5 practices (Invocation, Clearing the mind and Concentration, Breathing exercises, “Om” reciting and daily planning) to relieve the stress and insomnia of a 65 years old male (case 1) from the Netherlands who was suffering from stress and insomnia and a 70 years old Danish female (case 2) who had a sleep problem for more than 2 years. Both persons have undergone 2-4 sessions of practicing these five practices without any adverse effects. After receiving 4 sessions, case 1 experienced a decrease in stress symptoms and sleep problems and case 2 also got complete relief after 2 sessions. Furthermore, they also experience an improvement in their quality of life. Quality of life assessment using WHOQOL-BREF - 100 showed improvement in both cases. These practices are grounded in C5 concepts: concentration, calmness, confidence, contentment, and creativity and they explain the importance of these 5 principles for Quality of life. The results of this study suggest that these practices are effective, safe and can

be expand for others who require treatment for stress and insomnia. Implications for treatment and research are discussed.

Key words – stress, insomnia, quality of life, concentration, calmness, confidence, contentment, creativity

**Introduction:**

Now a day, the term “stress” is very common amongst all age groups. The adverse reaction on people due to excessive pressure or due to other types of demands placed on them is stress. This definition focuses on external stressors or hazards of work place. [1] Moreover, stress occurs when perceived pressure exceeds your perceived ability to cope. [2] The stress experience exists in all human beings at all ages in varying degrees [3]. Stress occurs in relation to stressors that are positive and negative. Stress also contributes to positive and negative effects. As a result many people face a problem of insomnia. Stress is common in insomnia [4]. It affects nearly 10% of individuals [5]. Depression and insomnia are correlated. An epidemiological study has shown that persons with insomnia are at a higher risk of developing depression and anxiety disorders. [6] Study also shows that depression also predicts future insomnia. [7] Insomnia is mainly described by persons experiencing difficulty initiating or maintaining sleep, early awakening, and interrupted or non-restorative sleep. Chronic insomnia can also lead towards

ischemic heart diseases. Insomniac persons are also three times more likely to develop frequent debilitating headaches. [8] A research conducted at University of Chicago in 1999 and published in the Lancet found that after only 4 hours of sleep for 6 consecutive nights healthy young men had a blood result that they had a huge drop in their insulin level and they had an elevated level of stress hormone.

Non pharmacological treatment for insomnia is stimulus control therapy, sleep restriction, sleep hygiene, paradoxical intention and progressive muscle relaxation. Pharmacological treatment for insomnia include medicines such as antihistamine preparations, benzodiazepine receptor agonists, sedating anti-depressants, neuroleptics and melatonin and herbal medicine such as valerian. Evidence suggests that non pharmacological treatment alone or possibly in combination with drugs produce clinically significant and durable improvement. [9] These drugs also have some side effects like day time sedation, motor incoordination, cognitive impairments and related concerns about increases risk of motor vehicle accidents and injury from falls. [9] Chronic hypnotic exposures can also carry additional risk of physical or behavioral dependence withdrawal rebound insomnia, and increased mortality. [10,11,12,13,14]

Typically, people suffering from prolonged stress consult their doctor for help. Some people try to relax

by watching TV, eating in excess start smoking or take to drink – as a result they could lose control of the situation. In my opinion, these methods of relaxation only relieve the stress for the time being. After some time, the tiredness and stress will recur and grow worse because the strength and potential of the person's mind, body and intellect have not been stimulated. A more powerful solution for the problem is needed. It is possible to learn to control response to these situations and develop techniques that will reduce the effects of stress on mental and physical health. It is my view that the 5 techniques described in this paper can give rest to the mind and the body and help people to explore the full potential of the mind. These techniques are grounded in C5 concepts (concentration, confidence, calmness, contentment, creativity). These 5 practices have been integrated in my life for more than 50 years. I taught the 5 techniques to many advocates and judges with positive effect. However, the following two success stories have encouraged me to make the benefits of these techniques known to a wider public.

**Case Introduction:**

The first case is 65 years old. He was working as the scientific director of a university institute with a staff of more than 100 people, consisting of full, associate and assistant professors, PhDs and Post-docs, and technical, managerial

and administrative staff. He retired in 2011. He is having a problem of insomnia and stress.

The second case is 70 Years old lady from Denmark. She retired as a teacher so that she can be at home with her husband. But her husband died after a year due to cancer.

**Presenting Problems:**

First case: He cannot sleep well; he has problems in concentrating during the day and he often feels tired. He has a problem of insomnia and stress. Work life imbalance is the main cause for case 1.

Second case: She has a problem of insomnia. She woke up several times and thought she had heard strange sounds in her house. These sounds were enough to wake her up and keep her awake for hours. She honestly said, she had never been a sound sleeper, but after her husband's death it had become worse.

**History:**

First case: He has an extremely burdensome workload with work days of 12 to 16 hours. He always works seven days a week. He has no time to relax. In short, he has an interesting life. However, there is also a dark side to this way of life may be due to this busywork life or a demanding job he has problems of exhaustive

tiredness, insomnia and stress. He contacted a medical doctor; the response was: “you have burned out”. He commented “I have a nice job which I really like and it is not too hard. It just takes a lot of time and the problems have adverse effects on efficient function”.

Second case: She stopped teaching in 2001 to be at home with her husband who died a year later in 2002. She did not have any children. It takes a lot of time to her getting used to be alone. Her complaint was that she did not feel very safe in her house after her husband’s death. Especially at night she woke up and had problem in sleeping but after her husband’s death it had become worse.

**Course of treatment and assessment of progress:**

**For first case:**

In the first session series all the necessary preparations were made for case 1’s introductory C5 weekend session in Denmark. Case 1 had all kinds of thoughts about what could happen that weekend. These ranged from difficult physical and/or yoga exercises to mental training in concentration and meditation. He was uncertain and nervous: uncertain because he was still skeptical about whether these practices would help him overcome his tiredness and his sleeping problems, and nervous because he sometimes felt that he was escaping into the unknown just

because he hoped that his problems could be reduced. He could not rely on himself but he had to trust and he found it hard to do this without any proof that the experiment would have a positive effect. We started the first session early morning and walked slowly until we found an excellent location for our body and mind and start with the five Practices. It was a very successful first experiment and he was satisfied with the results. Then after this first session he started with a kind of relaxing exercise, walking outside, and early morning exercises. He spent 20 minutes every morning and 30 minutes doing some meditation and concentration practices in the evening and closed the day by summarizing the plans. We started doing regular telephone sessions. After a week he reported that clearing the mind and getting fully concentrated was far from easy. He should be less impatient. He should take walks in the country side. After some preparation, he starts with concentration practice again. After some telephonic consultations, he spent a considerable amount of time on the practices but after a week he felt that there was not much change. But he carried on. In next few days, he proceeded with the exercises in the early morning and during working hours. The week later there had been some progress. According to case 1, after four months the clearing practice went automatically, while the full concentration practice were still too difficult. After five to six months case 1 thought that the full concentration practices were

also going a bit smoother and they gave him more satisfaction. He felt good about doing this series of tasks every day. However, his sleeping problems had still not been resolved. There were some nights where he slept a bit better, but he still not had a full night's sleep. But he was satisfied with the fact that he felt better and this was in itself an achievement.

In the second session series, was organized in a castle resort in the Netherlands, after a year. We started in morning with walking and stopped on a selected point. Repeated the same practice and also included a new task planning” You should always start the day in a good mood, you should fulfill the duties of the day in a good mood and you should close the evening in a good mood”. The positive elements of the daily task gave him greatest satisfaction. However, his sleeping problem had still not been solved. After this weekend session, the morning session was extended by 10 minutes by adding topics to be handled that day. The concreteness of daily tasks in a morning and evening session helped in his daily life in a better way and to create a positive attitude towards his colleagues and family. However, his mind is till now not able to rest.



In 2006 he got a heart attack. He had experienced a 'near death' moment that made a huge impact on his life. This extreme experience laid the foundation for giving priority to working toward giving his mind more rest. During his rehabilitation program, he was advised not to do any physical activity.

In Third session series in 2006, the first day started a little emotionally, he was not happy and that unhappy feeling cannot be solved by fanatical exercising in daily sessions. There is only one solution, to ensure that the body and mind are in equilibrium.

After 2½ years, Case 1 joined the hospital's heart rehabilitation program where cardiologist, physiotherapists and sport doctors were available. Case 1 learned via body sensing just how much physical training he could stand. This gave him a lot of self-confidence. He found that physical exercise such as biking, rowing, and team sports like handball, volleyball, under the supervision of specialists, helped him to get rid of his daily worries and gave him renewed trust in being able to stay active, physically as well as mentally. With this physical exercise program, he maintained his morning and evening concentration and meditation session. He found that combination of physical and mental programs brought his mind more

rest. He has now managed to achieve a state of total balance in his body and mind and he is sleeping better, he is more active nowadays throughout the day and has resumed his duties, working 40 to 60 hours per week without any problem. Now he is active in various international education and research programs. Every day he is grateful that he feels now healthy and happy.

<b>No of sessions</b>	<b>Exercise</b>	<b>Symptoms</b>	<b>Improvement</b>	<b>Complication/Problem/Feeling</b>
1. First session series in 2004	Start with first practice followed till fifth practice daily in morning as well as in evening for 20 to 30 minutes	Always a feeling of tiredness, improper sleep, not able to concentrate properly	Start feeling good about doing this series of tasks every day,	Uncertain about what to expect. There was nothing concrete at the start. There was only the actual proof of the concept knowing the quality of life of Ramjee (proff.) and his demanding duties.

<p>2. Second session series in 2005</p>	<p>Same series of these five practices.  To start day with a good mood or planning</p>	<p>Improper sleep</p>	<p>positive attitude, sleeping problem  had still not been resolved</p>	<p>Telephone sessions are not the same as annual session series but they help in case personal feelings on progress in C5 are negative.</p>
<p>3. Third Session series in 2006</p>	<p>To combine physical and mental exercise</p>	<p>Not happy, little emotional</p>	<p>Combination of physical and mental exercises gives mind more rest, total balance in his mind and body, more activeness, better sleep</p>	<p>Due to heart attack a drastic change in life occurred. Experience with physical exercises combined with mental exercises gave much satisfaction. When the body feels good it is easier to work on the mind as well. The combination had big impact on finding</p>

				solutions for the problems
4.Fourth till ninth session series from 2007-2012	5 practices plus every day planning plus evaluation. Physical and mental exercises extended to 60 minutes per day	No symptoms	Better control over the C5 practices. Have no points for improvement since my quality of life has reach an optimum	Feeling happy in family life and work, have largely improved efficiency in work, feel much stronger now. On days spent less time on C5 feel less happy. Fortunately, the number of days without spending time on C5 is limited

.

**Second case:**

In August 2008, training was started for her. First thing was to find for her a short term objective and give her eight days to think over it. Case 2 found out that a possible stress factor could be that she never got to the bottom of clearing out her

mind. Case 2 remembered that she always asked her pupils in school to clear up the classroom and their private drawers before they went home, because external mess very often was a sign of inner mess! And here she was not able to live up to her own principles. As, she had not cleaned up her cupboards and drawers since long time. It was a very good idea and she started clearing out a little bit every day, some days only for half an hour. At the same time as Case 2 started clearing out her house she began practicing the 5 practices. Case 2 does the practice in the morning before she starts her other daily activities. The best thing is to practice outside, but as it is often rather cold and windy, she practices in front of an open door.

<b>No of sessions</b>	<b>Exercise</b>	<b>Symptoms</b>	<b>Improvement</b>
1. First session series in August 2008	Start with first practice followed till fifth practice daily in morning as well as in evening for 25 to 30 minutes	Always a feeling of tiredness, improper sleep,	Start feeling good about doing this series of tasks every day, clear up her house, help to make her feel self-confidence, less selfish and egoistic, help for understanding her fellow beings

<p>2. After 8 months of continuous practice</p>	<p>Same series of these five practices.  To start the day with a good mood or planning</p>	<p>No such symptoms</p>	<p>Very often she fall asleep (better sleep), more happier, more observant, better listener, calmer</p>
---	--	-------------------------	---

She performed these five exercises daily. Before she goes to sleep at night she goes through all the things she has experienced during the day, trying to think what effect these events have had on her mind, what positive and negative feelings they gave her and why. Very often she falls asleep before she has gone through all her day. After practicing it 8 months regularly Person 2 feels now there have been new developments, especially concerning meditation. She finds it easier to reach a state of mind where she can let her thoughts wander without control and where she is totally relaxed. When the alarm clock sounds after 15 minutes, she really feels that she has been away in another world. After ‘coming back’ there is a period of time where she finds that she is able to concentrate much more on what she is doing, and she sees things more clearly than before. It is a state of mind that can last for a couple of hours, and then it vanishes again. Of course there are still days where she

feels less concentrated, but they are rarer now. She is looking forward to starting her second phase of exercises. She thinks that it is important not to rush into it and that the more time spent on each phase, the better the impact will be. You have to have a total grasp of one phase before you move on to the next.

### **Assessment**

There was reduction in the symptoms in both cases. Quality of life assessment was done using world health organization WHOQOL-BREF – 100 questionnaires. The results for case 1 is shown in figure 1 and for case 2 is shown in figure 2. It has been observed that there was improvement in physical psychological and environmental status for both cases.

### **Five Practices (figure3):**

In this section I will discuss about the five practices recommended to both of these cases.

#### **First Practice: Invocation**

Invocation involves the use of some verses and hand actions as shown in figure 4. Initially, by opening the palms of hands, person has to repeat the sentence “Wish to have the full strength to run successfully all the operations for which I am responsible for achieving my mission.” Uttering this sentence will give a sense of

wellbeing to a person, by joining the palms together and repeat “Generator of the universe”. By folding the palms together and say: “Operator of the universe”, by clenching the palms and say: “Destructor of all evils”. It is recommended to repeat the process for 5 times all together or by repeating each sentence for 5 times.

**Second Practice: Clearing the mind and concentration**

Clearing the mind is a preparatory phase of concentration. You just have to think of nothing to clear the mind from preoccupied mental task; it is beneficial to allow the mind and body to rest during working hours. By listening to restful soothing music, by reading something relaxing, by nurturing a hobby, by playing a musical instrument or by walking in free nature also gives rest.

After clearing the mind, person is being ready for concentration. Following phases are explained to improve the time span and quality of the concentration.

C5 phase	Concentration without any break	Time for achieving the concentration without any break	Types of practice	Difficulty



Phase 1	5minutes	Up to 2 years	Memorize past happy events in your life. Focus on one point	Very difficult to focus
Phase 2	10 to 15 minutes	Up to 1 years	Visualize past happy events or future plans	Practice makes it easier
Phase 3	More than 30 minutes	Up to 9 months	Focus and think of innovative ideas	Full concentration for a longer period becomes easy

After 8 months of continuous practice Case2 is at the end of phase 1 and Case 1 has past phase2.

**Third Practice:**

This practice consists of deep abdominal breathing (Diaphragmatic breathing) in the ratio of 1:2. It involves slow and deep inhalation through the nose, usually to a count of 5, holding the breath and followed by slow and complete exhalation for a double the count e.g. 10. The process may be repeated 5 to 10 minutes, several times a day.

**Fourth Practice:**

Reciting “OM”- “OM” sound should be recited with long phonation for eleven times. Later on uttering of “OM” should be done by closing the lips together for 11 times.

**Fifth practice:**

It consists of daily planning of your activity in the morning. I advised to both Case 1 and Case 2 “Set the new plans for the day and review the plans that you have set. Prioritize important tasks of the day. Implement the topics designated for that day as best as you can. At the end of the day, look back with satisfaction on the things that you have accomplished. Do not place the topics which you did not complete in a negative corner, but put them on the list for tomorrow and try to do the remaining topics the next day. You should strive to limit the number of points on your list so that it is realistic and you can manage to deal with all the tasks on the day’s list. Take pauses to rest and exercise while you are at work”. Moreover, Following activities are important during the day

- Physical exercise program with counting the repetitive movements.
- Always be positive and think about positive events from our own life.

- Before going to bed, review all the experience during the day and recall the positive and negative event of the day. Try to think the effect these events have had on your mind
- Concentrate fully on an important topic for five minutes (figure 5). And think about the topic in its full depth, then work it out, evaluate it in detail and try to transform thoughts into actions to be performed in the days to come. The selected subject should contribute positively to daily life activities and lead to positive and pleasant actions

Daily regular practice of all above (Practice 1-5) exercise is recommended for at least 30 minutes every morning or evening.

### **Discussion:**

Finding from these studies shows that these above five basic exercises are beneficial for people experiencing problems with stress and insomnia and the effects from the practices need vigorous scientific research to understand its impact on mind and body. Stress and insomnia are two main factors which distracts the mind. Stress reduces health and wellbeing. [15] It is well known that stress affects the immune system [16], and influences the onset and progression of disease

conditions. [17] Stress is associated with the development of physical and physiological balance. Insomnia is the difficulty in falling sleep and waking up frequently during the night.

These exercises have been seen beneficial in my whole life and also for Case 2 and Case 1. Both cases have reduced their symptoms of stress and insomnia. Case 1 has improvised his work and life balance a lot and also largely improved in efficiency in work, feel much stronger now. Case 2 also feels happier, more observant, better listener, and calmer now.

Concentration meditation is intended to help individuals retrain their minds from habitually engaging in self-related preoccupations (such as thinking about the past or future, or reacting to stressful stimuli) to more present moment awareness.[18] In concentration meditation condition, there was relatively less activation in meditation in the PCC and left angular gyrus. [19]. Breathing exercises is known to have a relaxing effect. [20] By performing the breathing exercises 'tension anxiety' and fatigue of the group suffering from cancer were relieved to a level. [21]

### **C5 Essentials:-**

For Quality of Life (QoL), the integral part is the C5 (Creativity, Contentment, Confidence, Calmness, Concentration) concept. By performing above 5 practices regularly, both Case 2 and Case 1 achieved five "C" in their life, consequently

leading to their wellbeing (figure 6): Sustaining these qualities in everybody's life is important for QoL. Many people are familiar with these words, still innovative focus is challenging. Many researches are carried out to demonstrate the importance of these cognitive functions of mind. Some are relevant to this case study and we would like to discuss it in details.

**Creativity** is the most basic element of C5 concept. It refers to ability of a person. Every person is creative in his own way but it varies from person to person. Creativity starts in childhood so parents play a very important role in making their children creative. Creativity has itself been associated with wide range of cognitive style [22] and has been correlated with age, work experience and travel experience [23]. It is more closely associated with a holistic mode of thinking than with the one that is analytical [24] and is positively correlated with psychosocial growth. [25] Creativity is intellectual strength and it helps in increasing positive emotion. [26]

**Calmness** is a special feature. Calmness is the key towards peace. Calmness increases strength and helps to achieve aim. Calmness brings peace to everyone and it is a sign of peaceful world.

**Contentment** is positive emotion.[26]) It follows from creativity and from Fredrickson's (1998) initial broaden-and-build model, which comprises three

positive emotions, namely joy, interest, and contentment.[27] Self-oriented dispositional emotions (i.e. joy, contentment, pride, and love) are mainly related to curiosity, bravery, zest, love, gratitude, hope and humor. [26] Contentment and related emotions (e.g., serenity, tranquility, and relief) arise in situations appraised as safe and as having a high degree of certainty and a low degree of effort [28]. This emotion is distinct from mere satisfaction, or the pleasure that derives from a good meal or otherwise meeting bodily needs.

**Confidence** is an essential point for C5. Confidence means the belief in your ability to succeed. Confidence is the will power of a person to keep on going towards the goal. Confidence is the heart of all creativity. For a confident person, the sky is the limit. Confidence helps in increasing positive emotions. [26] )

**Concentration** is the process of bringing mind to zero speed. It means focusing on one and one thing only. It is gateway towards success. Concentration is a big achievement in itself. Concentration involves focusing attention on an object, which progressively leads to the restraining of the natural wanderings of the mind. These five are known as C5. C5 especially confidence, creativity and contentment will help in increasing positive emotions. [26] And Positive emotions will lead towards a stress free life. [26] that is towards quality of life.

## References:

1. Health and safety executive. Tackling work related stress a manager's guide to improving and maintaining employee health and wellbeing, Suffolk HSE 2001.
2. Palmer S, Cooper C, Thomas K, Creating a balance managing stress, London British library, 2003.
3. Holmes.T.H.,&Rahe, R.H. (1967) The social readjustment rating Scale. Journal of psychometric research,11, 213-218
4. Insomnia Associated with Depressive Disorder: Primary, Secondary, or Mixed? Ravi Gupta, Vivekananda Lahan
5. Pigeon WR. Diagnosis, prevalence, pathways, consequences and treatment of insomnia. Indian J Med Res 2010; 131:321-32.
6. Taylor DJ, Lichstein KL, Durrence HH, Riedel BW, Bush AJ. Epidemiology of insomnia, depression,
7. Buysse DJ, Angst J, Gamma A, Ajdacic V, Eich D, Rössler W Prevalence and anxiety. Sleep 2005; 28:1457-64, course, and co-morbidity of insomnia and depression in young adults. Sleep 2008; 31:473-80.

8. How to be healthy Insomnia Thomas, Pat The Ecologist; Jun 2007; 37, 5; ProQuest Research Library pg. 58.
9. Insomnia Sateia, Michael J;Nowell, Peter D The Lancet; Nov 27-Dec 3, 2004; 364, 9449; ProQuest Research Library pg. 1959
10. Roehrs T, Roth T. Hypnotics: efficacy and adverse effects. In: KrygerMH.,Roth T, Dement WC eds. Principle and practice of sleep medicine. Philadelphia: W B Saunders,2000:414-18
11. Kripe DF, Garfinkel L, Wingard DL, et al. Mortality associated with sleep duration and insomnia: report of the Institute of medicine Committee on the efficacy and safety of Halcion. Arch Gen Psychiatry 2002; 59:131-13.
12. Ray WA, Thapa PB, Gideon P. Benzodiazepines and the risk of falls in nursing home residents. J Am GeriatrSoc 2000;48:682-85
13. Roehrs T, Merlotti L, Zorick F, Roth T. Rebound insomnia in normal and patients with insomnia in normal and patients with insomnia after abrupt and tapered discontinuation. Psychopharmacology 1992, 108:67-71
14. Roth T, Roehrs TA. A review of the safety profiles of benzodiazepine hypnotics. J Clin Psychiatry1991; 52 (suppl) : 38-41.
15. P.A. Thoits, Stress and Health: Major Findings and Policy Implications, Journal of Health and Social Behavior 51(S) (2010), S41-S53.



16. J.E. Graham, L.M. Christian and J.K. Kiecolt-Glaser, Stress, Age, and Immune Function: Toward a Lifespan Approach *Journal of Behavioral Medicine* 29(4) (2006), 389-400.
17. V.M. Figueredo, The Time Has Come for Physicians to take Notice: The Impact of Psychosocial Stressors on the Heart, *The American Journal of Medicine* 122 (2009), 704-712.
18. Gunaratana H (2002) *Mindfulness in Plain English* (Wisdom Publications, Somerville, MA).
19. Meditation experience is associated with differences in default mode network activity and connectivity Judson A. Brewer<sup>a,1</sup>, Patrick D. Worhunsky<sup>a</sup>, Jeremy R. Gray<sup>b</sup>, Yi-Yuan Tang<sup>c</sup>, Jochen Weber<sup>d</sup>, and Hedy Kober<sup>a</sup> <sup>a</sup>-Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06511; <sup>b</sup>-Department of Psychology, Yale University, New Haven, CT 06510; <sup>c</sup>-Department of Psychology, University of Oregon, Eugene, OR 97403; and <sup>d</sup>-Department of Psychology, Columbia University, New York, NY 10027 Edited by Marcus E. Raichle, Washington University in St. Louis, St. Louis, MO, and approved October 4, 2011 (received for review July 22, 2011)

20. Gobel BH. Symptoms of alterations in coping. In: Yarbrow CH, Frogge MH, Goodman M, editors. Cancer symptom management. 3rd ed. Massachusetts: Jones & Bartlett Pub; 2003. p. 658.
21. The effects of deep breathing on 'tension anxiety' and fatigue in cancer patients undergoing adjuvant chemotherapy Yuka Hayama\*, Tomoko Inoue  
Osaka University Graduate School of Medicine, Division of Health Sciences, 1-7 Yamadaoka Suita, Osaka 565-0871, Japan
22. Guastello, S.J. Shissler, S. Dirscoll, J. & Hyde, T. (1998). Are some cognitive styles more creatively productive than others? *Journal of creative behaviour*, 32(2) 77-91.
23. Zhang, L.F (1999) Further cross cultural validation of the theory of mental self government *Journal of Psychology* 136 (3), 245-261.
24. Zhang, Z.F. (2002) Thinking styles & modes of thinking: Implication for education & research *Journal of Psychology*, 136(3) 245-261.
25. Helson, R., & Pals J.L (2000). Creative potential, creative achievement & personal growth. *Journal of personality*, 68(1), 1-2
26. Are only Emotional Strengths Emotional? Character Strengths and Disposition to Positive Emotions *aphw*. Angelika Güsewell and Willibald Ruch University of Zurich, Switzerland..

27. Fredrickson, B.L. (1998). What good are positive emotions? Review of General Psychology, 2, 300–319. doi:10.1037/1089-2680.2.3.300
28. Ellsworth, P. C., & Smith, C. A. (1988). Shades of joy: Patterns of appraisal differentiating pleasant emotions. Cognition and Emotion, 2, 301–331 for contentment.

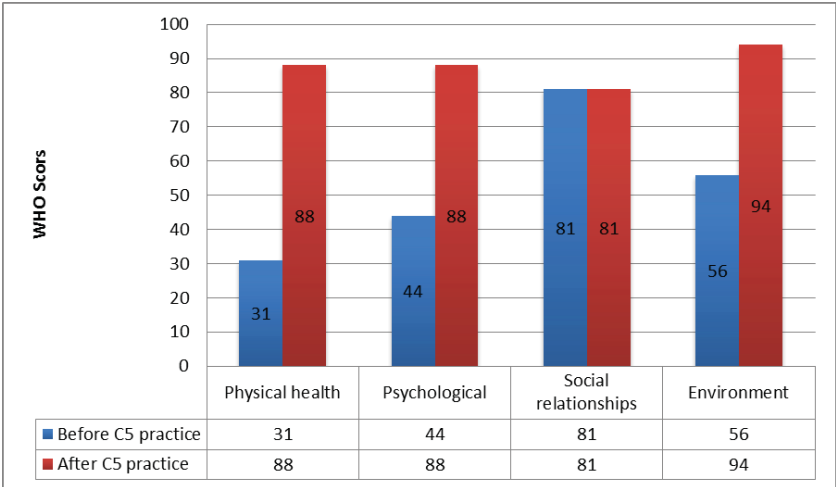


Figure 1: WHO QoL score for Case 1

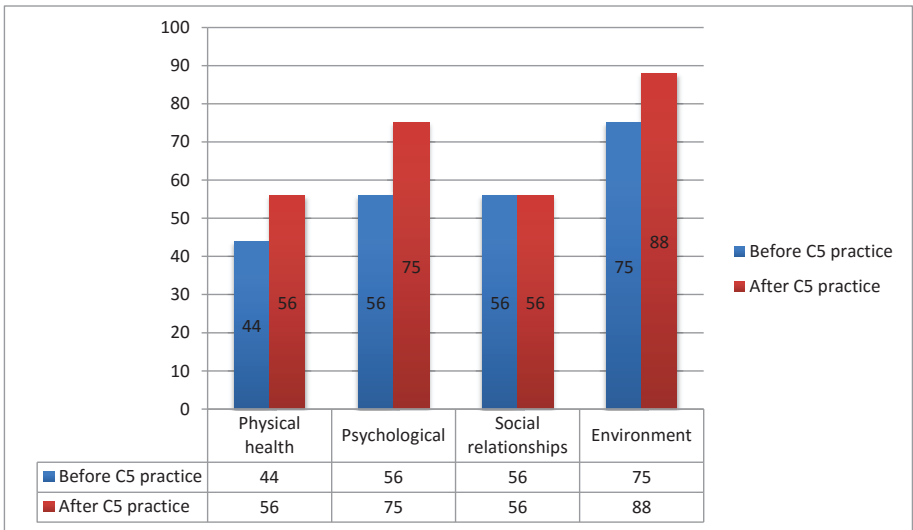


Figure 2: WHO QoL score for Case 2

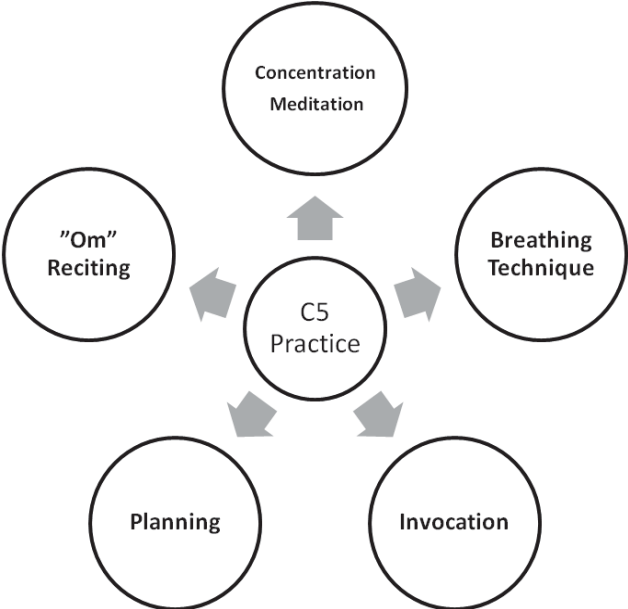


Figure 3:- Five basic exercises



**1**

“Wishing to have the full strength to run successfully all the operations for which I am responsible for achieving my mission.”



**2**

“Generator of the universe”



**3**

“Operator of the universe”



**4**

“Destructor of all evils”

Figure 4:- Invocation



Figure 5: Case 1 and Case 2 in concentrated mood



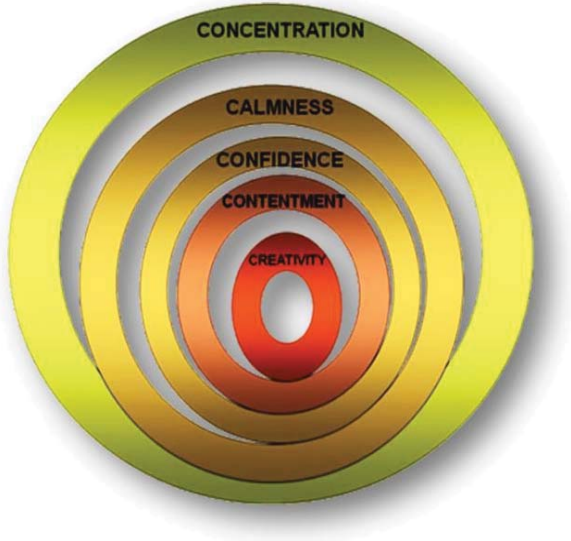


Figure 6: C5 concept

# The Practice and Impact of ATMAN on Human Mind and Body

Shobha Lala  
Center for TeleInfrastruktur Global Capsule (CGC),  
Aarhus University,  
Aarhus, Denmark,  
lala\_shobha@yahoo.com

Prof. (Dr.) Ramjee Prasad  
Center for TeleInfrastruktur Global Capsule (CGC),  
Aarhus University,  
Aarhus, Denmark,  
ramjee@btech.au.dk

**Abstract**— The era that we are living in, is the era of technology, technological advancements are faster than ever before. Everyday new models of phones, cars, computers, and other electronic equipment's are being launched. Communication technology has taken such a shape that everything is available on the tips of the fingers. Satellite communication, Global Positioning System (GPS), Internet and social networking have become an inseparable part of the daily life. There is a lot of pressure on the human mind to perform in different fields like education, jobs and social circle and so on. The employers want the employees to meet the goals and achieve targets at the earliest. The competition is very high in every field, be it the routine job, sports, films, education, culture, science or arts. However, for the human body, that took a long process of evolution, it has become difficult to maintain pace with the speed at which changes are taking place which is evident in the form of disease, frustration, fatigue and sleepless nights. Dependency on medicines, alcohol and drugs are increasing to get away from the stress and fatigue. In this paper we are investigating a novel method of natural process, resources and breathing pattern through a process ATMAN for development of C5 (Concentration, Confidence, Creativity, Calmness and Contentment).

**Keywords** □ □man Mind and □ody; ATMA□; □□;

## I. INTRODUCTION

ATMAN [1] is a one-hour process consisting of following processes,

1. Sun Gazing
2. Physical Activity
3. Relaxation
4. Laughter
5. Breathing Practice and,
6. Chanting [3-7]

ATMAN can be practiced by individuals from the age group of four and above. Regular practice of ATMAN brings discipline in life and positive values like love, compassion, harmony and peace can be inculcated. Children who practice ATMAN will have good values as well as a higher growth of C5 [2]. They will grow up to be confident with a clear vision and will be a compassionate lot who will contribute to the well-being of the society and on a global level. We have been practicing ATMAN for the past five years and have observed

significant changes in our life style, outlook, attitude and general conduct of life. We feel confident, concentrated, creative, calm and contented. We have grown positive traits like love, compassion, control and a sense of well-being. Negative traits like anger, jealousy and hatred seem to have been washed away from the body and mind. We feel healthy, happy, and at peace with a sound and undisturbed sleep at night.

Having felt the significant changes in our life we wanted to help others also to be free of stress, anxiety, fatigue and other negative traits. We started classes for practice of ATMAN. Housewives and school going children joined and started practicing ATMAN. All those who practiced ATMAN have felt significant changes in their lives. They observe improvement on physical as well as mental levels. We can say they are on the path of Healthy body and mind.

In this paper the impact of ATMAN is presented in the life of individuals at physical and mental level in context with C5 (Concentration, Confidence, Creativity, Calmness and Contentment). The impact observed has been based on the interview with the individuals and their families/friends. The impact is based on observation during the class, verbal interaction with practitioners and their families. These observations are noted on people from the age of 4 years to 75 years of age. The subjects are school going children, housewives, working professionals and retired persons. The subjects practiced ATMAN under our guidance for different durations the minimum being two weeks and the maximum being two years. The observations noted on the nine individuals are presented in this paper.

The paper is organized as follows. Section II presents the ATMAN concept. The case studies have been explained with individual and group results in Section III. Section IV gives the result of overall impact of ATMAN finally Section V concludes the paper.

## II. ATMAN CONCEPT

The first five years of an individual are the years of Innocence and the age up to 20 years is the age of learning. During this period the children develop concentration,

confidence and creativity. At this time, they have the maximum physical energy and it can be utilized to mould their personality. They have the potential to work for the utmost growth of positive traits. The human mind and body is like clay and a good potter or a good guide can mould them into exquisite pieces of art. This is the age to inculcate C5 in the individuals.

The age 21 to 50 years is the age to put the C5 concept into practice. This is the age when one faces a lot of challenges on the work front and the family front. The individuals become partners and create families and face the challenges of bringing up children as well as maintaining their work life and work hard to balance their personal and professional lives. They face the maximum stress during this time and are at times frustrated for not being able to maintain the balance. This period is the test period of putting C5 into practice.

The age 50 years onwards is the age of involvement when the individuals can share their experience of life with others and be role models for younger ones. These people can become the potters or the guides to the younger generations. Calmness and contentment are at the peak with which they enjoy the successes in their lives and can cherish the experiences of life.

The age of 75 and above or infinity is a full circle of life. The individual is like a child again, innocent, enthusiastic, care free and tension free. The life circle is complete and the individuals are ready to drift into another world happy, burden free and as light as feather and fresh like the new born baby.

III. CASE STUDIES

The study has been conducted considering the following format for 9 individuals:

- Name
- Age
- Gender
- Duration of the ATMAN practice
- Brief description about the individual
- General observation
- C5 growth presented through a bar graphs
- Age wise grouping and,
- Overall observation and representation of C5 growth through a pie chart based on the data obtained from the practitioners.

The chosen groups with details are given in Table 1.

Table 1. Group Details

Group	Age	Number of Participants
1	4-20 yrs	3
2	21-50 yrs	3
3	51-75 yrs	3

We will now describe the case studies with individual and group impact on practice of ATMAN.

Group 1

Subject 1

Name

Age

Gender

Duration of practice

Hiranya Wadhwa

four Years

Female

Four Weeks

Hiranya is the youngest individual to practice ATMAN. She was always full of energy, active and naughty. She had the most flexible body and had the perfect postures. She could even stand on her head for two minutes. It inspired the other kid Sheryl also and she also followed the practice of standing on the head and the two of them wanted to outdo each other. Hiranya loved the laughter exercise the most and enjoyed chanting too. Her presence in the class was like a dew drop, fresh and pure. Her C5 graph has been prepared on the basis of information provide by her mother Ms Jyoti Wadhwa. She told me that Hiranya was very keen on attending the class and made it a point to ask her to wake her up in the morning and make her ready for the class. She is keen on practicing ATMAN in future too. The impact of ATMAN on C5 for Subject 1 is depicted in Fig 1, showing maximum growth in the parameter of ‘Contentment’.

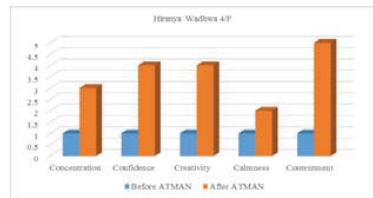


Figure 1. Subject 1 Results

Subject 2

Name

Age

Gender

Duration of practice

Sheryl Malik

Eight Years

Female

Four Weeks

Sheryl is a school going girl studying in class III. She practiced ATMAN a year ago during her summer vacation. She had some behavioral issues at that time. She used to be aggressive and at times rude to her elders and seemed to lack concentration. She practiced ATMAN for four weeks. Her C5 graph has been prepared with the help of her mother. Mrs Malik insisted that Sheryl should be contacted for the feedback and she herself should explain the changes that she felt. The C5 graph that was prepared for her is her own interpretation and the marking has been done by Sheryl herself. Her mother told us that after the practice of ATMAN she acquired a kind of discipline and her behavior also changed towards the elders, she became more obedient and calm and she also performed better at her school. She would definitely like to continue the practice of ATMAN in future too. The impact of ATMAN on C5 for Subject 2 is depicted in

Fig 2, showing maximum growth in the parameters of 'Contentment' and 'Calmness'.

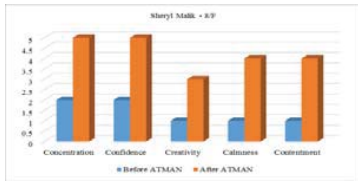


Figure 2. Subject 2 Results

**Subject 3**

Name  
Age  
Gender  
Duration of practice

*Rachit Ohri*  
*Thirteen Years*  
*Male*  
*Four Weeks*

Rachit Ohri is the son of Ms Neelima Ohri and started the practice of ATMAN at the insistence of her mother. He practiced ATMAN a year ago during summer vacation. He had a good flexible body and was able to do the physical activities effortlessly. He was eager to learn more postures and was able to achieve the postures without much difficulty. In his words the breathing exercise made him feel relaxed and calm and he was able to concentrate more on his studies. His grades improved and he started having a smile on his face which otherwise was expressionless. The practice gave a boost to his confidence and he was able to confront boys who earlier bullied him. Rachit has been practicing Atman on a regular basis. The impact of ATMAN on C5 for Subject 3 is depicted in Fig 3, showing a maximum growth in the parameters of 'Concentration', 'Calmness' and, 'Contentment'.

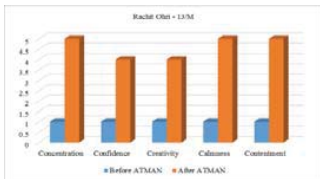


Figure 3. Subject 3 Results

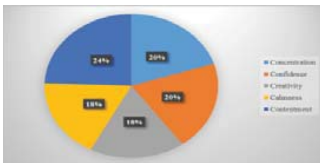


Figure 4. Group 1 Results

The overall improvement in C5 parameters for Group 1 is shown in Fig 4.

**Group 2**

**Subject 4**  
Name

*Purnima Lala Mehta*

Age  
Gender  
Duration of practice

*Twenty-Nine Years*  
*Female*  
*Six Weeks*

Purnima is a working professional and is pursuing her PhD. She has a hectic schedule of teaching at an Engineering College managing classes consisting of 50-60 students who are only a few years younger to her. She used to be exhausted and suffered a lot of fatigue, anxiety and sleep disorder. The long working hours, domestic chores and study work had taken a toll on her body and mind. During her break from her job, she joined the class and practiced ATMAN. She has observed significant changes in her body and mind. Her body has become more flexible and she has a better control over the body and she feels more energetic. With activities like Sun Gazing, Breathing exercise, relaxation and chanting she feels a significant growth of C5. Her sleep pattern has improved significantly and she has sound and peaceful sleep. In her words she feels very light and happy and at peace. She has been practicing ATMAN on a regular basis. The impact of ATMAN on C5 for Subject 4 is depicted in Fig 5, showing a maximum growth in the parameters of 'Concentration' and, 'Calmness'.

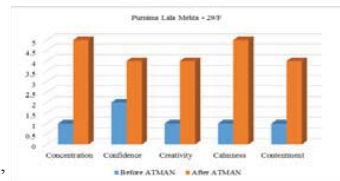


Figure 5. Subject 4 Results

**Subject 5**

Name  
Age  
Gender  
Duration of practice

*Neelima Ohri*  
*Thirty-Seven Years*  
*Female*  
*Six Months*

Neelima is a homemaker with two sons. She is responsible for taking care of the home and her parents in law. She had been suffering from a poor health after the birth of her younger son. Bringing up the baby and meeting the demands of the teenager son and other household chores had left her over worked and stressful. She felt her body was very stiff and felt pain in her knees and shoulders and she had a disturbed digestive system. She was suffering due to a disturbed sleep pattern. After the practice of ATMAN she observed a remarkable change at physical and mental level. She was so satisfied with the results that she asked her teenage son also to start the practice of ATMAN. She would like to continue the practice of ATMAN. The impact of ATMAN on C5 for Subject 5 is depicted in Fig 6, showing a maximum growth in the parameter of 'Calmness'.

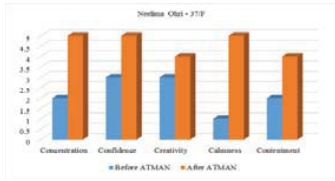


Figure 6. Subject 5 Results

**Subject 6**  
 Name  
 Age  
 Gender  
 Duration of practice

*Deepti Kureja*  
 Fifty Years  
 Female  
 Two Years

Deepti is a housewife and has to manage a family consisting of two grown up children, an ailing mother in law and a brother-in-law and his wife (both disabled). She has been multi-tasking to meet the requirements of the big family. She had been over worked, stressed, and frustrated. She joined the class to practice ATMAN two years ago and has been practicing since then. The practice of ATMAN has helped her to recognize herself as an individual and has made her aware of the hidden talents. She has gained a lot of concentration and confidence which has been reflected in her behavior and ability to handle situation. We have noticed a multi-dimensional growth in her behavior and personality. She has learned to play musical instruments and also joined hobby classes. She has a keen interest for cooking and making chocolates. Her creativity is evident in her personality. She says, "I would like to practice ATMAN till my last breath". The impact of ATMAN on C5 for Subject 6 is depicted in Fig 7, showing a maximum growth in the parameter of 'Calmness'.

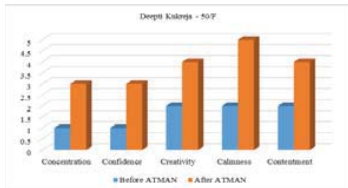


Figure 7. Subject 6 Results

The overall improvement in C5 parameters for Group 2 is shown in Fig 8.

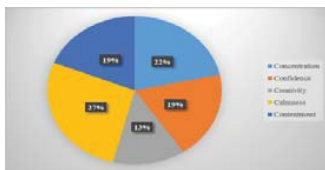


Figure 8. Group 2 Results

**Group 3**  
**Subject 7**  
 Name  
 Age  
 Gender  
 Duration of practice

*Vinod Bala*  
 Sixty-Three Years  
 Female  
 Eight Weeks

Vinod Bala is a retired school teacher who had been teaching for more than thirty years. She had spent a very hectic life earlier as she had to travel to her school in the bus for two hours. At school also she had to manage a large number of children and participate in various activities. She joined the class out of curiosity and once she started the practice of ATMAN, she was totally into it. She was very inquisitive and asked a lot of questions. She particularly enjoyed the laughter and chanting exercise. In her words she used to be very anxious, worried and uncomfortable at handling situations and certain things earlier. After Practice of ATMAN she gained confidence to handle situations in a confident and calm way which gave her a lot of contentment too. The level calmness that she has achieved is remarkable. She would like to continue practice of ATMAN on the bed of green grass, under the open sky surrounded by the chirping of birds. The impact of ATMAN on C5 for Subject 7 is depicted in Fig 9, showing a maximum growth in the parameter of 'Calmness'.

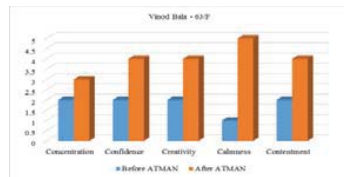


Figure 9. Subject 7 Results

**Subject 8**  
 Name  
 Age  
 Gender  
 Duration of practice

*Sheel Sangal*  
 Seventy-Five Years  
 Female  
 Eight Weeks

Sheel Sangal is the oldest, in age, practitioner of ATMAN. Her enthusiasm and curiosity was like that of a baby. She felt the happiest after breathing exercise and chanting. She was not able to perform all the physical activity but she always gave it a try and put her 100% effort in whatever she did. She said that she discovered her potential and feelings of confidence and calmness for the first time in life and she was glad that she started the practice of ATMAN. She felt calmer than before and was able to concentrate better. She would like to practice ATMAN for the rest of her life. She persuaded others to follow the practice of ATMAN. The impact of ATMAN on C5 for Subject 8 is depicted in Fig 10, showing a maximum growth in the parameter of 'Contentment'.

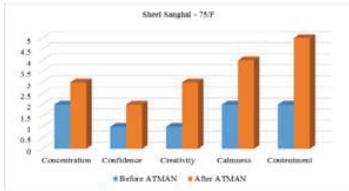


Figure 10. Subject 8 Results

Subject 9  
 Name  
 Age  
 Gender  
 Duration of practice

Geeta Gupta  
 Fifty-Two Years  
 Female  
 Six Weeks

Geeta Gupta is a homemaker and has two married daughters, one son and three grandchildren. She wanted to do something different and experience the change in her personality. She practiced ATMAN for six weeks. After the practice her outlook towards life, as a whole, changed completely. She discovered the positive traits of her personality. She said that she had never experienced such concentration and contentment and calmness in her life before. She feels very confident in handling situations and her conduct has changed and has revealed the positive sides of her personality. The changes have been noticed by the family and friends too. She keeps encouraging others to practice ATMAN and to know about their real self. She feels that maximum benefits can be achieved if ATMAN is practiced in open space - a park or a forest.

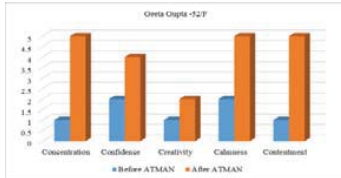


Figure 11. Subject 9 Results

The overall improvement in C5 parameters for Group 3 is shown in Fig 12.

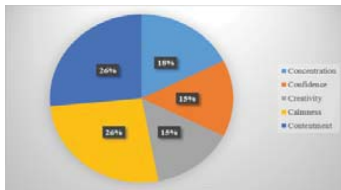


Figure 12. Group 3 Results

IV. PRACTICE RESULTS

The Fig 13 below depicts the average growth of nine individuals, divided in three groups. C5 growth of every

individual is presented by a bar graph and a comparative growth study is depicted by a pie chart for each group. Average growth of all the groups is presented through a pie chart. The average improvement of all the groups shows highest growth in Contentment and Calmness (23%), followed by Concentration (20%), Confidence (18%) and Creativity (16%).

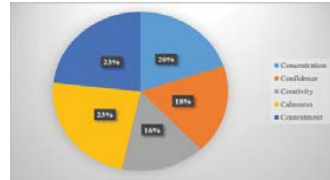


Figure 13. Average Growth Results

V. CONCLUSION

Impact of Atman has a relation with the duration of practice. The maximum impact has been observed on children, especially Rachit (13 years of age). Maximum growth has been observed in Calmness and Contentment followed by Concentration, Confidence and Creativity. Practice of ATMAN in open space like park or forest assures of maximum growth. All the practitioners who had sleep disorders experienced improved sleep patterns and had a peaceful sound sleep.

The “FEEL GOOD” factor was evident in the opinion of every participant. Participants felt more energetic, happy, contented and calm during and after practice of ATMAN. All the participants would like to continue the practice of ATMAN and would encourage others also to start the practice. ATMAN is a simple naturalistic method on the road of human evolution for creating peace and harmony in the world.

REFERENCES

- [1] Shobha Lala,Ramjee Prasad,'A Novel Naturalistic Concept Proposal "ATMAN" towards Human Evolution for Personal Bliss, submitted at Global Wireless Summit 2016.
- [2] Ramjee Prasad, 'Unlock Your Personalization', Aalborg University Press, 2012.
- [3] Bettina Hauge, 'Lives under the Sun', The Senses and Society Journal ,Vol. 10, Iss. 1, 2015.
- [4] Maria Fernandez-del-Valle, Tyrel S. McCravens,' Resistance Training Guidelines for Active Females Throughout the Lifespan: Children, Adolescents, Adult Women, and the Aging Woman', Springer Science Business Media, New York 2016.
- [5] Jonathan C. Smith, Roosevelt University Alberto Amutio University of DeLlsto, Spain John P. Anderson and Leslie A. Aria Roosevelt University.' Relaxation: Mapping an Uncharted World Biofeedback and Self Regulation', Regulation of Mental States and Biofeedback, vol 21 no. 1, 1996.
- [6] Rickard, K. B., Dunn, D. J., & Brouch, V. M.,' Breathing techniques associated with improved health outcomes', Virginia Henderson Global Nursing eRepository (2015, June 29).
- [7] Bernie,W., Candace, H., & Sydney, H. ' A Healthy Dose of Laughter and the Arts: The Role(s) of Play,Laughter and the Arts in Promoting Good Health and Maintaining Quality of Life'. Springer International Publishing Switzerland 2016.

# Supporting sustainable use of persuasive technology in the creation of sustainable business models

Annabeth Aagaard

Department of Business Development & Technology

Aarhus University

Herning, Denmark

aaa@btech.au.dk

**Abstract**— Sustainability has become a new premise for doing business - moving it from political discourses into company boardrooms. However, applying business model innovation as a way to create sustainable value requires several alterations of our ways of understanding and evaluating businesses and their business models. The potentials of persuasive technology in facilitating of sustainable business models are vast. The aim of this article is therefore to discuss the key areas of how persuasive technologies can be actively applied in a sustainable fashion to support sustainable business model innovation. Case examples are utilized to provide an understanding of the concept empirically. The findings and contributions of this study are summarized in propositions to be explored further in future research.

**Keywords**— sustainable business model innovation, sustainable business, persuasive technology

## I. INTRODUCTION

In the rise of globalization, sustainability has become a new premise for doing business - moving it from political discourses (Dryzek, 2005) into company boardrooms and company strategy (Bisgaard, 2009). However, the interrelation of business models, sustainability and innovation is still at an early stage (McGregor and Fontrodona, 2008; Preuss, 2011), and announced by many researchers as the next wave of innovation (Hall and Vredenburg; 2003; Maxwell, 2009; Boons et al., 2013). These new rules of the game force businesses to emphasize more sustainable business, sustainable business models and sustainable innovations in order to gain legitimacy and to reduce unsustainable practices as a minimum (Hall and Vredenburg, 2003; McGregor and Fontrodona, 2008). According to Hall and Vredenburg (2003) it is no longer enough for companies to create new radical innovations, if leading customer groups perceive these innovations and business as unsustainable.

In addition, solving persistent environmental and social problems is also considered as a new source of inspiration for businesses in pursuing innovative opportunities (Hansen et al., 2009; Aagaard, 2016). The constant need for businesses to stay at the innovative forefront has led them towards the most unexpected areas, where leading companies increasingly realize that complex social and environmental problems can

convert into new business opportunities (Kanter, 1999; Venn and Berg, 2013). In this context it is no longer enough for a business model to be novel and original in its technical features it has to be novel and original in terms of environmental or social sustainability as well (Phills et al, 2008). In the other perspective sustainability is merged into innovation where sustainable problems are seen as sources of inspiration for companies in generating new innovations and business opportunities (Perrine, 2013). The research question of this paper therefore emphasizes:

*“In what ways can we identify and develop sustainable business models using persuasive technologies?”*

## II. SUSTAINABLE BUSINESS & SUSTAINABLE BUSINESS MODELS

The study of sustainability is challenged by the fact that the concept is fragmented, where some researchers even question, whether sustainability is a concept or a political discourse (Dryzek, 2005) or an artefact (Faber et al., 2005). Although there is a growing body of literature in analyzing and discussing sustainability and sustainable development on the political and society-level (Dryzek, 2005), the operationalization of the concept in relations to business and on corporate level is still rather weak (Bansal, 2005; Stubbs & Cocklin, 2008; Zink et al., 2008; Carroll and Shabana, 2010; Aagaard, 2016).

The contribution of this study is therefore to provide a better understanding and operationalization of the concept of sustainable business models (SBM) through the sustainable use of persuasive technologies, while creating new sustainable opportunities and business models on corporate level (Holling, 2001; Newman, 2005; Schaltegger et al., 2012).

The most common translation of sustainability in business is the triple bottom line, which consists of three sustainable dimensions: people, planet and profit (Elkington, 1997) and is described as three equally important managerial principals (Bansal, 2005; Hansen et al., 2009; Bradbury-Huang, 2010; Schaltegger and Wagner, 2011). Based on the triple bottom line, Charter and Clark (2007) offer a definition of sustainable innovation embracing all of these elements:

*“Sustainable innovation is a process where sustainability considerations (environmental, social and*

*financial) are integrated into company systems from idea generation and development (R&D) and commercialization. This applies to products, services and technologies, as well as to new business and organizational models (Charter and Clark, 2007:9)".*

In defining and exploring the theoretical concept of sustainable business model innovation (SBMI) the starting point would be state of the art definitions of business model innovation. Business models appear in many different forms. They can be applied as a core unit of analysis extending beyond the business boundaries (e.g. Zott and Amit, 2007). In addition, business models may be viewed as a construct between strategy and implementation (Baden-Fuller and Morgan, 2010). Business models can also be a mean for commercializing new technologies (Chesbrough and Rosenbloom, 2002; Chesbrough, 2007; 2010) and as an intermediary between different innovation actors such as businesses, financiers, research institutions, etc., i.e., actors who shape innovation networks (Doganova and Eyquem-Renault, 2009). Business models can therefore be subject to innovation themselves, or a template for implementing managerial initiatives (Zott and Amit 2010).

Baden-Fuller and Morgan (2010) stress that from a holistic and systemic concept a business model perspective may be expected to contribute to a sustainable business model innovation agenda by opening up new approaches to overcoming internal and external barriers. Birkin et al. (2009a, 2009b) identified in their study on North European and Chinese businesses that societal and cultural demands of sustainable development evolve outside the economic sphere as drivers for business model change in businesses. Earlier work reveals the first developments in mapping the concept of and movements towards sustainable business model innovation (Demil and Lecocq, 2010; Chatterjee, 2013).

Lovins et al. (1999) propose four steps in designing SBM: increase of natural resources' productivity; imitation of biological production models; change of business models; and reinvestment in natural capital. One interesting and recent contribution in mapping SBM comes from Boons and Lüdeke-Freund (2013), who define three different types of sustainable business models, which create social value and maximize social profit focusing on three different areas (p. 14-15):

**Technological innovation:** creating a fit between technology characteristics and (new) commercialization approaches that both can succeed on given and new markets.

**Organizational innovation:** implementing alternative paradigms that shape the culture, structure and routines of organizations and thus change the way of doing business towards sustainable development.

**Social innovation:** that helps in creating and further developing markets for innovations with a social purpose.

### III. PERSUASIVE TECHNOLOGIES IN SUPPORT OF SUSTAINABLE BUSINESS MODELS

Persuasive technologies can play an active role in the creation of sustainable business models and in relation to all

three types of SBMs as previously suggested by Boons and Lüdeke-Freund (2013). Basically, persuasive technology is designed with the intent to change a particular aspect of human behavior in a predefined way (Dominic et al., 2013). Numerous authors emphasize that persuasive technology is one of the most effective tools to help change others' attitudes or behaviors (Fogg, 2003; Dillard and Pfau, 2002; Chatterjee and Price, 2009; Khalil and Abdallah, 2013).

According to Fogg and Nash (1997) using available present-day technology to assist persuasion is almost as effective as persuasion itself. As tools, persuasive technologies can increase people's ability to perform a target behavior by making it easier or restructuring it (Fogg, 2003). For example, an installation wizard can influence task completion and complete tasks such as installation of additional software not planned by users (Aagaard and Lindgren, 2015).

In an empirical exploration of the use of persuasive technologies in generating sustainable business models, the framework of Boons and Lüdeke-Freund (2013) is applied. Their first category, technological innovation, and applied in a persuasive context primarily emphasize the commercial/business side of what persuasive technology can do for sustainable business modeling.

This type of persuasive SBM is here named: business-driven SBM. A case example hereof is LifeLink, which is a water pump business system created by Grundfos that allows for communities with limited access to clear water to be provided/buy access to water. The persuasive technologies incorporated include among other sensors, persuasive mobile communication and robotics and easy access water-stations with chip-cards that only has to be inserted in a water-automat to provide clean water.

The system feeds back information about e.g. the water quality and water usage of each community, family and person, which are applied to improve the pumps and the water supply. However, another value of the business model is water/health data, which is shared with/sold to NGO's as big data on e.g. water quality status and health situations across communities and borders. As this example reveals the persuasive SBM examples are seldom only business-driven, as a business model may only truly be considered sustainable if it benefit more than just the business. However, the main objective of this type of persuasive SBM is business.

The second category, organizational innovation, addresses the ability to change organizational behavior as part of the sustainable business model. This category of persuasive SBM is here called Organization-driven SBM. Examples hereof could be Hygiene Guard, which is a surveillance system that tracks employees' hand washing using sensor in e.g. the employees' ID badges, in the restrooms, canteen etc. The applications help ensure better hand hygiene, which minimize number of employee's sick days and foodborne diseases in canteens. This example has both elements of business- driven, organization-driven and social driven SBMs and stresses that persuasive SBMs often have multiple focus areas.

The latter category, social innovation, emphasize the opportunities to change social behaviors and conditions as part



of the SBM, and this category of persuasive SBM is therefore here called, Social-driven SBM. An example hereof is persuasive-technology- based interventions such as Telecare. The business model of Telecare system emphasizes diminishing social isolation and low physical activity among older adults as this represents a significant societal challenge and health risk in which persuasion offers potential solutions.

Through door, room and individual sensors, communication systems as well as persuasive training devices, the elderly person and their families can not only feel more secure, but also increase their social interactions and physical activities. As this example reveals Social-driven SBM emphasizes a social issue and provides social solutions, where the business/commercial element is of less value.

#### IV. DISCUSSION AND CONCLUSION

The concept of sustainability changes and evolves over time and across contexts, so what may be viewed as a sustainable and as a sustainable business model today may be considered un-sustainable over time. This does not imply that the unique opportunities of applying persuasive technologies in facilitating more sustainable behaviors in business, organizations and social environments should be ignored. However, the level of sustainability of the business and business models should be evaluated continuously across Elkington's three P's (people, planet and profit) in accordance to what is considered sustainable in the specific period of time and in the specific context, where the business model is integrated.

**Proposition 1:** For persuasive technology to support sustainable business models, the level of sustainability should be evaluated on the prevalent standards of sustainability in the given time and context of the actual integration and application of the sustainable business model.

The main objective of persuasive technology is to change behavior. However, who is to decide what is considered sustainable behavior? How can persuasion be carried out in ways that ensure a transparent, sustainable and ethical use of persuasive technologies?

**Proposition 2:** In ensuring sustainable business models through ethical applications of persuasive technologies, a transparent communication and control device has to be installed to protect stakeholders from the unsustainable and unethical use of persuasive technologies.

The findings reveal that persuasive-driven SBM's seldom only focused on one area, but are combinations of business-driven, organization-driven and social-driven SBM's. However, the empirical applications of persuasive technology in forming SBM typically have a main emphasis – business, organization or social from the start.

**Proposition 3:** In developing sustainable business models through the use of persuasive technologies, the SBM typically has one primary focus (business, organization or social) and secondary focus, which is sustainable and beneficiary, but not the main emphasis of the business.

Finally, another interesting discussion is, who should benefit from the persuasion, for a business model to be sustainable? In business applications persuasion typically benefits the businesses alone selling and persuading their customers to buy and/or use their specific products. However this is a one-way benefit. It is therefore suggested that for a persuasive business model to be sustainable, the business of the business model has to benefit multiple stakeholders to be truly sustainable.

**Proposition 4:** For persuasive business models to be sustainable, they have to benefit more than the business (one-way), but multiple stakeholders (multiple-ways).

All of these propositions are to be explored in a multi-case-study across various industrial contexts and PT applications in exploring the development of SBM through PT.

The managerial implications of the study reveal three overall approaches to be applied in mapping and designing sustainable business model innovations (technological, organizational and social) as well as a way to evaluate the level of sustainability of the different dimensions of the business model through Elkington's three P's.

The limitations of the study also reveal new venues for further research. For one, this paper is conceptual, so the suggested approach in determining and evaluating sustainable business models should be explored further through a cross-industry case study to ensure the applicability of the approaches across contexts.

Secondly, other metric systems may have to be suggested and applied, which can measure the level of sustainability both a qualitative and quantitative level. Finally, different organizational factors like size and/or maturity of company/industry may affect the choice of SBM approach, which would also be of interest in a further research study.

#### REFERENCES

- [1] Aagaard, A., 2016. Sustainable Business – Integrating CSR in Business and Functions. River Publishers.
- [2] Aagaard, A. & Lindgren, P., 2015. The Opportunities and Challenges of Persuasive Technology in Creating Sustainable Innovation and Business Model Innovation. *Wireless Personal Communication*, 81, 1511–1529.
- [3] Baden-Fuller, C. & Morgan, M.S., 2010. Business models as models. *Long Range Planning* 43, 156-171.
- [4] Bansal, P., 2005. Evolving sustainability: A longitudinal study of corporate sustainable development. *Strategic Management Journal*, 26, 197-218.
- [5] Birkin, F., Cashman, A., Koh, S.C.L., Liu, Z., 2009a. New sustainable business models in China. *Business Strategy and the Environment* 18, 64-77.
- [6] Birkin, F., Polesie, T., Lewis, L., 2009b. A new business model for sustainable development: an exploratory study using the theory of constraints in Nordic organizations. *Business Strategy and the Environment* 18, 277-290.
- [7] Bisgaard, T., 2009. CSI. Corporate Social Innovation - Companies participation in solving global challenges. Erhvervs- & Byggestyrelsen.
- [8] Boons, F. & Lüdeke-Freund, F., 2013. Business models for sustainable innovation: state-of-the-art and steps towards a research agenda. *Journal of Cleaner Production*, 45, 9-19.

- [9] Boons, F., Montalvo, C., Quist, J. & Wagner, M., 2013. Sustainable innovation, business models and economic performance: an overview, *Journal of Cleaner Production*, 45, 1-8.
- [10] Bradbury-Huang, H., 2010. Sustainability by collaboration: The Seer case, *Organizational Dynamics*, 39, 335-344.
- [11] Carroll, A. B., & Shabana, K. M., 2010. The Business Case for Corporate Social Responsibility: A Review of Concepts, Research and Practice. *International Journal of Management Reviews*, 12(1), 85-105.
- [12] Charter, M. and Clark, T., 2007. Sustainable Innovation – Key conclusions from Sustainable Innovation Conferences 2003-2006. The Centre for Sustainable Design. University College for the Creative Arts. May 2007. www.cfsd.org.uk
- [13] Chatterjee, S., & Price, A. 2009. Healthy living with persuasive technologies: Framework, issues, and challenges. *Journal of the American Medical Informatics Association*, 16(2), 171–179.
- [14] Chatterjee, S. 2013. Simple Rules for Designing Business Models. *California Management Review*, 55, 97-124.
- [15] Chesbrough, H. and Rosenbloom, R., 2002. The role of the business model in capturing value from innovation. *Industrial and Corporate Change*, 11(3): 529–556.
- [16] Chesbrough, H., 2007. Business model innovation: it is not just about technology anymore. *Strategy and Leadership*, 35 (6), 12-17.
- [17] Chesbrough, H., 2010. Business model innovation: Opportunities and barriers. *Long Range Planning*, 43, 354-363.
- [18] Demil, B. & Lecocq, X., 2010. Business Model Evolution: In Search of Dynamic Consistency. *Long Range Planning*, 43(2), p. 227-246.
- [19] Dillard, J., & Pfau, M. 2002. *The persuasion handbook: Development in theory and practice*. Thousand Oaks, CA: Sage Publications.
- [20] Doganova, L., Eyquem-Renault, M., 2009. What do business models do? Innovation devices in technology entrepreneurship. *Research Policy* 38, 1559-1570.
- [21] Dominic, D., Hounkponou, F., Doh, R., Ansong, E., & Brighter, A. (2013). Promoting physical activity through persuasive technology. *International Journal of Inventive Engineering and Sciences (IJIES)*, 2(1), 16–22.
- [22] Dryzek, J. S., 2005. *The politics of the Earth: Environmental Discourses*, Oxford University Press.
- [23] Elkington, J., 1997. *Cannibals with forks: Triple Bottom Line of 21st Century Business*, Capstone Publisher Limited, Oxford.
- [24] Faber, N., Jorna, R. & Van Engelen, J., 2005. The sustainability of "sustainability" – A study into the conceptual foundations of the notion of "sustainability", *Journal of Environmental Assessment Policy and Management*, 7(1), 1-33.
- [25] Fogg, B. J., & Nass, C. 1997. Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*, 46(5), 551–561.
- [26] Fogg, B. J. 2003. *Persuasive technology: Using computers to change what we think and do*. San Francisco, CA: Morgan Kaufmann Publishers.
- [27] Hall, J. & Vrendenburg, H., 2003. The challenges of innovating for sustainable development, *MIT Sloan Management Review*, 45, 61-68.
- [28] Hansen, E.G., Große-Dunker, F., Reichwald, R., 2009. Sustainability innovation Cube - a framework to evaluate sustainability-oriented innovations. *International Journal of Innovation Management* 13, 683-713.
- [29] Holling, C.S., 2001. Understanding the Complexity of Economic, Ecological, and Social Systems, *Ecosystems*, 4, 390-405.
- [30] Kanter, R.M., 1999. From spare change to real change: the social sector as a beta site for business innovation. *Harvard Business Review* 77, 123-132.
- [31] Khalil, A., & Abdallah, S. (2013). Harnessing social dynamics through persuasive technology to promote healthier lifestyle. *Computers in Human Behavior*, 29(2013), 2674–2681.
- [32] Lovins, A.B., Lovins, L.H., Hawken, P., 1999. A road map for natural capitalism. *Harvard Business Review*, 1-14 (HBR paperback reprint 2000).
- [33] Maxwell, I. E., 2009. Managing sustainable innovation: The driver for global growth. Springer. McGregor, S. P. & Fontrodona, J., 2008. Exploring the fit between CSR and innovation, Center for Business in Society, Working paper.
- [34] Newman, L., 2005. Uncertainty, innovation, and dynamic sustainable development, *Sustainability, Science, Practice & Policy*, 1(2), 25-31.
- [35] Perrine, F., 2013. The complementarity of corporate social responsibility and innovation: Evidence from Belgian firms, *Global Journal of Business research*, 5(5), 99-113.
- [36] Phills, J. A., Deiglmeier, K. & Miller, D. T., 2008. Rediscovering social innovation, *Stanford social innovation review*.
- [37] Preuss, L., 2011. Innovative CSR: A framework for anchoring corporate social responsibility in the innovation literature, *Journal of Corporate Citizenship*, 42, 17-32.
- [38] Schaltegger, S., Wagner, M., 2011. Sustainable entrepreneurship and sustainability innovation. Categories and interactions. *Business Strategy and the Environment*, 20 (4), 222-237.
- [39] Schaltegger, S., Lüdeke-Freund, F., Hansen, E.G., 2012. Business cases for sustainability - the role of business model innovation for corporate sustainability. *International Journal of Innovation and Sustainable Development* 6 (2), 95-119.
- [40] Stubbs, W. & Cocklin, C., 2008. Conceptualizing a sustainability business model, *Organization & Environment*, 21(2), 103-127.
- [41] Venn, R. and Berg, N., 2013. Building competitive advantage through social entrepreneurship, *South Asian Journal of Business Research*, 2(1), 104-127.
- [42] Yunus, M., Moingeon, B., Lehmann-Ortega, L., 2010. Building social business models: lessons from the Grameen experience. *Long Range Planning* 43, 308-325.
- [43] Zink, K. J., Steimle, U. & Fisher, K., 2008. Human factors, business excellence and corporate sustainability: Differing perspectives, joint objectives. In Zink, K. J. (Eds). *Corporate sustainability as a challenge for comprehensive management*. Physica-Verlag.
- [44] Zott, C. & Amit, R., 2007. Business model design and the performance of entrepreneurial firms. *Organization Science* 18, 181-199.
- [45] Zott, C., Amit, R., 2010. Business model design: an activity system perspective. *Long Range Planning* 43, 216-226.

## “Human Bond Communication in a world of persuasive business models embedded with Disruptive Technologies”

Peter Lindgren & Annabeth Aagaard

Aarhus University, Business Development and Technology, Denmark

[Peterli@btech.au.dk](mailto:Peterli@btech.au.dk) and [aaa@btech.au.dk](mailto:aaa@btech.au.dk)

### ABSTRACT

*Disruptive and persuasive technologies are being developed with exponential speed and are coursing disruption of businesses and business model (BM) ecosystems every day. Businesses put emphasis on human bonds related to persuasive BMs and how they can be innovated and operated to ensure the full business potential but do they – or we – actually also secure sustainable human bond communication (HBC) or are we just victims of a persuasive disruptive technology race. BM ecosystems [15] and longtime based built up human bonds can vanish within split seconds by these technologies – and how is and will this change the bonds between humans, humans and machines and humans and nature. How will this evolvement influence humans and machines ability to “sense”, “relate” and “communicate” each other. The paper addresses the exponential development of persuasive technologies and persuasive BMs and how it disrupt existing human bonds and HBC - but potentially also create new ones.*

Keywords: Human Bond, Human Bond Communication, Disruptive Persuasive Technology, Disruptive Persuasive BMs, BM Innovation,

### 1. Introduction

Sensors, wireless and persuasive technologies in our everyday life have increased and developed exponential lately. We are moving towards a world of persuasive BMs, which are embedded with persuasive technologies [5],[6][4], [23],[19]. Robots that are “persuaded” or more positive learns from humans, humans that are “persuaded” or learns from machines – where Pokemon is just the latest example of such development.



Figure 1. Human to machine and machine to human persuasion – or learning?

The development and innovation of persuasive BMs embedded with persuasive technologies is inevitable – and is still an uncharted research territory with security, ethical and trust issues. On behalf of workshops between 2010 – 2016 [24], [25], [19], lab experiments combined with literature studies of state of the art persuasive BM and persuasive technology (PT) research, we provide a conceptual discussion and outlook on behalf of our learning of what we have experienced that we can expect from persuasive BMs embedded with disruptive

persuasive technologies related to the topic HBC. The paper addresses an interdisciplinary context and with an interdisciplinary approaches the question:

*How to secure and develop HBC competences for the future betterment of human in a world of disruptive persuasive BMs?*

During the last decades disruptive technologies have changed humans, human to machine communication to become ...



**Figure 2. Technologies as .. or as a friend or a partner in BM innovation**

Why focus on BM Innovation (BMI), Multi BMs and disruptive persuasive technologies in this context? “BM Innovators earns an average premium that are several times greater than that enjoyed by product or process Innovators” [2],[7], [8]. “BMI delivers return that are more sustainable than product and process innovation”. However we do not today know much about the “value” the persuasive BMs embedded with disruptive technologies gives us as human being and to our businesses [19].

We know there is and hopefully always will be a strong dependency between Technology and BMI. “Disruptive Technology” have until now always been in need of a BM or actually a multitude of BMs. “No Disruptive technologies” - neither very advanced disruptive technologies embedded in persuasive BMs – have been able “to go or do without a BM” – or actually many BMs”. However and maybe humans have forgotten or have neglected how the BMs behind the disruptive technologies actually look like and actually are related.

Why – because humans and business still basically lack a language and framework concept of BMs. Many researchers have attempted to define the BM concept [1], [21], [9], [4], [11], [7], but a general language and framework concept of a BM is still missing [26], and not agreed upon. To have this language to be useful and related to persuasive BMs, it has been argued that a persuasive BM framework must be as reasonably simple, logical, measurable, comprehensive, as well as operational and meaningful as a general BM description [28],[14]. A persuasive BM embedded with disruptive persuasive technologies should meet the same qualifications for humans to understand the values, relations and impact of persuasive BMs as for the general BM framework.

In a preview for the role of the persuasive BM [14] we claimed that this type of BM will try to attach to anything, anybody, anywhere and anytime [10] alone but maybe more important together with other BMs. Soon everyone and everything at any time and in any place will have the possibility but also the risk of acting and being persuasively interfered with a persuasive BM. BM based on disruptive persuasive technologies will “represent an important and exiting vehicle for BMI” “maybe also a source of BMI in and of itself” in the future. “BM based on disruptive persuasive technologies gives us a new dimensions of BMI, distinct, albeit complementary, to traditional dimensions of BMI, such as product-, process- or organizational innovation” (Inspired by Mazza and Tucci) [17]

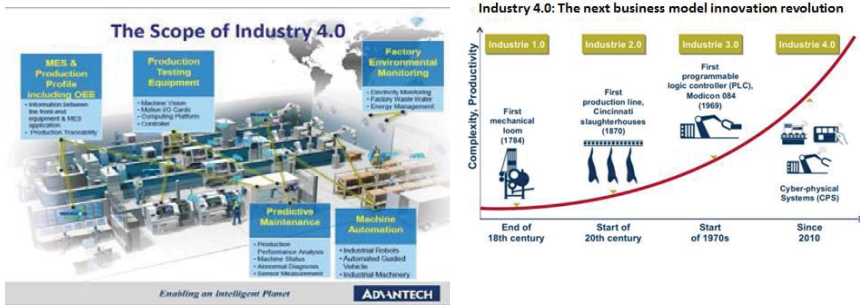


Figure 3 Industry 4.0 and cyber-physical systems (CPS) related to persuasive BMs

However where is the human and human bonds related to these persuasive BMs. In the last decade we have seen the human slowly vanish from some of the earlier most essential BMs to humans – e.g. production of food.



Figure 4 From manual to .....“robot” based production to ... artificial production with few or no humans.

As the businesses and BMs evolves towards cyber physical systems – and the BMs becomes more and digital and virtual – then “the platform” for persuasive BMs and persuasive technologies are laid.



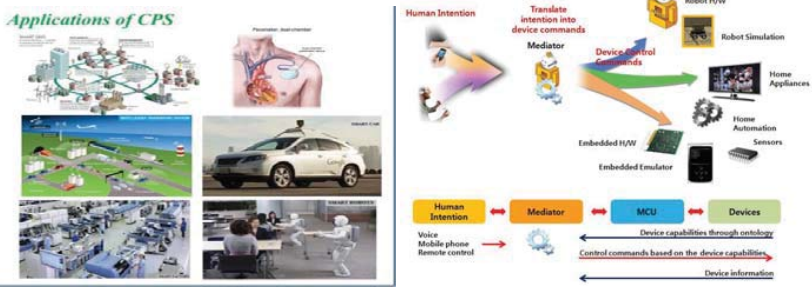
Figure 5. From robot based production to .... Value based Production ... to persuasive based production.

As our society moves fast towards a Cyber-Physical world and Cyber-Physical Systems (CPS) where integrations of computation, networking, and physical processes and behavior becomes embedded with computers and networks monitor and increasingly controls and motivate the physical processes, with manifolds of feedback loops where physical processes affect computations and vice versa the questions raises where are the human being in this “game” and what are the impact to the human being [16]. “Are we leaving the woods” and/or are we losing control and our ability to “sense” through our by nature given 5 senses.



Figure 6. Human sensing and learning in cyber physical systems

Or are we maybe learning some new “sensors” and learning some new ability to sense the world – developing some new sensing competences that can be highly valuable in future



BMI.  
Figure 7. CPS and human interaction with mobile devices.

We are still “learning” to some extent what disruptive persuasive technologies “technical” can do and be used for and do -but we know very little about what the impact is on our “AS IS” BMs and How we can create our “TO BE” persuasive BMs. We know also very little about what we can expect of these persuasive BMs and what impact they will have on the human being, human bonds and HBC. What we can observe at the moment is that humans are highly influenced by PT and persuasive BMs – in some case actually also seeing that humans are hurting and stressing itself because of the persuasive “values” that the persuasive BMs “creates”, “delivers”, “demands” and the human “receives”, “captures” and “consumes” – tangible and/or intangible values.



Figure 8: Persuasive BMs interacting with humans

## 2. Discussion

Businesses, “technologies” and humans will eventually – at an optimum - have to develop new and persuasive BMs in symbiosis - that can leverage the business potential of these technologies and BMs in the betterment of the human being. Many predict that the extent to which persuasive technologies will disrupt and affect our future lives will be extensive and go along with high expectations for better business and a noticeable impact on the global economy. [6], [20]. Particularly, health care, and well care sectors are expected to be able to diminish their high economic burden with the help of these technologies. But will it actually be in the betterment of the human being? – and how can we ensure this vision can come true?. The gaming, event and retail industry will soon be some of the sectors to value from this development, but also the Information and News industries are expected very soon to be influenced. Researchers, businesses and public players alike are devoting themselves to understanding how these technologies and BMs might be designed, so that remunerative technologies, behavior and BMs are within reach and can be created, captured, delivered, received as well as consumed in new, and hopefully secure and sustainable ways. [10], [14]. In explanation, pure disruptive PT has digital components embedded and is situated within two dimensions. The first dimension contains knowledge about how people and BMs can be triggered to change behavior – thereby create, capture, deliver, receive and consume persuasive values. A second dimension contains the PT, in which advanced Information Communication Technology [ICT] is in some way embedded and can “communicate” with the human, thing or both in combination. in an integrated physical, digital and virtual world. PT and persuasive BM however get much more powerful when they are capable of influencing all worlds, all five senses of the human being and “machines” and this in combination. In practice, this could be a disruptive game sending certain persuasive values – “hidden pokemons” [19], so that a certain reaction is triggered by young students at a university – and even drivers at a highway. It could be a “cupboard that vanish in the ceiling” and “hides” the young girls dresses – and when the mobile phone “sense” it then the right dress for the right “mood” to the right context turns up. It can also be technology that changes colors or forms of equipment in an elderly persons home so that fall is prevent or the elderly women is saved from taking wrong pill.

This implies that PT and the knowledge on behavioral change behind it can steer a process in detail from starting point to desired goal on behalf of big data, created by sensors, captured by sensors, delivered by IOT, received by sensors – both technology based sensors and human sensors and consumed by both machines and humans – and in combination. Possible negative scenario could be businesses taking control and persuading others to act against what is best in their interests, and that as a result people and even businesses could get hurt economically – but not least at their competences – technologies, human resources, organizational systems and culture. Also that humans loses their ability to sense [16] – the tangible ability to sense what is around them but maybe more critical the intangible sensors that we yet know so little about [16]. Previously we claimed [14] that tomorrow’s network based persuasive BM will consist of BMs with all kinds of security structures, strategies/objectives and offers, including virtual, digital and physical value propositions and security systems. Thus, persuasive BM providers and developers are facing a world, where we innovate and operate persuasive BMs within a multitude of BMs and a variety of security systems [18].

Nevertheless, we still have challenges of security, trust and ethics to overcome before we reach the point where all BMs become persuasive [25],[19]. An even stronger focus on security, as personal security and network based security technologies, and BMs that are persuasive, in process and changing continuously in different BMES context, sets businesses and researchers under high pressure to find solutions. Solutions, that both technologically and business-wise, will have to meet the needs of all kind of stakeholders for increasing persuasiveness, agility, flexibility, individualization, and privacy. [14]. We believe that

persuasive BM that is based on advanced security technology will provide those businesses with key competitive advantages – in the betterment of the human.

### 3. Conclusions

Disruptive Persuasive technologies are and will be increasingly integrated in more and more advanced persuasive BMs. Persuasive BMs can be operating physically, digitally and virtually – integrated, connected and with anybody, anything, anywhere and anytime in the BMI process. The persuasive BM can create, capture, deliver, receive and consume value propositions, wherever and whenever the user, customer, network partner, employees, things and businesses model demand it – want it or need it. Future persuasive BMs will not only be a matter of security, ethics and trust on the ‘surface’ of things, places, people and time, but also as part of the BMs related to both inside and outside things, bodies and time [25], [19]. How persuasive BMs can be controlled and secure in this context is still open to investigate. As pointed out in this article, there are still many unanswered questions to be explored before reaching an enhanced understanding of how businesses and human can value from persuasive BMs and how they can ensure security, ethics and trust in the process for the betterment of the human. In this paper we have shown some examples of how to understand and explore disruptive persuasive BMs and how it potentially can influence HBC with human and things.

### References

- [1] Afuah, Allan, and Christopher L. Tucci. (2001) *Internet BMs and Strategies: Text and Cases*. Boston: Irwin/McGraw-Hill, 2001.
- [2] Boston Consulting Group (2009) Business Week/BCG Innovation Survey.
- [3] Chesbrough, H. “The Role of the BM in Capturing Value from Innovation: Evidence from Xerox Corporation’s Technology Spin-off Companies.” *Industrial and Corporate Change* 11, no. 3 (June 1, 2002): 529–55.
- [4] Chesbrough, H. (2006) *Open BMs: How to Thrive in the New Innovation Landscape*. Boston, Mass: Harvard Business School Press
- [5] Fogg, B. J. *PT: Using Computers to Change What We Think and Do*. The Morgan Kaufmann Series in Interactive Technologies. Amsterdam ; Boston: Morgan Kaufmann Publishers, 2003.
- [6] Fogg, B. J., ed. *Mobile Persuasion: 2007 Perspectives on the Future of Behavior Change*. Stanford, Calif: Stanford Captology Media,
- [7] Gassmann, Oliver; Frankenberger, Karolin & Csik, Michaela (2013) The St.Gallen BM Navigator. Working Paper: ITEM-HSG
- [8] IBM (2012) IBM Global CEO Study – The Enterprise of the Future.
- [9] Johnson M.W., Christensen, C. and Kagermann, H. (2008) Reinventing your BM, Harvard Business Review, vol. 86 No. 12, pp. 50-59
- [10] Lindgren, Peter, Morten Kamøe Sondergaard, Mark Nelson, and B.J. Fogg. “Persuasive BMs.” *Journal of Multi BMI and Technology* 1 (2013).
- [11] Lindgren, Peter, Yariv Taran, and Harry Boer. “From Single Firm to Network-Based BMI.” *International Journal of Entrepreneurship and Innovation Management* 12, no. 2 (2010): 122.
- [12] Lindgren, Peter, Yariv Taran, and Kristin Margrethe Saghaug. “A Futuristic Outlook on BMs and BMI in a Future Green Society.” *Wireless Personal Multimedia Communications Symposia Proceedings*, August 10, 2010, 23.
- [13] Lindgren, Peter; Aagaard, Annabeth “The Sensing BM”  
In: “*Wireless Personal Communications*”, Vol. 76, No. 2, 2014, p. 291-309
- [14] Lindgren, Peter (2015) The secure persuasive BM IEEE Conference paper Hyderabad 2015.
- [15] Lindgren, Peter (2016) The BM Ecosystem Journal of Multi BMI and Technology.
- [16] Louv Richard (2013) Last Child in the Woods: Saving our Children from Nature-Deficit Disorder Atlantic Books Ltd, 4. jul. 2013 - 300
- [17] Mazza and Tucci (2013) BMI The Oxford Handbook of Innovation Management Chapter 21.
- [18] Mucchi, Lorenzo, Luca Simone Ronga, and Enrico Del Re. “Physical Layer Cryptography and Cognitive Networks.” In *Trustworthy Internet*, edited by Luca Salgarelli, Giuseppe Bianchi, and Nicola Blefari-Melazzi, 75–91. Milano: Springer Milan, 2011.
- [19] NJIT 2016 - Kickoff Operation of CTIF Global Capsule (CGC) jointly with the Round Table Discussions on Knowledge Home, New Jersey Institute of Technology (NJIT) September 8, 2016 sponsored by IEEE
- [20] Onur, Ertan. “TU Delft Presentation SW2010 Held at the Strategic Workshop on Sensor Networks in Florence, July 2010.”
- [21] Osterwalder, A., Pigneur, Y., & Tucci, C. (Vol. 15. May 2005). Clarifying BMs. Communications of AIS.
- [22] Prasad Ramjee 2014 5G: 2020 and Beyond. River Publishers, 2014. 126 p. (The River Publishers' Series in Communications).
- [23] Seiter, John S. and Gass, Robert H. *Persuasion, Social Influence, and Compliance Gaining*. Fifth edition. Boston: Pearson, 2014.
- [24] SW2010. Strategic Workshop 2010 Held at Hotel Relais Certosa, Florence, Italy. Twelfth Strategic Workshop Distributed and Secure Cloud Clustering (DISC). May 26-28, 2010.
- [25] SW2015. Strategic Workshop Held at Villa Mondragone Via Frascati, Monte Porzio Catone, Napoli, Italy.” Seventeenth Strategic Workshop 2015. HBCs (HBC). May 18-20, 2015.
- [26] Taran, Yariv, Harry Boer, and Peter Lindgren. “Managing Risks in BMI Processes. Proceedings of the 11th International CINET Conference Practicing Innovation in Times of Discontinuity.” Zürich, Switzerland: Continuous Innovation Network, 2010.
- [27] Whinston, Andrew B., Dale O. Stahl, and Soon-Yong Choi. *The Economics of Electronic Commerce*. Indianapolis, IN: Macmillan Technical Pub, 1997.
- [28] Zott, C., R. Amit, and L. Massa. “The BM: Recent Developments and Future Research.” *Journal of Management* 37, no. 4 (July 1, 2011): 1019–42.



---

## BUSINESS MODEL INNOVATION COMPETENCES

### “What interdisciplinary competences can really value and do Business Model Innovation?”

---

Jane Flarup<sup>1</sup>  
Signe Stagstrup Jensen<sup>2</sup>  
Peter Lindgren<sup>3</sup>

<sup>3</sup> M-BIT Research Group Business and Social Sciences, AU Herning, Aarhus University, Denmark

#### Abstract

*Societies are lacking competences that are able to create, capture and deliver Business Models (BM) to present or new Business Model Eco Systems (BMES) (EU 2016 ). Many BM do not even reach BMES successfully and if this continues too many times businesses – both established and non-established (startups) businesses are in risk of falling behind, falling out or not even reach the BMES. Access to excellent Business Model Innovation (BMI) competences are therefore extremely important and more and more critical to businesses at all stages in their lifecycles. Valuable BMI competences can be found among university students and BMI competences can be increased in the favour and in the value of our businesses and society. The paper intends to commence the journey of building up knowledge, language and a screening framework on BMI Competence. On behalf of a literature study on BMI competences combined with many data from a quantitative and qualitative interdisciplinary research study of business engineering students at Aarhus University in Denmark a preliminary BMI competence profile related to BMI is carried out. Special focus has been on BMI projects in teams BMI Competences, characteristics and potentials.*

**Keywords:** Business Model Innovation Competence, Business model Innovation in teams, Business Model Innovation competence screening tools

#### 1. Introduction

In BM's we have earlier (Lindgren 2013), inspired by Prahalad and Hamel (Prahallad 1990), divided BMI competences in to four groups — technologies, human resource, organizational system and culture.

**Technologies** according to (Sanchez 1996, 2000, 2001) we divided into:

1. Product- and service-technologies
2. Production technology — both "Product- and Service-production technologies"
3. Process technology — process technologies that runs and steers the production technologies so that the product and service technologies can be created, captured, delivered, received and consumed

**Organizational system** is the system that the business models use to organize the use and “mixture” of technologies and human resource to carry out the Value Chain functions. The organizational system can also be so unique that the organizational system is a core competence.

**Culture** is the “soft” part of the competence dimension. We claim that any BM has a specific culture. The culture can be adapted one to one from the business or other BMs but can also be incremental even radically different to these. No need to explain the culture of a BM can also be a core competence to a business.

**Human Resources** are the people — either white collar or blue collar (Peters 1999) that the BM use or can use to carry out its Value Chain functions. The human resource, the mix and the use of human resource can also be so unique that human resource too is rendered as a core competence. Our focus in this paper is in particular the Human Ressource competence component.

## 2. Networkbased BMI – sharing BMI competences

In any BMI project we consider the competence dimension to be a mixture of technology, human resource, organizational system and culture but it is also about sharing competences (Porter 2011) with different BM’s competences inside and outside the business (Lindgren 2013, Lindgren 2016) - “pooling” their competences with ours. The sum of these competences forms the “shared competences” available in any business, BMI team to BMI projects like those we studied in this paper at the university of Aarhus.

BMI is and requires a multifaceted of competence – that are strongly related to the HR competence dimension or what we have called the BMI competence component. BMI competence related to the human competence component is about the ability - or competence - of the human to be open, willing and able to learn, create, capture, deliver, receive and consume knowledge at all tangible and intangible levels in the business and in the BMI project in focus (Teece 2016). It is about the ability to learn, create, capture, deliver, receive and consume knowledge with and through all tangible and intangible sensors of the human - with all possible human bonds (Prasad 2016) available. Technology competences can in this case support the human in this process. BMI competences and hereunder the human sensors can be supported with technology sensors - important BMI competences of all BMI projects that is expected to be even more important –maybe even a scarce competence in the future. Therefore there is a need to be able to find these BMI competences fast and to develop these if possible.

## 3. Design/Methodology/Approach

There is much knowledge about BM and BMI (Zott 2010, Teece 2011, Lindgren 2013). There are until now however not an accepted BMI competence language or framework developed.. There is little knowledge and research about those “BMI competences” that are so critical to our society – and valuable to our BMI projects. How do they really operate and work as value-adding mechanism – “objects” or “species” in the BMI process? Nor is the term BMI competence and BMI sensors defined and described in the BM and BMI Literature. The aim is to find a framework which can enable us to “screen” very early HR for BMI competences. Maybe further even in beforehand find the best, necessary and in the context required human BMI competences. This with the aim to improve BMI, do more BMI and maybe also even faster BMI? So

*How can BMI competences related to the Human being be defined and characterized?  
What are the BMI characteristics that businesses should look for?*

On behalf of a literature study on BMI competences and psychometric competence screenings systems together with a quantitative and qualitative study of six groups with business engineering students from Aarhus University was carried out. The six student groups each consisted of either two or three individuals. Out of fourteen students in all, ten were business engineering students while four were not. The data and competence profiles were taken up in the time period from 2010 – 2016. The Insights Discovery Evaluator (IDE) test was used as the psychometric framework and it gave us the opportunity to have a big set of data of personal competence profiles.

#### 4. BMI Competence screening

There is much knowledge and many tools for competence screening - but in our literature study we did not find screening systems particularly focusing on BMI competence. This can be due to there is still not an agreed upon business model language (Zott, 2010). We took our point of entry from the several psychometric test frameworks that exist. Psychometric tests seek to impose measurement upon operations of the mind. By combining different information with a specific personality characteristic, such tests can give answers to what characterizes a person – in our sense BMI HR competences has.

#### 5. Findings

Six businesses and/or innovative BM ideas which were chosen out of the total data material for their perspective of establishing and directing a business and related BM's are shown in table 1.

Business name	Business	Owners	Insight Profile colour	Status of the Business/idea
Grejfreak <a href="http://www.grejfreak.dk">www.grejfreak.dk</a>	Outdoor, hunting, military, fishing, camping equipment.	Kristian Casper	24 24	Established in 2008. Growth in turnover 8.3 mio., (2015), in squaremeters (900 m2) and people. Situated in Viby J., DK. 2 owners and 7 employees.
Benefittech (changed from Beerbot) <a href="http://www.benefittech.dk">www.benefittech.dk</a>	Digital serving of draft beer.	Jonas Jakob Mikkel	23 56 49	The original idea is closed down and substituted by a new idea which forms the base of a business: Benefittech. The idea is a digital solution for measuring personal results in a fitness center. The original founders (Jonas and Jakob) plus a newcomer work close together with an investor: LOOP fitness (no 2 in DK).
Dynello (Changed from Rew Strap) <a href="http://www.dynello.com">www.dynello.com</a>	A device for rewinding cargo straps.	Jens Emil	43 47	The product is changed into a similar product for other segments. The founder of Rew Strap (Jens) has included a newcomer (Emil) and changed the name to Dynello. A Start Up business in the auspices of Start Up Factory, Aarhus.
Cobraid <a href="http://www.cobraid.dk">www.cobraid.dk</a>	Digital solutions of a business's collection of big data.	Alexander Sune Morten	47 43 47	Growing. The founders started up the business before BE and developed the idea during the study. A Start Up business in the auspices of Start Up Factory, Aarhus.

Aarhus Cable Park No web. No Facebook profile.	Extreme water sport activity.	Tim Tobias	28 27	Helping Aarhus Festival week 2016 with ideas to extreme water sport events, and helps in the context of the Blue Rambla in Aarhus. Plan to structure the business as a non-profit organization with public subsidies in 2017 and from here to go for a commercial water sport business. Aarhus Cable Park is still in a fragile situation (mostly because of the cable park's need for a location by the waterside in Aarhus) if the founders don't get any sponsors or closer collaborators than they have at the moment.
Lyager Design <a href="http://www.lyagerdesign.dk">www.lyagerdesign.dk</a>	Interior design	Nikolaj Marie Louise (	24 52	The business was started by Marie Louise (architect student) and has been run as a side job to the studies for the two founders. The business is growing with a wide net of dealers in DK and some abroad.

Table 1: Six businesses and BMI projects created out of the student mass

Looking at the Insights Discovery profile position numbers we primarily look at the first and the second colour in each profile, because these two colours are representative for the dominating preference of a team member. They are informations about how best to collaborate with this team member. As can be seen in the profiles of the student projects and their businesses we found that they all possess the extraverted attitudinal colours of Fiery Red and Sunshine Yellow either as the first colour or the second colour in each individual's profile. The figure 2 shows the Conscious Persona Team Overview for all six BMI teams.

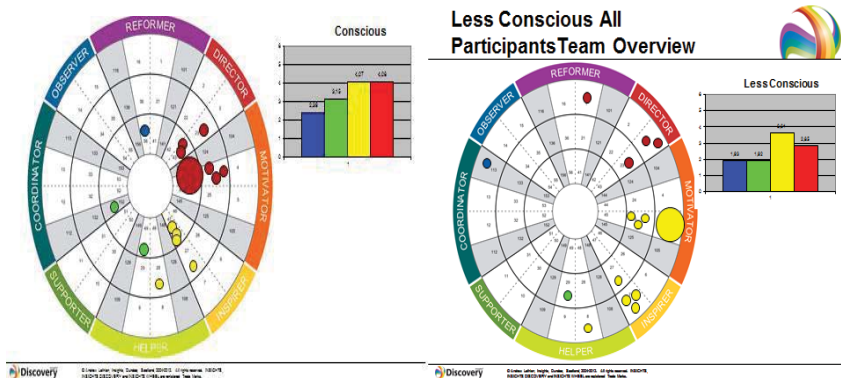


Figure: 2 Conscious All 6 BMI Team Overview *Fiery*  *Less*  *Conscious*  *All*  *MI*  *Or*  *Team*

Here it is demonstrated that almost all individual profiles are positioned on the Extravert hemisphere of the wheel. The big Fiery Red dot placed in the Directing Motivator position represents their combined energies – this is their average position, representing extraverted Intuition. The Conscious Persona represents: “Who you are”, “Who you ideally want to be” and “You, when you live up to other peoples’ expectations”. Looking at the Less Conscious Persona Team Overview the figure 3 shows for all six teams that almost all individual profiles – again - are positioned on the Extravert hemisphere of the wheel. The big Sunshine Yellow dot represents their average colour energies, once again extraverted Intuition. The less conscious persona represents “Your preferences when you react instinctively”, “Your behaviour when you are under pressure”, “You, when you are not conscious about the signals you send to others”.

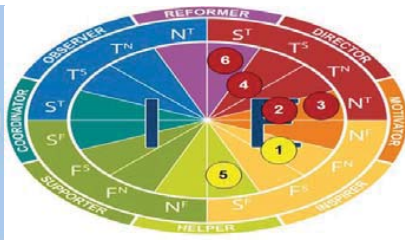
Remembering that the attitude means the way we react to outer and inner experiences opens for a hypothesis: That an innovative and entrepreneurial individual profile with BMI competences is extraverted. The introverted Cool Blue and Earth Green - we found - are not the dominant colours overall seen in any of our six student projects. Though in two businesses we found one blue and two green, but these three introverted colours have either red or yellow as their second colour. Secondly we found that none of the profiles are positioned in the Focused ring (the outer ring) which indicates that their colour expressions are multifaceted and not focused on one single basic way of communication. Eight of the 14 people were Adaptive (the inner ring) and six are Classical (the middle ring) individual types. In short – we found – that this leads to a tendency that all profiles were referring to more than their own colour (Classical types refer to two colours and Adaptive to three colours. According to (Insight 2006) more than 54 % of all individuals are Classical and 43 % are Adaptive. Only 3 % of all individuals are Focused.), which means that they have a preference for stretching behaviourally towards other colours. They have a natural tendency to adapt and connect. Looking at the Preference Flow (see figure 4) we see how much the six teams stretch on different colour energies (see explanations above about the Preference Flow). The Preference Flow varies within a continuum from -67% to +67% in general. In our teams we found that the combined energies of the participants results in a positive Preference Flow. They are mostly stretching on the Earth Green colour and Fiery Red colour, indicating that they want to make results, but in a way where they remain friends and in harmony. The combined Preference Flow of 32% of all six start-up businesses indicates that they are willing to stretch in all behavioral directions in order reach their goals. The 32% is significantly higher than average for a group of this size, indicating these participants show extra energy and that they are similar types.

As described above we also studied the four functions: Thinking, Feeling, Sensing and Intuition in the individuals and in their teams. The illustration below shows the individuals on a “wheel” that both describe the dominating function and the auxiliary function in the individuals. Where the dominating functions can be seen as the trunk of a tree, the auxiliary functions represent the branches. The dominant function is the most developed cognitive function within a person’s psychological dynamics, and the auxiliary function plays a supporting role, less dominant, but also visible.



**Figure 4: An illustration of the dominant and auxiliary functions in each individual and in each of the 6 businesses.**

**Note.:** 1) Cobraid; 2) Grejfrek; 3) Dynello; 4) Benefittech; 5) Aarhus Cable Park; 6) Lyager Design



**Figure 5: An illustration of the positions of each of the businesses compared to dominating and auxiliary function.**

We observe that 5 persons had a dominating Intuitive approach, 4 have a Feeling approach, 3 have a Thinking approach and 2 have a Sensing approach. Out of these 14 dominating functions 11 dominating functions are extraverted (Figure 5). Looking at the individual profiles and aggregating their individual colour scores within each of the businesses, we find that they are all extraverted (Figure 6). Furthermore 4 out of 6 businesses are placed in the Fiery Red quadrant. The extraverted attitude within each of the business dominates fully over the Introverted attitude.

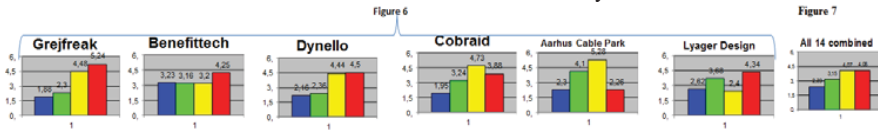


Figure 6: All the businesses in deeper detail regarding their functions and Figure 7: All of the individuals combined

When looking at the businesses in deeper detail we observe that in all the 6 businesses that the combination of Fiery Red and Sunshine Yellow dominates over the combination of Cool Blue and Earth Green. Extraverted Intuition is on average the dominating function. The combined energies of all the participants (Figure 7) confirm the observation extraverted Intuition (the combination of Fiery Red and Sunshine Yellow) is a dominant trait throughout the startup businesses. The position of Figure 7 is called the Adaptive Directing Motivator.

## 6. Discussion

The preliminary traits and indications shows that a “Helping Inspirer”, “a Motivating Director” and “a Directing Motivator” are the primary individual profiles and BMI competences to look for. This means that being Extraverted has to be present for almost all human beings in a BMI team. The colours red and yellow have to be very present either as first colour or second colour. These two colours give the qualities: Fiery Red – structure, goal, control, speed and results. Sunshine Yellow – creativity, innovative new ways, working together, visions, making a difference for other people. This facilitates the following vital BMI competence components:

BMI Competence	Description
<b>Speed</b>	Both colours are extraverted. Restless, with a high level of activity and energy, time is money (red), constantly looking for meaning (yellow). Both are constantly asking themselves: What is in it for me?
<b>Out of the box thinking creativity</b>	Especially the Sunshine Yellow will never work with copy-paste projects. They prefer walking on the thin ice, finding new ways, the unproven. The Fiery Reds don't respect authorities if it gives no sense – the Fiery Red make their own rules and procedures.
<b>Structure</b>	Sunshine Yellow detest structures, routines, details. They are unstructured and constantly innovative. Fiery Reds go for structure of time and organizing people. No waste of anything, especially not the Fiery Reds' life.
<b>Goal, results</b>	The Sunshine Yellow is ready to change everything if it makes sense. New visions, new projects. Results are less important, they constantly look for the concept and meaning with the project. Fiery Red fix a target and go for it – constantly looking for results and progression. My way is the high way. The Fiery Red can get rather brutal if the goals are not met.
<b>Courage</b>	Both colours are courageous – they don't like routines unless it gives sense to the projects. They are not afraid of change and the unknown.
<b>Action</b>	Lots of action. They are restless. If they feel waste of meaning, waste of time, they will disappear for you. The work hard if they feel that they are in the eye of the storm. If things get boring they find something else to do.

<b>Decision making</b>	Sunshine Yellows have lots of ideas fertilizing each other, so they can have problems in making decisions. There is always a new and a better idea. They live in a world of dreaming and fantastic ideas. Fiery Red can make decisions too fast – my target is the most important and I have no time to waste. They are not afraid of anything – decisions, arguments, discussions is a part of the game.
<b>Adaptive</b>	Three colours over the mid-line. Meaning that a person have a more open communicative approach – the person see the world from more angles and other people recognize three colours in his behaviour. Under pressure they are Focused (the outer ring) meaning that they react very much Fiery Red or Sunshine Yellow (the two colours we found in almost all profiles).
<b>Classical</b>	Two colours over the mid-line. Meaning that a person have a communication which is led by these two colours. The outer world will recognize these two colours. Under pressure a Classical profile stay in the middle ring meaning that he or she still show two colours. Under pressure a Classical will have a more open communication than a Focused profile (with only one colour over the mid-line).
<b>Communication</b>	Being Adaptive or Classical you show three or two colours to the world and you are recognized as such. You find it easier to understand and the communicative needs of others because these colours/needs are in your own persona. Your opposite colour is the colour you understand the least (this colour irritates you) but you will use all your colours to embrace it. Under pressure - as mentioned above – the Classical seems to have a more open communication as he/she keeps two colours above the mid-line whereas the Adaptive only shows one colour over the mid-line (Focused) thus demonstrating one colour (most often Fiery Red or Sunshine Yellow).
<b>Tolerance</b>	In all our businesses and BMI Projects we saw that they are aware of tolerance and of stretching themselves towards understanding of other people’s needs <i>and</i> working goals. In the analysis of the combined Preference Flow of all six we saw (see above and appendix 1) that they want to make results, but in a way where they remain friends and in harmony. We saw that they are willing to stretch in all behavioural directions in order reach their goals. The 32% of the Preference Flow is significantly higher than average for a group on this size, indicating these participants show extra energy and that they are similar types.
<b>Extraverted Intuition</b>	Intuition means that they interpret reality intuitively. They are not led by something sensed and registered in the outer world but they follow something from the inside. Sometimes their behavior may seem strange - but perhaps they listen to inner more or less spiritual guidelines.

**Table. 2 BMI Competence components and Descriptions**

### 7. Conclusion and acknowledgement

The aim of the article was to commence to do the very first findings of BMI competences that can act and do BMI. Our findings indicates that through interdisciplinary research and big data analysis it is possible to screen for BMI competences and potentials - even before the students enter the university or before members of a BMI group enters a BMI project. This can be of high value to businesses as well as to universities- those teaching and “forming” our students and BMI competences of the future. Society could benefit from this research as it is our hypothesis that some students have BMI competences that are not discovered in beforehand and maybe not treated and “nursed” as they could be with preference - to become and be developed to value able business model innovators. We acknowledge very much AU Interdisciplinary fond for supporting our interdisciplinary research at the MBIT LAB.

### **Referencer**

Benton, S., Schurink C.E & Desson S. (2008). An Overview of the Development, Validity and Reliability of the English version 3.0 of the Insights Discovery Evaluator. University of Westminster’s Business Psychology Centre. Brunceala, D’Estab & Salter, 2010.  
 Costa Paul T. and Robert R. McCrae (2010) NEO Inventories for the NEO Personality Inventory-3 (NEO-PI-3), NEO Five-Factor Inventory-3 (NEO-FFI-3), NEO Personality Inventory-Rvised (NEOPI-R). Copyright 2010 by PAR, 16204 N. Florida Ave., Lutz, FL 33549.  
 Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources.  
 Etzkowitz, Webster & Healey, 1998  
 EU 2016 - <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0795>  
 Flarup, J. & Wivel H. (2013). Designing Process Enablers to Strengthen Professional Skills in Project Work. Conference paper. MIT and Harvard University. See [www.edio.org](http://www.edio.org).

- Hogrefe (2014) Dansk udgave er oprindeligt oversat af Dorte H. Silver. Vejledningen er bearbejdet til ny udgave af Martin Stolpe Andersen/Hogrefe Psykologisk Forlag A/S. Dansk vejledning Psykologisk Forlag A/S 2014. 1 udgave. 1 oplag 2014.
- Insights (2006) Discovery Learning Library. Handbook for Insight Discovery Practitioners. www.insights.com.
- Jacobi J. (1973). *The Psychology of Jung*. Yale University Press.
- Lindgren, P. (2013). *The Business Model Cube*. Journal of Multi Business Model Innovation and Technology. River Publishing
- Lindgren, P. O.; H. Rasmussen and K. F. Saghaug (2013) Business models relations to Intellectual Capital - How to release Intellectual Capital from business BM's relations? IFKAD Conference Zagreb Croatia Conference Proceedings.
- Lindgren, P O. H. Rasmussen and K. F. Saghaug (2013) Business models relations to Intellectual Capital - How to release Intellectual Capital from business BM's relations? IFKAD Conference Zagreb Croatia Conference Proceedings
- Lindgren, Peter (2016) *The Business Model Ecosystem* Journal of Multi Business Model Innovation and Technology.
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI Manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Perkmann, M & Kathryn Walsh, 2007; University-industry relationships and open innovation: Towards a research agenda International Journal of Management Review 22 November 2007 Volume 9, Issue 4 December 2007 Pages 259-280
- Peters Tom (1999) *The Circles of Innovation* Knopf/Doubleday Publishing Group. Copyright.
- Porter, M (2012), *Shared Values* Harvard Business Review January 2011
- Prahalad, C. K. and Hamel, G. (1990). *The core competence of the corporation*. Harvard Business Review, May-June, 79-91.
- Prasad, Ramjee. *Human Bond Communication*. In: *Wireless Personal Communications*, Vol. 87, No. 3, 04.2016, p. 619-627.
- Quenk, N.L. (2000). *Essentials of Myers-Briggs Type Indicator assessment*. New York: Wiley.
- Santucci Silvia C, Heinz Amenitsch, Dan Cojoc, and Christian Riekel 2012). *Optical Tweezers for Touchless Sample Manipulation in Synchrotron Radiation Experiments*.
- Chapter Book (Ch. 6 pp. 183-213) in *Synchrotron Radiation and Structural Proteomics*, Eds. E. Pechkova and C. Riekel, Pan Stanford Publishing Pte. Ltd., Singapore, (2012), eBook ISBN: 9789814267939
- Siegel DS, D Waldman, A Link (2003) *Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study* Research policy 32 (1), 27-48
- Teece, D. J. (2011) *Business Models, Business Strategy and Innovation* Long Range Planning 43/2-3 April/May 2010
- Teece, D.J., Margaret Peteraf, Sohvi Leih (2016) *Dynamic Capabilities and Organizational Agility Risk, Uncertainty, and Strategy in the Innovation Economy* California Management Review, Vol. 58 No. 4, Summer 2016; (pp. 13-35) DOI: 10.1525/cmr.2016.58.4.13
- Zott, C.; Amit, R. & Massa, L. (2010). *The business model: Theoretical Roots, Recent Developments, and Future Research*. University of Navarra: IESE Business School.



# Biogas in Denmark and the Wireless Business Model - Discussions on Potential Value Upgrading

Morten Karnøe Søndergaard

Aarhus University, Department of Business Development and Technology, Multi Business Model Innovation and Technology, Birk Centerpark 15, 7400 Herning, Denmark. E-mail: [mks@btech.au.dk](mailto:mks@btech.au.dk)

**Abstract**— This paper investigates the status and focus of the biogas sector in Denmark, with a particular emphasis on wireless business models. A status on the current use of wireless sensors is presented, and a number of institutional barriers for further adoption are indicated. Furthermore, areas of potential value upgrading using wireless technologies are discussed focussed on what best fits and approaches might look like. Finally, suggestions made, are weighed against common institutional barriers and constraints for a potential best-case scenario.

**Keywords**—*biogas; business models; wireless technology; institutional barriers; integrated systems*

## I. INTRODUCTION

In recent decades technological developments in the Danish biogas sector has mainly been, focused on activities and experiments with biomass separation of nutrients and adaptations of biogas technology to varying local conditions. In the early stage of this development biogas plants were mainly based on low temperature systems - mesophilic biogas plants - with a process temperature of about 38 degrees. In recent years developments has tended to focus on high temperature - thermophilic plants - with an operating temperature of 52 degrees, which are slightly more gas-efficient [1].

The historical focus on especially biomass can be attained to the very large livestock population in Denmark, which has been a major element in Danish biogas plants using manure, unlike countries such as Germany, where energy crops especially corn has been the main biomass for biogas plants.

The Danish biogas model is based on waste and residues and seizes no feed - and food products to any great extent [2]. The profitability of biogas plants have been based on utilization for

power generation and utilization of the residue, which is warm, for either power plants or heating in the agricultural sector or industry.

The Danish biogas model remains very sustainable and competitive compared to prevalent biogas models in other countries. Biogas production means not only the production of environmentally friendly renewable energy, but also a beneficial treatment and recycling of animal manure and waste. At the same time, there are great advantages in the production, such as reducing greenhouse gas emissions, reduction of odours from manure, better utilization of nutrients in fertilizer, reduction of nutrient and sanitation and recycling of waste from industry and households. With recent developments - where biogas displaces fossil fuels- biogas has been given a different status, as a CO<sub>2</sub>-neutral fuel with a very favourable climate accounting [3].

The cost of producing energy in the form of biogas has historically been about twice as high as one kWh produced in a power plant, which basically had the consequence that biogas production needed to be subsidized to be competitive. Until 2012, the rate of contribution to the production of biogas was at a level, which did not in itself guarantee a significant development in the area. There has also been uncertainty in connection with the EU Commissions approval of the (Danish) Energy Agreement subsidies rates for biogas. These factors combined for a long time prevented new biogas-projects to have project-economics, finance and marketing agreements finalized [4].

Lack of standardization and low-key technology without trained staff characterized the biogas sector for many years. At the same time, the sector has experienced bottlenecks in

---

This paper is part of the Biogas2020 ([www.biogas2020.se](http://www.biogas2020.se)) project supported by EU Program, Interreg Öresund-Kattegat-Skagerrak ([www.interreg-oks.eu](http://www.interreg-oks.eu)). Interreg Öresund-Kattegat-Skagerrak provides support for joint Scandinavian projects developing society. Areas of support are innovation, low-carbon economy, transport and employment.

marketing and lengthy authorization procedures. In recent years, the sector has experienced rapid development as a result of political momentum and desire for expansion and a number of plants are being planned or erected currently, which can lead to a multi-fold increase in the biogas production.

It is a prerequisite for sound economics in the biogas plant that there is access to sufficient quantities of suitable biomass with a sufficient gas output at an affordable price. In relations to manure the challenge has been to ensure sufficient quantities within short transport distances, and high enough solids content, for it to be profitable to degas. In connection with the planning of biogas plants the content of dry matter have often been overestimated – i.e. the gas potential - and no practice of using other solids sources such as straw, has been incorporated. There are limited quantities, thus increasing competition and prices for waste-biomass, that can supplement manure.

Many plants have been dependent on the input of waste-biomass such as slaughterhouse waste and fat, and the reactors have not been designed for long residence times, which could ensure profitable turnover on other difficult biomasses. Energy crops are a solution only to a limited extent, as they are expensive because of high grain prices, and the permissible usage for biogas production will be limited in 2015 because of sustainability requirements. Other types of biomass, such as for instance straw and harvest from natural areas, have so far been used in relatively small amounts [5].

To overcome some of the current challenges in the biogas sector, i.e. access to suitable biomasses, gas efficiency and production planning the adoption of wireless technologies needs to be considered in the business model. This paper addresses the relevant areas and reflects on the potential value upgrading, which might be obtained.

## II. WIRELESS TECHNOLOGY IN BIOGAS

The use of wireless technology i.e. sensors in the Danish biogas sector is relatively well known. This especially goes for systems, which monitors gas-flows and process developments. However, the limitations of technical solutions and the work-effort related to sensor maintenance, such as calibration and cleaning, cause a less than optimum use and general poor acceptance. One survey, conducted in 2014, showed insufficiencies with sensors capacities in more than 400 farm-sized-plants in Germany [6]. In Denmark, no similar studies has been conducted, but all major gas plant producers on the

market, today offers some type of sensor or surveying system. Different input from industry agents suggests, that; malfunctions, lack of use or time-consuming calibrations, are equally commonplace in the Danish biogas sector [7].

The overall indication from both recent studies and industry agents is, that the commonplace restriction to primarily monitor the gas phase composition is likely insufficient, because disturbances in the process are discovered with a significant time delay. Despite of improvements it seems, that in current systems, changes of concentrations in off-gas stream is simply discovered too late. In order to overcome this shortcoming, in a recent study focused on a multiposition sensor technology it is suggested that focus is put on the liquid phase of the biogas process.

While, such an approach might be more efficient, as the study clearly indicates, several constraints have to be considered; the device has to withstand harsh conditions during digestion, the device has to be affordable, the device need to be manageable to the staff of the plant, and finally disturbances has to be discovered in zones, in which disturbances can be discovered as early as possible [8].

It is a well known issue from industrial sized plants that distribution of substrate and dissolved gases is highly heterogenic. A bulk of research efforts has been put into better the understanding of the gasification process and especially how to better gas-performance, while at the same time have a fairly standardized and “even” process throughout [9]. Despite a certain ability to control the range of biotechnological processes, the way biogas digestion is conducted varies strongly from industrial plants to farm-sized plants and from plant to plant. Therefore it is difficult to design a standard procedure that would “fit all”. Presently the application of stirring is a pivotal focus in industrial research concerning biogas due to the indications of economic benefits from this procedure [10].

However, in an overall perspective results from stirring are unevenly distributed from lager to smaller plants, which again indicate the difficulties in developing a uniform fit. In this perspective a wireless multi-sensor system makes good sense, as more components can be observed and evaluated and warnings can be issued without significant delays.

Results from the studies conducted concerning a wireless multi-sensor system clearly indicates, that such a system could be both cost-efficient and easy to install, thus meeting some of the pivotal requirements for implementation [11].

An important questions remaining is of-course whether lab-scale results will easily transfer into industrial practices. Using Denmark as the case and judging from previous developments, it seems fair to expect some institutional barriers. The development of the Danish biogas sector has been a very hands-on and knowledge-based process. Trial and error has been the leading paradigm and a significant gap exists between lab-design and daily life practices. Having a wireless tool at hand, which might contribute considerably to value upgrading, is not enough by itself. The tool also needs to show capabilities in practice and plant managers need to feel comfortable with its operation, otherwise there will be a lack in adoption or a sub-optimal use.

III. DISCUSSION ON VALUE UPGRADING

In the above section it has been indicated, that currently wireless sensors are not being put to their optimum use in the Danish biogas sector. A multisensory system with an emphasis on early stage detection of gas concentration changes might be a next step, when seeking to better the economic benefits of operations. Before any significant steps are made in this direction it is likely that more institutional barriers has to be reduced, especially concerning expectations on operational skills.

Monitoring the biogas process is far from being the only option, when seeking to adopt wireless technology in the optimization of the biogas sector. Adopting an international perspective the adoption of wireless technologies, especially sensors, is quite widespread in biogas. Apart from wireless sensors in the biogas process, the use of wireless sensors in landfill supervision is also quite commonplace. Wireless technology is also adopted in mixing biomass and wireless technology is offering live updates on plant status concerning needs for maintenance [12].

In many instances, what can be observed is that sensors are used in one or maximum two places in the process from crop (or manure) to final gas production. A generalised comparison between the different system approaches shows, that each system is very localized and fitted to specific needs, which again builds to the point, that developments in biogas are marked by local conditions, instability and is a very difficult task indeed.

Another challenge, when assessing the wireless system components being adopted today, is the lack of integration. In

a world, where it is agreed, that smart city ecosystems, backed by recent ICT technologies and services can create better, more cost efficient and sustainable environments, it stands out, that the biogas sector is lacking integration. There might still be many key-points to solve for a full wireless integration in biogas, but a vital point here is that it might be the sector itself halting developments, as the gap from lab to practice still needs bridging. Imagining a scenario, where this is achieved, there is essentially very few limitations to the level of control, which can be obtained, hence creating the stability needed both in production and yield.

Crops can potentially be designed and grown in labs, which will fit perfectly for the available manure and slurry mixtures. Wireless sensors will tell exactly when crops are ripe for a best fit of liquids, multisensory systems in plants will tell us if the processes are running smoothly and can tell us exactly, what yields will be. There are equal wins on both sides of such a system, as a further adaptation of wireless technology would also allow for us to have a better supervision of developments in the field, plus what biomasses are available “right now” (whether on tank or in the field) and what mixtures would be the best fix. In the other end of the system wireless technologies might also help create prognosis on where and when the gas is needed, i.e. for transportation, thus creating a best fit between production, price and customers.

While, a fully integrated wireless system supporting the biogas sector from end to other might still be a futuristic scenario, it is fairly simple to create an overview of what areas might be next step, when adopting wireless technology into the biogas sector (see Table 1)

TABLE I. AREAS OF NEXT POTENTIAL WIRELESS VALUE UPGRADING IN THE BIOGAS SECTOR

Type	Wireless ready		
	□ort	□ow	□e.T
Sensor	Tank and digester supervision	x	
Sensor/ Phone	Plant Maintenance	x	
Sensor	Biomass Mixing	x	
Sensor/ Phone	Crop-analysis		x
Sensor	Best “live” mix analysis		x
Sensor	Best fit; mix, demand, customer		x

Currently a number of “next phase” wireless options present themselves within the biogas section. It could be fully argued, that on going efforts still needs better integration before any next steps are tested, but evaluating the possibilities in Table 1 it could be stated that a natural next step would be pursuing a

best live mix analysis, encompassing surveys of tank capacities in a given local area (both fixed and on wheels) and fitting this knowledge into a smart system with on going biogas mixes, to see where the optimum results might be achieved.

Setting up a such a system, would require a fixed geographical setting and specific knowledge on transportation in the area (see Table 2)

TABLE II. PHASES AND COMPONENTS IN A WIRELESS "BEST MIX" BIOGAS SYSTEM

Phase	Components		
	ort	ye	ef
1	Tank (Stationary and wheels overview)	Sensor	GPS
2	Local Biogas Plants Online (ongoing production overview)	Sensor	CPU
3	Tank and traffic flows		GPS
4	Near-future crop and manure predictions	Sensors	
5	Early warning systems to secure best use of assets	Sensors	

More phases in a wireless best mix system will likely unfold, as the system gets up and running, but it seems likely, that establishing such a system could contribute to a valuable next step in biogas value upgrading.

IV. CONCLUSIONS

In recent years the biogas sector in Denmark has seen strong developments in relations to both political support and technological advances. In this process the use of wireless technology has been broadly adopted, but it still remains to see a real efficient use with an early stage discovery of changes in the concentration in the off-gas flow. A system for such an approach has been developed (at the lab level), but whether it will become widely adopted still remains to be seen. Notably a number of institutional barriers, especially concerning the lab and practice gap, will need to be reduced for this to happen.

Building on current use of wireless technology use in the Danish biogas sector, in the discussion section more possibilities for upgrading are pointed out. Most of these reflects on international developments and remain scattered. Therefore a suggestion is made, that in Denmark future developments might be focussed on developing integrated wireless systems, which "live-track" the full process from biomass production to end consumer. Further studies needs to be implemented to build such an approach, but preliminary

indications are, that this is where the value upgrading for next generation biogas is found.

ACKNOWLEDGMENT

Apart from the full list of references, throughout, this paper is based on knowledge obtained from working with several biogas agents in recent years. This effort has been divided between business efforts and research efforts and supplied a strong understanding of the need to bridge the gap between practice and research in this particular area.

Thanks to all agents and stakeholders for contributing valuable knowledge throughout mutual efforts and for contributing to the present paper. I equally acknowledge the valuable contributions, which has been brought to my attention via my colleagues with the Biogas2020 project.



REFERENCES

- [1] O. Rasmussen, P. Hesselbaek, M. K. Soendergaard, "Developments in the Danish Biogas sector and Current Status" Unpublished.
- [2] O. Rasmussen, "Comparing European Biogas Models with a special Focus on Sustainability" Unpublished.
- [3] M. K. Soendergaard, "The Danish Biogas Model – Presentation for the Nordic Investment Bank (NIB)" Unpublished.
- [4] Afdækning af muligheder for at fremme investeringer i biogas. Status i relation til muligheder og betingelser i forbindelse med finansiering af biogasanlæg (Energistyrelsen og Deloitte), 2013.
- [5] O. Rasmussen, "Biomass potential in Denmark, types and use", unpublished.
- [6] Weiland, P. Biogas-Messprogramm II: Fachagentur Nachwachsende Rohstoffe e.V.: Gülzow, Germany, 2009.
- [7] O. Rasmussen, "Preparatory notes for assessing the Danish Biogas Model", unpublished.
- [8] E. Kielhorn *et al.*, "Multiposition Sensor Technology and Lance-Based sampling for improved monitoring of the liquidphase in biogas processes", 2<sup>nd</sup> International Scientific Conference Biogas science (special issue), American Chemical Society, 2015.
- [9] P. Neuberger, S. Junne, "Scale-down simulators for metabolic analysis of large-scale bioprocessors" *Curr. Opin. Biotechnology*, 2010, 21, 114-121.
- [10] A. Lemmer, H-J. Naegle, "How efficient are agitators in biogas digesters? Determination of the efficiency of submersible motor mixers

- and incline agitators be measuring nutrient distribution in full-scale agricultural biogas digesters” *Energies*, 2013, 6, 6255-6273.
- [11] J. A. Álvarez, J.A. Otero, J. M. Lema, “A methodology for optimizing feed composition for anaerobic co-digestion of agro-industrial wastes. *Bioresour. Technology*. 2010, 101, 1153-1158.
- [12] B. Jun, Z. Shiping, Z. Hu, “Design and Research of Wireless Monitoring System for Solar Energy and Biogas Plant”, *Journal of Agricultural mechnization research*, 2014, 11.

# BUSINESS MODEL STRATEGY IN A WORLD OF PERSUASIVE TECHNOLOGIES

Troels Christian Korsgaard Andersen

BSS-BTECH  
Aarhus University  
Denmark  
Troelsck@btech.au.dk

When adopting ICT and/or data in the form of persuasive technologies into ones' business, it usually implies changes towards the business model, but also in terms of changing the business strategy. This is typically inherent in a business transformation process from AS-IS business model (product and service oriented business models) towards managing multiple TO-BE business models (data-driven business models). This article reviews the relation between strategy and business models when adopting ICT and/or data (persuasive technologies) the business, using research conducted in a European Electronic Manufacturing business, that are undergoing a transformation process from product and service-oriented business model, towards applying ICT as means to create, deliver and capture value through data into data-driven business models.

**Keywords**—  open strategy;  customer and/or user-oriented strategy;  data driven business models;

## I. INTRODUCTION

Recent studies within the field of business model, business model innovation and business strategy implies that the role of technology and its rate of advancement changes how we think and do business today (Chesbrough and Rosenbloom, 2002, Magretta, 2002, Markides and Charitou, 2004, Alexander et al., 2005, Johnson et al., 2005, Chesbrough, 2007, Mintzberg et al., 2008, Chesbrough, 2010, Doz and Kosonen, 2010, Teece, 2010, Zott et al., 2011, Lindgren and Rasmussen, 2013, Abraham, 2013). The role of business model as a concept and the process of business model innovation has increased significantly both in academia and in the world of businesses (Johnson et al., 2008, Chesbrough, 2010, Casadesus-Masanell and Ricart, 2011, Zott et al., 2011, Lindgren and Rasmussen, 2013). The use of technologies has never been greater than it is today, more and more are depended on their technology devices in their daily activities, which it turns out creates new business opportunities (Ljubomir and Marina, 2008, Lindgren and Abdullah, 2013) e.g. smartphones, iPads, computers, applications and user interfaces etc. The advancement of information and communication technologies (ICT) (Casadesus-Masanell and Ricart, 2010, Fleet, 2012) has created opportunities for small start-up businesses, entrepreneurs and alike to disrupt whole ecosystems e.g. Netflix, Uber or Tesla Motors etc. These disruptions have changed the competitive positioning within their respective

ecosystems in such a way that the 'old and mature' businesses struggle to survive and compete. Why? Because their business strategy and business models are not structured to be agile or even able to change their core business towards offering the same values as their disrupting competitors. As mentioned in the previously articles (Andersen and Bjerrum, 2016, Bjerrum et al., 2016), manufacturing businesses are struggling to compete and faces commoditization against development countries.

These mature businesses had experienced growth over a period of time, resulting in expansion of their organization, with some on investing heavily in tangible infrastructure such as buildings, production technologies or increased staff etc. This development has caused many manufacturing businesses to become less mobile (just imagine the difference between a tankship shifting its course versus a speedboat) and in facing the effect of commoditization, they downsize and cut costs to avoid bankruptcy. Many of these manufacturing businesses have adopted and used classic business strategy that guides these business through their organizational growth and how to compete within their respective 'industries' (Oliva and Kallenberg, 2003, Brax, 2005, Lindberg and Nordin, 2008, Vargo and Lusch, 2008). Unfortunately, there were numerous problems that could be identified as the cause of these manufacturing businesses challenges and ultimately downfall. Three of which is first the lack of understanding on the advancement of ICT and/or other trends occurring in the environment. Second, the lack of a better strategy that fits with the behavioral patterns in the environment (Markides and Charitou, 2004, Chesbrough and Appleyard, 2007, Casadesus-Masanell and Ricart, 2010, Doz and Kosonen, 2010, Tan et al., 2010, Teece, 2010), and third the lack of understanding and integrating the customer and/or user' in the centre of the innovation process occurring within the business. The latter will be the focus of this article.

This article builds upon the foundation of the previously articles (Andersen and Bjerrum, 2016, Bjerrum et al., 2016) with a focus on addressing a customer and/or user driven strategy when managing the service and data driven multi business model platform (SDBM-platform) in the attempt to create data-driven business models.

II. METHODOLOGY

This conceptual paper is the third link extension from previous research on the subject of Servitization on the transformation process from a product-oriented logic towards service dominant logic, through ICT and the creation of the SDBM-platform (Andersen and Bjerrum, 2016) with a focus on persuasive technologies. These studies were conducted as participatory action research on a single in-depth case study (European Electronic Manufacturing company - EEM) at a managerial level.

This paper builds on the qualitative data collected at the EEM and the findings from the previous article, hence the SDBM platform.

A case study design has been chosen to conduct in-depth qualitative interviews with 14 selected respondents on managerial level. It facilitates a holistic understanding of an complex environment and phenomena of strategy and business model innovation that is not necessarily easy to separate from its contextual setting (Eisenhardt, 1989). Thus allowing the author to build new theory and/or extend existing theoretical contributions.

Furthermore, the participatory action research approach allowed the author to test the suggested framework in conducted workshops throughout a 6 months' transformation process from a product and service-oriented logic towards incorporating data-driven business models.

III. THE SERVICE AND DATA DRIVEN MULTI BUSINESS MODEL PLATFORM

The previous article concluded that in order to structure the transformation process from a product dominant-oriented logic towards a service dominant-oriented logic, a platform was needed. In addition, the creation of business models and managing multiple business models also indicated the need of a platform that represents the business dimensions and how value could be created, delivered and captured using data as the main source to create data-driven business models. This platform was created on the foundation of the business model cube framework (Lindgren and Rasmussen, 2013) which suggest that any business consists of seven dimensions. Fig. 1 - Some of these were combined into sub-platforms; The value proposition (the product and service system), creating value (customer and/or users'), using resources (value chain functions and competences), network, value formula and relations as seen below in figure 1. Overall, the role of ICT in this platform is to systemize and structure data as means to produce data-driven business models that incorporate how users interact with the ICT (user data) and process this through the different sub-platforms into value that fits with the customers and/or users' behavioral patterns and needs.

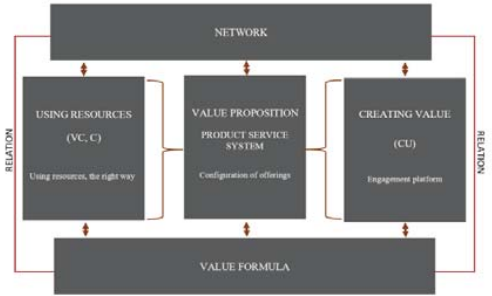


Fig. 1 The service and data-driven multi business model platform (Andersen and Bjerrum, 2016)

IV. STRATEGY – THE SERVICE AND DATA DRIVEN MULTI BUSINESS MODEL PLATFORM

Recent contributions within the research field of business models advocates for defining the relation between business strategy and business models (Chesbrough and Appleyard, 2007, Casadesus-Masanell and Ricart, 2010, Teece, 2010). (Casadesus-Masanell and Ricart, 2010) argues that there exists a differentiation between strategy and business models, hence a business model is a reflection of the realised strategy in which the business model is conceived as (a) *the concrete choices made by management about how the organization must operate, and (b) the consequences of these choices* – (Casadesus-Masanell and Ricart, 2010): p. 198. The article contributes with the two-stage competitive process framework: the strategy stage – *The firm chooses the business model through which it intends to compete*, and the tactics stage – *Tactical choices made amongst those available, depending on business model choice at first stage* – (Casadesus-Masanell and Ricart, 2010): p. 196. The purpose with this framework is to make a distinguish between the three concepts: **Business models** as the logic of a business, how it operates and how it creates value for stakeholders, **Strategy** as the choice of business model and how to compete in the marketplace, and finally **Tactics** as choices that are available for the business, after choosing the business model (crucial for determining a business' value creation and capture).

Another perspective on the relation between strategy and business models comes from Henry Chesbrough. In the article (Chesbrough and Appleyard, 2007), Chesbrough argues for the need to look differently at both business models and business strategy by opening up the business' "boarders" to the degree necessary in their respective situations. In that, the author also places great empathy on the need to receive values from the business network actors such as network partners, stakeholder and customers, because otherwise they might shield themselves from important inputs that would contribute

to a greater innovation and business model innovation process. This also includes the necessity to monitor the environment for emerging trends such as technological advancement e.g. ICT that might change the competing conditions within their ecosystems.

The open strategy approach advocates for knowledge creation through open invention, meaning that the business model should make use of the existing knowledge that lies within its network (capture value). Network partners and customers might have unique and tacit knowledge that could contribute significantly to the business model e.g. Linux using its innovation community to create the OS. Furthermore, the need to use 'open coordination' is seen as a great contributor to the innovation process adding more value to the business by utilizing the ecosystem to create joint ventures with other businesses e.g. IBM on opening the personal computer architecture to the ecosystem spawning new businesses such as Compaq and set the standard for the PC industry, through cooperation between Microsoft (OS) and Intel's (microprocessor technology). The article concludes that open strategy can balance the value creation forces found in creative individuals, innovation communities or collaborative initiatives with the need to capture value to sustain continued innovation with other parties, see figure 2 below.

Fig. 2 - The below figure illustrates the relation between strategy, the SDBM-platform, data-driven business models and ICT. It is divided into two stages. The strategy stage includes the SDBM-platform and possible data-driven business models to adopt as the outcome of that platform, hence strategy as the contingent plan of action as to what business model to use. In particular, both the platform and the data-driven business models is the result of open innovation and open strategy as they are user-driven by including the customers and/or users to take part in the innovation process of creating and delivering values through ICT e.g. e-business, products or services. The tactics stage involves choices that become available from the previous stage. These will

determine how the individual data-driven business model operate (compete) within the ecosystem with the purpose of capturing value from the customers and/or users. The arrows illustrate the exchange of value e.g. data, knowledge and information between the different stages.

V. CONCLUDING DISCUSSION

The previously article raised a question regarding how the business should adopt the SDBM platform and by that how it would fit with their business strategy. As mentioned above, the traditional ways of doing strategy does not comply with how business models are to be developed today, using ICT and customers and/or user as a means to create, deliver and capture values instead of merely focusing on manufacturing and transactional exchange of products. More importantly, the SDBM platform does not take into account that the EEM would need to open up their business or business segments to some degree in order for the platform to function properly, because it depends on inputs (knowledge, information and data) from the business environment (ecosystem), through ICT that leads to creating data-driven business models. This means that the EEM on their transformational journey needs to make some strategic changes on their organization such as resource allocation into the network (opening the boarders) regarding building an ICT system that can process the large amount of data created from the users of their products. They would also need to create a customer and/or user driven strategy that implies the customer as center of their innovation process including the making of tactical decisions that implies to that strategy by attracting customers and/or users to take actively part in innovating the desired business models.

This advocates for another strategy, open strategy.

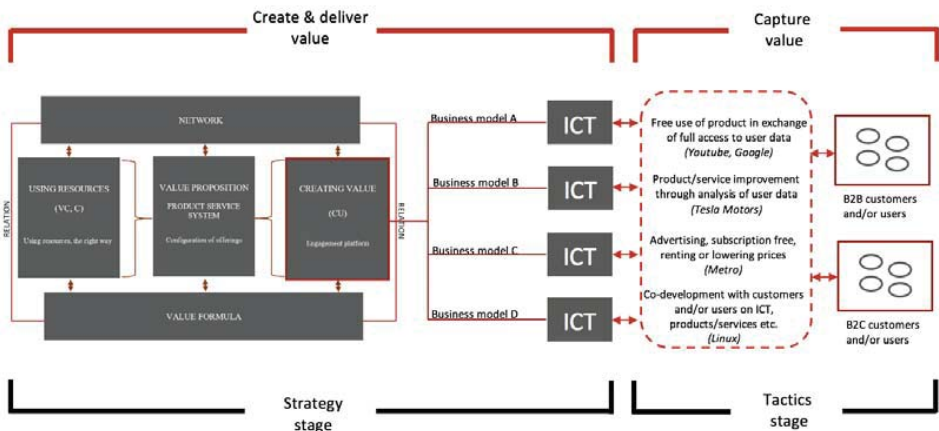


Fig. 2 Modified Strategy business models and tactics framework: adapted from (Casadesus-Masanell and Ricart, 2010)



REFERENCES

- [1] ABRAHAM, S. 2013. Will business model innovation replace strategic analysis? *Strategy & Leadership*, 41, 31-38.
- [2] ALEXANDER, O., YVES, P. & CHRISTOPHE, L., TUCCI 2005. Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, 16, 1.
- [3] ANDERSEN, T. C. & BJERRUM, T. C. B. 2016. Service and Data Driven Multi Business Model Platform in a World of Persuasive Technologies. *Journal of Multi Business Model Innovation and Technology*, 4, 47-60.
- [4] BJERRUM, T. C. B., ANDERSEN, T. C. & AAGAARD, A. 2016. Facilitating Data Driven Business Model Innovation - A Case study.
- [5] BRAX, S. 2005. A manufacturer becoming service provider - challenges and a paradox. *Managing Service Quality*, 15, 142-155.
- [6] CASADESUS-MASANELL, R. & RICART, J. E. 2010. From Strategy to Business Models and onto Tactics. *Long Range Planning*, 43, 195-215.
- [7] CASADESUS-MASANELL, R. & RICART, J. E. 2011. How to design a winning business model. Harvard Business School Publ. Corp./Harvard Business Review.
- [8] CHESBROUGH, H. 2007. Business model innovation: it's not just about technology anymore. *Strategy & Leadership*, 35, 12.
- [9] CHESBROUGH, H. 2010. Business Model Innovation: Opportunities and Barriers. *Long Range Planning*, 43, 354-363.
- [10] CHESBROUGH, H. & ROSENBLUM, R. S. 2002. The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11, 529-555.
- [11] CHESBROUGH, H. W. & APPELYARD, M. M. 2007. Open innovation and strategy. *California management review*, 50, 57.
- [12] DOZ, Y. L. & KOSONEN, M. 2010. Embedding Strategic Agility. *Long Range Planning*, 43, 370-382.
- [13] EISENHARDT, K. M. 1989. Building theories from case study research. *Academy of management review*, 14, 532-550.
- [14] FLEET, G. J. 2012. EVIDENCE FOR STALLED ICT ADOPTION AND THE FACILITATOR ECOMMERCE ADOPTION MODEL IN SMEs. *International Journal of the Academic Business World*, 6, 7-18.
- [15] JOHNSON, G., SCHOLES, K. & WHITTINGTON, R. 2005. *Exploring corporate strategy*. Harlow, Prentice Hall/Financial Times.
- [16] JOHNSON, M. W., CHRISTENSEN, C. M. & KAGERMANN, H. 2008. Reinventing Your Business Model. (cover story). *Harvard Business Review*, 86, 50-59.
- [17] LINDBERG, N. & NORDIN, F. 2008. From products to services and back again: Towards a new service procurement logic. *Industrial Marketing Management*, 37, 292-300.
- [18] LINDGREN, P. & ABDULLAH, A. M. 2013. Conceptualizing strategic business model innovation leadership for business survival and business model innovation excellence. *Journal of Multi Business Model Innovation and Technology*, 1, 115-134.
- [19] LINDGREN, P. & RASMUSSEN, O. H. 2013. The business model cube. *Journal of Multi Business Model Innovation and Technology*, 135-182.
- [20] LJUBOMIR, D. & MARINA, M. 2008. E-BUSINESS AND GLOBALIZATION. University of Zagreb, Faculty of Economics and Business.
- [21] MAGRETTA, J. 2002. Why Business Models Matter. *Harvard Business Review*, 80, 86-92.
- [22] MARKIDES, C. & CHARITOU, C. D. 2004. Competing with dual business models: A contingency approach. *Academy of Management Executive*, 18, 22-36.
- [23] MINTZBERG, H., AHLSTRAND, B. & LAMPEL, J. 2008. *Strategy safari : a guided tour through the wilds of strategic management*, Hemel Hempstead, Prentice Hall.
- [24] OLIVA, R. & KALLENBERG, R. 2003. Managing the transition from products to services. *International Journal of Service Industry Management*, 14, 160-172.
- [25] TAN, A. R., MATZEN, D., MCALOONE, T. C. & EVANS, S. 2010. Strategies for designing and developing services for manufacturing firms. *CIRP Journal of Manufacturing Science and Technology*, 3, 90-97.
- [26] TEECE, D. J. 2010. Business Models, Business Strategy and Innovation. *Long Range Planning*, 43, 172-194.
- [27] VARGO, S. L. & LUSCH, R. F. 2008. From goods to service(s): Divergences and convergences of logics. *Industrial Marketing Management*, 37, 254-259.
- [28] ZOTT, C., AMIT, R. & MASSA, L. 2011. The Business Model: Recent Developments and Future Research. *Journal of Management*, 37, 1019-1042.
- [29] ABRAHAM, S. 2013. Will business model innovation replace strategic analysis? *Strategy & Leadership*, 41, 31-38.
- [30] ALEXANDER, O., YVES, P. & CHRISTOPHE, L., TUCCI 2005. Clarifying Business Models: Origins, Present, and Future of the Concept. *Communications of the Association for Information Systems*, 16, 1.
- [31] ANDERSEN, T. C. & BJERRUM, T. C. B. 2016. Service and Data Driven Multi Business Model Platform in a World of Persuasive Technologies. *Journal of Multi Business Model Innovation and Technology*, 4, 47-60.
- [32] BJERRUM, T. C. B., ANDERSEN, T. C. & AAGAARD, A. 2016. Facilitating Data Driven Business Model Innovation - A Case study.
- [33] BRAX, S. 2005. A manufacturer becoming service provider - challenges and a paradox. *Managing Service Quality*, 15, 142-155.
- [34] CASADESUS-MASANELL, R. & RICART, J. E. 2010. From Strategy to Business Models and onto Tactics. *Long Range Planning*, 43, 195-215.
- [35] CASADESUS-MASANELL, R. & RICART, J. E. 2011. How to design a winning business model. Harvard Business School Publ. Corp./Harvard Business Review.
- [36] CHESBROUGH, H. 2007. Business model innovation: it's not just about technology anymore. *Strategy & Leadership*, 35, 12.
- [37] CHESBROUGH, H. 2010. Business Model Innovation: Opportunities and Barriers. *Long Range Planning*, 43, 354-363.
- [38] CHESBROUGH, H. & ROSENBLUM, R. S. 2002. The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11, 529-555.
- [39] CHESBROUGH, H. W. & APPELYARD, M. M. 2007. Open innovation and strategy. *California management review*, 50, 57.
- [40] DOZ, Y. L. & KOSONEN, M. 2010. Embedding Strategic Agility. *Long Range Planning*, 43, 370-382.
- [41] EISENHARDT, K. M. 1989. Building theories from case study research. *Academy of management review*, 14, 532-550.
- [42] FLEET, G. J. 2012. EVIDENCE FOR STALLED ICT ADOPTION AND THE FACILITATOR ECOMMERCE ADOPTION MODEL IN SMEs. *International Journal of the Academic Business World*, 6, 7-18.
- [43] JOHNSON, G., SCHOLES, K. & WHITTINGTON, R. 2005. *Exploring corporate strategy*, Harlow, Prentice Hall/Financial Times.
- [44] JOHNSON, M. W., CHRISTENSEN, C. M. & KAGERMANN, H. 2008. Reinventing Your Business Model. *Harvard Business Review*, 86, 50-59.
- [45] LINDBERG, N. & NORDIN, F. 2008. From products to services and back again: Towards a new service procurement logic. *Industrial Marketing Management*, 37, 292-300.
- [46] LINDGREN, P. & ABDULLAH, A. M. 2013. Conceptualizing strategic business model innovation leadership for business survival and business model innovation excellence. *Journal of Multi Business Model Innovation and Technology*, 1, 115-134.
- [47] LINDGREN, P. & RASMUSSEN, O. H. 2013. The business model cube. *Journal of Multi Business Model Innovation and Technology*, 135-182.

- [48] LJUBOMIR, D. & MARINA, M. 2008. E-BUSINESS AND GLOBALIZATION. University of Zagreb, Faculty of Economics and Business.
- [49] MAGRETTA, J. 2002. Why Business Models Matter. *Harvard Business Review*, 80, 86-92.
- [50] MARKIDES, C. & CHARITOU, C. D. 2004. Competing with dual business models: A contingency approach. *Academy of Management Executive*, 18, 22-36.
- [51] MINTZBERG, H., AHLSTRAND, B. & LAMPEL, J. 2008. *Strategy safari : a guided tour through the wilds of strategic management*, Hemel Hempstead, Prentice Hall.
- [52] OLIVA, R. & KALLENBERG, R. 2003. Managing the transition from products to services. *International Journal of Service Industry Management*, 14, 160-172.
- [53] TAN, A. R., MATZEN, D., MCALOONE, T. C. & EVANS, S. 2010. Strategies for designing and developing services for manufacturing firms. *CIRP Journal of Manufacturing Science and Technology*, 3, 90-97.
- [54] TEECE, D. J. 2010. Business Models, Business Strategy and Innovation. *Long Range Planning*, 43, 172-194.
- [55] VARGO, S. L. & LUSCH, R. F. 2008. From goods to service(s): Divergences and convergences of logics. *Industrial Marketing Management*, 37, 254-259.
- [56] ZOTT, C., AMIT, R. & MASSA, L. 2011. The Business Model: Recent Developments and Future Research. *Journal of Management*, 37, 1019-1042.

## Applying Persuasive Design for Reducing Urban Unrest

Kristian Løbner and Morten Karnøe Søndergaard

Department of Business Development and Technology, Aarhus University  
 Research Group for Multi Business Model Innovation and Technology (MBIT)  
 Herning, Denmark

krlb@btech.au.dk ; mks@btech.au.dk

**Abstract**— Urbanization has been on the rise for decades. In recent years, the urbanizations is further fueled by the increase in immigration to Western societies. Post-war planning of social housing areas in the outskirts of Western cities – and the following lack of maintenance – has led to socially derived areas with diverse, fractionated and unstable social contracts, and urban unrest is lurking. Applying wireless technologies and persuasive design for coping with potential urban unrest is possible, effective and doable in most Western societies. Increasing motivation, securing of abilities and application of triggers are key for changing people's behaviors from negative attitudes to positive engagement. This paper discovers a practical example of the application of wireless technologies and persuasive design for positive behavior changes in a social housing area in Aalborg, Denmark.

**Keywords**— *application of wireless sensors, persuasive technology, behavior change, social housing, urban unrest*

### I. INTRODUCTION

Throughout centuries, the world has experienced urbanization with more people moving from rural areas to cities. Since World War II, the urbanization has increased and put cities under pressure with increasing demands for space and resources. According to the United Nations World Urbanization Prospects [1], 30 % of the world's population was urban in 1950 and in 2014 the number was increased to 54 per cent – and by 2050, 66 per cent of the world's population is projected to be urban. Continuous urban growth are projected to add 2.5 billion people to the world's urban population by 2050. As the world continues to urbanize, social, economic and environmental challenges will be increasingly concentrated in cities. The juxtaposition of people often leads to innovation and economic growth as cultures interact. But, it can also lead to tensions and conflict, especially if cultural groups retreat into particular areas of the city [2].

Many western cities has experienced the following development: The increasing demand for space in the cities also led to increased value of urban property. Hence, urban renewal of former industrial inner city areas such as factories, docks or military areas transformed the inner city into dense, value-adding purposes as well as residential areas. The slum clearance of inner city areas paved the way for development of new residential areas on the brink of the city. As the passion for concrete, social housing and 'streets in the air' was eminent during the 1950 to the 1980's, the modern, dense and suburban apartments were praised as framing the new lives of the

middle-class. However, a gradual decline of the ideal of equality eroded the impression of social housing. In many Western countries, a neo-liberalism emerged and the former appraisal of 1960's architecture entered a trough of despair. Lack of connections to the outside world, lack of maintenance, building failures, decreasing rents, increasing immigration and lack of political awareness turned many social housing projects into ghettos of suppressed underclass with high crime rates, unemployment, health issues and immigrations problems [3]. Recent research suggests that conflicts are increasingly likely to occur in densely populated, electronically networked settings at the periphery of sprawling coastal cities in the developing world [4]. As experienced, tensions and the urban unrest sometimes boil over into riots as the disorders in Los Angeles in 1992, in various English cities in 2001 and 2011, the French Riots in 2005 and in Ferguson and Baltimore in 2014-2015.

In Scandinavia, there are also plenty of examples on the abovementioned urban development, and, to some degree, on urban unrest. However, as Denmark, Norway, Sweden, Finland and Iceland all are in top 20 of the most peaceful countries in the world [5], the Nordic Countries can be perceived as one of the world's most peaceful regions. Furthermore, the Scandinavian countries are developed to comprise massive amounts of detailed data on economics, health, education, insurance, traffic etc. If it is possible to change the derived social housing areas into prosperous urban development – and measure the effects – through persuasive peace innovation technology, it might very well be possible to use the tests and cases from Scandinavia in other Western countries as well. However, the problems, people, technological levels, quality of networks etc. should be taken into consideration in terms of transferability of the findings between socio-cultural contexts.

Peace technology is fundamentally a mediating technology, and acts as an intervening agent, augmenting our ability to engage positively with others [6]. Peace technology consists of sensors, communications, computation and actuators, and unlike previous technological revolutions, individuals can now design and deploy peace technology at scale almost anywhere in the world [6] – also in derived social housing areas.

In order to study how peace technology can be utilized innovatively in social housing areas, a case study has been conducted on the Aalborg East area. With app. 15,000 inhabitants in an isolated community with more social issues than other areas of the city of Aalborg, there is substantial potential for implementing peace technologies for coping with

urban unrest. The study was conducted in co-operation with Stanford Peace Innovation Lab and its Danish affiliated partners, Himmerland Housing Association, COWI and more than 100 local experts.

This paper is divided into four parts. The first part introduces the challenges with urban unrest in social housing areas and the themes of persuasive technology and a specific case. After this introductory part, the second chapter assesses how persuasive technology can cope with urban unrest. In chapter III, the case of Aalborg East is explained with special regard to fast adaptation to persuasive technology. In chapter IV the validity of the study is assessed with comparison to the specific work with persuasive technology in other social housing areas. Chapter V discusses and concludes of the validity of the study and proposes further studies for closing the gaps on application of peace technology in social housing areas.

## II. PEACE TECHNOLOGY

For the first time in human history we are now able to measure and record inter-personal interactions at a very high resolution and in real time – and many governments agencies, private companies and ordinary people are doing this to an increasingly extent. Through social software and mobile devices we are passively recording more social behavior every day. With the possibilities of measuring social behavior, we are also capable of shaping it into desired behaviors. Research has shown that the greatest enabler for sustainable behavior change is "positive behavior" rather than leading to "negative behavior". For this, and for ethical reasons, this article focuses on designing for pro-social behaviors.

Peace technology consists of a variety of Information and Communication Technologies (ICT) such as **sensors** that can measure human behavior, (cameras, GPS, microphones etc.), **communications technology** (wireless technologies, smart phones, tablets etc.) and **computation** (local or cloud-based). Add to the three components a fourth addition: **actuators**, which allows triggering and coordinating a desired behavior.

ICT plays a key role in improving communication, facilitating negotiations, increasing transparency, and building trust [6]. Due to its relatively recent development, ICT for peace-creating purposes does not yet encompass a clear analytical framework [6]. Some frameworks are cyclical and refers to conflicts as phases, which builds upon each other. Other frameworks focus on how ICT is used before, during and after a conflict. Historically, these phases of a conflict are seen as a chronological sequence. More recently, it is acknowledged that conflicts do not follow a linear path and a more integrated approach is needed. When it comes to conflicts in social housing areas it is clear that conflicts are much more intertwined and complex than separate phases.

However, in order to find viable peace technology solutions to urban unrest in social housing areas, it is proposed to merge a framework for behavior design into peace technology. B.J. Fogg, an American psychologist, defines persuasive technology as any interactive computing system designed for changing people's attitudes and/or behaviors [7]. According to Fogg, behavior change is caused when something is easy to do

to, the motivation is high and the trigger for behavior change is timely [8]. For further elaboration on the behavior model for persuasive design, see [7] and [8].

In social housing areas, designing for peace and coping with urban unrest is attached to two main purposes: either it is about policy and diplomacy or it is about philanthropy and charity. There exist three problems for developing derived social housing areas. The *first* problem is related to the scale of resources to the problem. In derived social housing areas, there exist a vast amount of resources allocated to the area through dozens of stakeholders such as the housing association, police, education, health system, insurance, banking etc. Many nationalities, languages, ages, income levels, religions etc. tends to disturb the picture of a uniform demography in social housing areas. The question is where to focus the efforts to reach the highest impact on peace.

The *second* problem related to allocating the resources effectively. With the great number of stakeholders and possible target groups, it can be hard to bridge resources effectively to create sustainable positive behavior. As encountered in many social housing areas, allocation of resources does not itself solve any problems. A specific target group, clear value propositions and measurement of effects are of great need – otherwise, scarce resources are spilt on politics or philanthropy.

The *third* problem is the absence of price signals. If a stakeholder invests money on peace innovation in social housing areas he should be able to gain profits within a relatively short period of time. If not, the investments can be seen as charity – which in itself does not solve the problems on the long range. Public or private investment to derived social housing areas are in great need, but the greatest impact is sustained if the resources are put into play in the local communities, by local citizens and for local purposes. The result of the three problems is sometimes perverse investments; building of large mosques (which without proper supervision turn into unsafe environments for radicalization); investments in ICT in schools (which without proper security are being looted); investments in gardening and arts (which without local ownership are being vandalized). A suggested solution is to create a profit-based peace technology platform that inspires stakeholders and citizens to positive pro-social behaviors.

As opposite to the challenges, cities and social housing areas comprise special possibilities and challenges for creating positive behaviors. *Firstly*, they enable people to grow social webs that encourage people to learn, specialize, and depend on each other in new and deeper ways [9]. On the other hand, the diversities of people, believes, social norms and actions can lead to instability and insecurity. *Secondly*, cities and social housing areas also increases the extent, quality and usage of virtual and physical infrastructure. On the other hand, such networks, especially physically infrastructure, can lead to segregated and divided urban areas. The virtual networks tend to encompass more peace-increasing possibilities – for example, see the Israel-Loves-Iran campaign, initiated by Stanford Peace Innovation Lab. Cities and social housing areas provide the physical networks that enable people to interact but the construction of actual social networks rely on the people themselves. *Thirdly*, the fittings and tensions of these networks,

and the tradeoffs among them, affect how prosperous a city or a social housing area is. That is, whether it fractures or fails, and whether people stay or move out. *Fourthly*, it is possible for planners and policy makers to craft urban constructs and policies that create positive social interactions at low cost. It is also possible to enable technologies that promote the ability of people to connect, trust, and engage to mitigate connection obstacles such as crime or segregation.

Based on Stanford Peace Innovation Lab's research [10] on peace innovation, the following methodology can be applied for peace innovation in social housing areas.

- 1) *Choose the target communities*
- 2) *Observe the technologies they use*
- 3) *Pick a positive engagement*
- 4) *Create fast prototype interventions*
- 5) *Measure impact*
- 6) *Optimize & repeat from (1)*

### III. DEVELOPMENT OF PEACE TECHNOLOGY IN AALBORG EAST

This chapter exemplifies how peace technologies can be applied in cities and social housing areas. This chapter dives into the setup, the process and the outcome.

#### A. The Setup: Aalborg East

Aalborg is the fourth largest city in Denmark with app. 135,000 inhabitants. In the 1960's, based on functionalistic approaches to urban planning, a new neighborhood was developed 10 km east of the city center. The functional design created effective infrastructure, green spaces and modern apartments for almost 15,000 inhabitants. However, the planning designs enclosed Aalborg East within infrastructures with few possibilities to enter and exit the area and divided the suburb sharply into functionalistic purposes. Today, more than 50 years later, the result is that residents of Aalborg East have lower incomes, lower education levels, higher birth rates, more health issues, greater unemployment, higher crime rates and more nationalities compared to citizens in the rest of Aalborg.

On the other hand, Aalborg East is one of Aalborg's largest and most dynamic suburbs with massive investments in urban development; a new University Hospital is under construction, the university is expanding heavily, business parks are being developed and new housing areas are being planned. There are also political intentions for developing Aalborg East to a modern and sustainable suburb.

In parallel to the municipal and private investments around Aalborg East, several of the area's major Housing Associations are initiating plans for physical re-development of the area with new infrastructures, demolishing of worn-down buildings, building of modern apartment blocks, developing recreational areas etc. If the life between the buildings shall change as well, local residents need to take more action in preserving and developing the urban life in Aalborg East. In other words, the timing for creating an intervention on peace innovation in Aalborg East was perfect.

In the Spring 2014, Himmerland Housing Association invited Stanford Peace Innovation Lab, Aalborg University and the consultancy, COWI, to host a 2-day workshop in Aalborg East. The purpose was to create systematic, social and economic impact while increasing life quality and citizen engagement. A location was found (the newly built Health- and Neighborhood Center) experts and stakeholders were invited (education institutions, banks, insurance companies, private business, social workers, police, Aalborg Municipality, the Major and other politicians, consultancies etc.) In total, more than 100 local and external experts participated in the kick-off event. The following two days of workshops was structured around the following framework, which also has been utilized by Stanford Peace Innovation Lab in other urban contexts around the world:

- 1) *Identification of positive and negative behaviors in the local community*
- 2) *Identifying stakeholders, incentives and economy*
- 3) *Targeting a specific change of behavior*
- 4) *Detailed mapping of a behavior plan*
- 5) *Identification of applicable persuasive technology*
- 6) *Developing prototypes*
- 7) *Validation of prototype*

The overall point was to show how persuasive technology can be applied for peace innovation in Aalborg East only within two days. After the two days, work was done to refine the prototype, conduct investment plans, develop a business around the new technology-application and plan for long-term business model innovation of peace technology in Aalborg East. The overall target was to establish a local branch of Stanford Peace Innovation Lab in Aalborg East in order to, on a daily basis, develop positive behavior directly in collaboration with the local residents, educations, businesses and public authorities.

#### B. The Process: Two days of workshops

The workshop was initiated by group discussions on positive and negative behaviors in the neighborhood of Aalborg East. A key result was, that local experts, consisting of inhabitants, schools, social workers etc. presented a positive picture of Aalborg East as nice a place to live – in sharp contrast to the dire picture painted by the media. It was acknowledged that:

- a) *The civil society and the volunteer associations are well established.*
- b) *The green environment has great recreational qualities.*
- c) *The diversity in the area is an advantage for social complexity and networks.*
- d) *The pioneering spirit and positive engagement are pivotal to make positive behavior change.*
- e) *There exists good collaborations between schools, police and social workers.*

- f) *There exists a variety of unused competences and experiences in the local community.*
- g) *There are well established structures for collaborations and clear decision making processes*
- h) *The area is experienced as progressive and under positive development with plans for new welfare entities and infrastructure.*
- i) *The inhabitants of Aalborg East has a high degree of tolerance.*

In order to narrow down the target group and target behavior, the list of positive identifications of Aalborg East was specified into three subjects by votes from the local experts. The results showed, that especially these elements were key to build positive and persuasive technology around:

- 1) *The quality of the surrounding areas*
- 2) *The perceived safety and handling of crime prevention*
- 3) *Supporting healthy habits.*

After identifying possible entities for positive behavior change opportunities for intervention points were discussed. Especially, the stories of Aalborg East plays important roles for the image and attractiveness of the area, and hence of the security and possibilities for prosperity. For example it was clear that:

- a) *The media rarely tells the successes in Aalborg East, e.g. that the crime rate in the area has decreased by 42 % in the period from 2011-2013.*
- b) *New physical development in the area contrasts hopelessness and social despair in many local families due to unemployment etc. As an example, only few new jobs are offered to local residents.*
- c) *Further enriching of the quality of the green environment was needed.*
- d) *It is needed to attract more activity to the community by non-residents.*

After the initial workshops, identifications of behaviors, the stakeholders, incentives and economics were discussed in terms of four aspects: People, Organizations, Places and Things. These were linked in multiple pairs of positive incentives.

**Example 1:** *Schoolgirls need other role models from the outside environment in order to adopt positive behaviors.* Relations could be built between school girls in the age of 10-15 years and male players from the local soccer club. The girls could be motivated and triggered to healthier eating habits but interacting with local soccer players, e.g. sharing recipes or images of their healthy foods with immediate responses via smart phones. To support the behavior, local supermarkets could engage in the shopping experience, schools could take up learning on nutrition and social workers could establish cooking classes.

**Example 2:** *Physical movement and social interactions between mothers and sons can substitute alcohol intake.* The basic idea is to substitute mothers' drinking time with movement time, bringing the mothers out in the society while reducing alcohol consumption. The app is a game where the boys in 4th grade compete to make their mothers move. The movement is measured by the GPS in the mothers' phones.

**Example 3:** *Relations between schoolboys from different grades builds relations rather than distance.* A mobile game, using QR-codes and tablets, is intended to trigger schoolboys from 1st and 5th grade to meet at the school at the proper time every morning – and further, to assist the younger with home works.

More than 50 examples of potential behavior changing ideas using persuasive technology materialized within hours. Many of the ideas had the potential of turning into real business cases with only a small amount of further development.

According to the Persuasive Design Loop, the prototypes were tested quickly. The initial test was conducted on familiar residents in the neighborhood, outsiders, etc. via Facebook, LinkedIn, Instagram etc. Within 60 minutes, a number of replies ticked in with feedback and other comments. Some of the examples of persuasive behaviors were tested directly via participants' smart phones etc.

After the tests, the prototypes were refined in terms of ensuring proper motivation, abilities and triggers for the technologies to change behavior as intended. Further testing were conducted and further refinements were applied – all within few hours.

### C. The Outcome: The Neighborhood App

The workshops in Aalborg East was conducted with regard to the Peace Innovation Design Loop as briefly presented in Chapter II. The Loop was conducted more than 50 times in various efforts in order to search, iterate and prototype various solutions. In this

#### 1) Choose Target Communities

The first task was to determine the target communities among the almost 15,000 inhabitants in Aalborg East. Firstly, it was discussed whether peace technologies could be developed to comprise all inhabitants. or if the solution only should involve a specific target groups. It was decided, that the first solution should involve as many inhabitants of Aalborg East as possible in order to gain the maximum effect and positive press coverage.

#### 2) Observe/choose the tech they use

It was observed, that most inhabitants of Aalborg East are using smart phones on a daily basis. Smart phones include Accelerometer, Gyroscope, Proximity sensor, Pedometer and a multitude of additional sensors. Hence, the smart phone was decided as the technological platform for persuasive design in Aalborg East.

3) *Pick a positive engagement behavior*

Throughout the many iterations it was clear, that the basic thing of 'knowing each other' was key for increasing social coherence, security and coping with potential urban unrest.

4) *Create Fast Prototype Interventions*

As explained in the previous section, more than 50 different prototypes were tested on various levels. It turned out that increasing the social factor of belonging, friend making etc. was one the most important issue. A specific prototype was developed during the workshop by the tech companies, SenseMate and IT Knows.



5) *Measure Impacts*

Some of the prototypes were sorted out, as their effects were hardly measurable. By applying means of data measurement, surveillance of technological usage etc. it was clear that the persuasive design was easily monitored and refinements could easily be based on quantifiable data.

6) *Optimize & repeat*

An optimization process of the final product could not be completed within the 2 days of workshops. However, the idea of an Neighborhood App was born. With greeting to newcomers, calendars of events, links to social services and so one, activities could be scheduled much faster and with much broader attention than earlier. The Neighborhood App still needs further refinements in order to work properly.

IV. FURTHER APPLICATIONS OF PEACE TECHNOLOGY FOR SOCIAL HOUSING

A range of other target groups, technologies and behaviors could be subject for applying persuasive technologies for reducing urban unrest. The following describes other examples, discussed on the workshops in Aalborg East and with other Danish Housing Associations.

A. *Social Network for Local Micro-jobs*

Economic incentives are great motivators for behavior change. Ownership to the local community is also a strong driver for reducing vandalism and increase positive behavior. Combined with the trigger on a smart phone providing new micro-jobs such as helping neighbors with everyday challenges

like babysitting, shopping, cleaning, gardening etc. are also possible persuasive designs for reducing urban unrest. Further, such a framework provide structured data on the positive behaviors. The solution could also transpire into gamification designs by applying a 'social score'.

B. *Sensors on Lighting Paths*

An issue discussed in Aalborg East was the lack of street lightings during night time. The dark paths felt unsafe, keeping inhabitants indoor. As vandalism, attacks and robberies often tends to happen at night time, street lights are necessary for reducing the fear. Another issue is the costs of electricity for the lampposts – if the lights should be on at all times, for example, the costs would lay on the Municipality. As a peace technology, sensors could be built into the lampposts. The electricity reducing issue could be dealt with while applying a technology, that detect a person or a smart phone, and would then turn on street lights as pedestrians or cyclists are passing by. For reducing costs to the sensors, it could be contractual mandatory for the supplier of lampposts to build in such technology.

Multiple other solutions for reducing urban unrest with peace innovation technologies are GPS on machineries for detection of vandalism or theft, sensors on doors for sensing unwanted intruders etc. Only the fantasy sets the limits.

V. DISCUSSION AND CONCLUSION

This chapter discuss and conclude on the case of applying persuasive design and wireless technologies for mitigating urban unrest.

Firstly, it can be discussed whether everybody can develop peace innovation using persuasive technologies. The experience is that it is vital to include as many of the local experts in the neighborhood, empowering them by giving them the responsibility for further development. It is unlikely that a top-down approach would be more appropriate than the bottom-up approach described in this case. It is acknowledged that peace innovation can be done in every place on Earth where the majority of local residents has access to wireless technologies for information and communication.

Secondly, it can be discussed if a two-day intervention really matters. We would argue that the invitation of local stakeholders and the inspiration of a two-day workshop with world-class experts implied some insight that applying persuasive technology for peace innovation is not rocket science – at least when it comes to the conceptual designs. It is not likely, that a week-long workshop would have improved the results. However, the lack of continuity due to the lack of sustainable business models was a major flaw.

Thirdly, one obvious flaw of this study is the absence of actual business model innovation. In only two days, it was nearly impossible to set up a sustainable innovation lab for developing persuasive business models. It can be discussed, if the value chain where too long. Hence, the time from implementing safety increasing solutions to the monetary benefit from the financial institutions was too long and uncertain. It would probably be easier to implement persuasive

technology for behavior change if the business model was much more simple, direct and based on micro-business.

Summing up it can be concluded that persuasive technology might be an effective way for mitigating urban unrest. However, completing the intervention by actually developing a local and sustainable framework for developing a perspective of multi business model innovation with direct and short-term value propositions is needed.

#### ACKNOWLEDGMENTS

The authors want to thank the following contributors to the findings:

- a) *Stanford Peace Innovation Lab and its affiliated Danish partners who provided thoughtful insights in persuasive design and for helping design the workshops in Aalborg East.*
- b) *COWI and Himmerland Housing Association for setting the scene, preparing the workshops and for thoughtful insights on urban planning and social housing areas*
- c) *SenseMate and ITKnows for leading the prototyping during the days and for thoughtful insights on applied and rapid prototyping*
- d) *All the participants and local experts who contributed with valuable knowledge about the communities.*

We are hoping to bring inspiring insights on how to cope with urban unrest by applying peace technologies and wish Aalborg East all the best in further developments and applications of persuasive designs for positive behaviors. As the workshops and other findings suggest, persuasive design is all about a 'just-do-it'-principle.

#### REFERENCES

- [1] United Nations, World Urbanization Prospects 2014 revision. Department of Economic and Social Affairs, 2014.
- [2] P. Knox and S. Pinch, *Urban Social Geography: An Introduction* 6th ed. Essex, Prentice Hall, 2010.
- [3] M. Dear. and J. Wolch, *Landscapes of Despair*. Princeton, NJ: Princeton University Press, 1991.
- [4] D. Kilcullen, *Out of the Mountains: The Coming Age of the Urban Guerrilla*. London: C. Hurst & Co., 2013.
- [5] Institute for Economics & Peace, *Global Peace Index* 2016.
- [6] D. Stauffacher, W. Drake, P. Currión and J. Steinberger, *Information and Communication Technology for Peace*. NY: The United Nations Information and Communication Technologies Task Force, 2005.
- [7] B. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. ACM, 2009.
- [8] BJ. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, Morgan Kaufmann. 2002.
- [9] L. M. A. Bettencourt, The origin of scaling in cities, in *Science*, vol. 340, pp. 1438–1441, Jun. 2013.
- [10] M. Quihuis, M. Nelson and K. Guttieri, "Peace Technology: Scope, Scale and Cautions", *Building Peace*, vol. 5, pp. 14-16.





# *Technological driven business model innovation: a pragmatic take on human bond communication*

John Bang Mathiasen & Jacob Steendahl Nielsen  
 Department of Business Development and Technology,  
 Aarhus University, Denmark  
 ”johnbm@btech.au.dk & jacobsn@btech.au.dk”

## **Abstract**

*This paper aims at increasing the prevalent understanding of how the communication processes in a supplier/customer relationship influence the implementation of a new Business Model (BM). The research draws on a case study conducted at a supplier company involving cross-functional and interorganisational communication within two rather different BMs; a well-established BM and a new BM. The results illustrate that; the communication in the new BM is highly influenced by the well-established BM “mind-set”; the new BM requires adaptations in the communication patterns; an essential prerequisite for a successful BM implementation is an adaptation of the social communication patterns to facilitate trust-buildings as well as the development and implementation of new communication technologies.*

*Keywords; Business models, pragmatism, communication, product development*

## I. Introduction

Technological driven business model innovations necessitate interorganisational and cross-functional communication processes [1]. To facilitate the communication processes companies can apply for instance the Business Models (BM) cube [2], BM canvas [3], or Baden-Fuller and Mangematin’s [4] four typological dimensions of BMs. The above valuable models seem to either advocates for a descriptive agenda focusing on exogenous matters or a cognitive agenda highlighting endogenous matters for facilitating the communication processes. The descriptive agenda does often consider the BM as something real, while the cognitive agenda emphasises that the managers’ reflective thinking determine the innovation of the BM [5]; i.e. communication processes. As for the latter endogenous matters, Prasad [6] introduces and elaborates the concept of “*human bond communication*” to grasp how human faculties influence the communication processes. However, drawing on pragmatism [7], the endogenous and exogenous matters reciprocally evolve, which means that human faculties are embodied yet contextual embedded. To the best of our knowledge, this complex two-way reciprocal interaction between the endogenous and exogenous matters has received little attention in BM research. Accordingly, the purpose of this research is to shed light on BM innovation in a supplier-customer collaboration developing wind turbine control applications and “*to which extent the employees’ communication process influence the implementation of a new BM*”.

This research applies the following approach. By adapting a Deweyan [8]; [9] pragmatic understanding of experience and inquiring and thereby subscribing to the practice-based viewpoint emphasising an analytical focus on “the doings” rather than “the individual actors” [10], we present empirical and analytical findings of an innovation of a new BM within situational practices from the point of view of pragmatism bringing the doings and consequences of the doing to the fore in the attempt to analyse the reciprocity between endogenous and exogenous matters.

## II. Theory

One approach to BM innovation is to provide a number of generically communicative descriptors of how a firm organises itself to create and distribute value in a profitable manner. For instance, a BM is a description communicating how companies functions [11], how companies collaborate with different stakeholders as for instance suppliers and customers [12] and how this inflow and outflow of information influence the BM innovation [1]. Osterwalder et al. [13] puts forward that “*A business model describes [communicates] the rationale of how an organization creates, delivers and captures value.*”. Osterwalder’s Canvas [3]; [13] approach suggests that nine matters influence the communication when doing the BM innovation, however the values proposition matter seems to be rather central in innovating a new BM. A group of researchers argues that the communication when doing BM innovation needs to take into consideration the innovation of other BMs and suggesting the concept of “multi BM” and in addition, that multi BM innovations need to have a more balanced approach to the influencing matters; for instance Lindgren and Rasmussen [2] rank alongside six matters and their relationships when doing multi BM innovation.

A group of BM researchers [4]; [5]; [14] departures from the above generically descriptive agenda and suggesting bringing cognition to the fore to reveal the communication processes. Perkmann and Spicer [14] consider BMs as recipes instructing the employees doings when they are involved in innovating a new BM. According to these two researchers, humans are often guided in their decision-making by cognitive frameworks (like [4] and [5]) that privilege certain courses of action to the exclusion of others. Indeed Baden-Fuller and Morgan [15, p. 168] suggest that “*Business models are not recipes or scientific models or scale and role models, but can play any - or all - of these different roles for different firms and for different purposes: and will often play multiple*

roles at the same time". Despite these cognitive inspired approaches regard BM as being representations that create material effects such as involving buyers and suppliers in BM innovations, persuading investors, and directing employees (see [14]), the models do not contribute to improve our understanding of how the complex and reciprocal interaction between endogenous and exogenous matters influences the communication processes.

Prasad [6] introduces a novel (and interesting) concept of "human bond communication", which addresses five human senses (seeing, hearing, smelling, tasting, and touching). This focus on human faculties expands the cognitive agenda discussed in the above and illustrates that useful communication when doing BM innovation should not be taken for granted. A human is not just a "cognitive phenomenon"; rather humans have different biological faculties, which influence the communication processes among the involved employees. However, as the Deweyan [9, p. 32] pragmatism reminds us "every organic function is an interaction [transactional relationship] of intra-organic and extra-organic energies, either directly or indirectly.... An organism does not live in an environment; it lives by means of the environment". In other words, Prasad's [6] "human bond communication" does solely address the endogenous matters – the human biological faculties, which implies that he only analyses one side of the coin in relation to communication processes. The other side of the coin is the exogenous matter.

Drawing on Dewey [8] [9] we take the position that endogenous matters are embodied in the humans' biological faculties such as sensing, thinking and feeling, while exogenous matters are embedded within practices in which the communication processes unfold. The two matters (in the word of Dewey, the two sources of energies) are not two separate entities; instead, they are completely embedded within the practices. Hence, the employees' communication processes do not occur "in" the practice; rather their communication processes unfold as consequences of actively applying usable exogenous matters "within" situational practice. This "lives by means of" the practice causes the exogenous matters within the practice to be intertwined with each employee's senses, thoughts and feelings. As Nicolini [10, p. 2] reminds us "Things seem to fall into place much better if we think of the fluid scene that unfolds in front of us in terms of multiple practices carried out at the same time".

Useful communication processes do not take place per se. First, humans are heterogeneous and consequently humans have different level of biological faculties as sensing, feeling, thinking and acting [9]; [16], which means that the employees involved in BM innovation have different faculties for communicating as well as grasping and utilising the communication to create new/valuable communication. Second, exogenous matters are heterogeneous; for instance a meeting room having different decor of technical equipment, IT systems, video conference facilities and furniture or a production facility as well as different accessibility of sketches, drawings and physical breaker panels and also the employees' laptops and mobile phones and sketches/drawings of BMs. These technologies (exogenous matters) have inherent properties; i.e., an IT-application has intrinsic properties enabling electronic based communication meaning that communication and information are embedded in

the physical and digital materials of technological devices. It implies that communication occurs in and the value of the communication are utilised in a reciprocal interaction involving humans' biological faculties and all exogenous matters within the situational practices.

### III. Method

The research is empirically driven and originates from constructionism [17]. The research approach extends a Deweyan conceptualisation of experience and inquiring and Nicolini's [10] conceptualisation of practices. The research strategy draw on Dubois and Gadde's [18] systematic combining – an abduction approach.

The data collection is conducted by one of the two authors. During a four-month period, one of the authors has a desk in the middle of an open-plan office available, which makes it possible to participate in small talk and to take the role as complete observer [19]. The small talk and traditional observations take place in the open-plan office, canteen and production area. Furthermore, individual and group interviews with 14 employees are conducted; all interviews are unstructured [19]. Prior to and during the period in which the unstructured interviews are conducted, one of the authors benefits from collaborating with a gatekeeper who helps to identify the informants to be interviewed. The criteria used for selecting the informants imply that different hierarchical levels and all functions influencing the innovation processes are represented. The interviews are not taped; instead, notes are taken. Immediately after each interview, a detailed summary is written. In addition, one of the authors participate in four meetings, including a video conference with the Polish subsidiary. During these meetings, the observer-as-participant [19] is applied. Notes dealing with the content of dialogues are written down; immediately after the meetings, a summary is made.

This research does only study two different BMs; one company's collaboration with two different customers. This limited empirical material may question the generalisability of the findings. In the same vein, many of the empirical statements from observations in the open-plan office, production area, canteen and meetings as well as from the unstructured interviews are omitted in the empirical description in this paper. Despite this omission, these absent interviews and observations have contributed to the authors' understanding, which all thing being equal challenges the credibility of the paper.

### IV. Empirical findings

Supplier (pen-name) has developed and produced Wind Turbine Control (WTC) applications ever since Christian Riisager laid the foundation to the modern wind turbine industry more than three decades ago. Supplier has been in the wind turbine industry since then and the company is well-established within the market for onshore and offshore WTC as well. Supplier is located in the middle of Jutland, Denmark. A production subsidiary was established in Poland to carry out batch production. Furthermore, in cooperation with Chinese Corporation, a Chinese Joint Venture (JV) was established in November 2008. All prototypes are produced in Denmark, while the mainstream production has been transferred to Poland.

### A. The background for a new BM

Supplier has two groups of customers; the key customer is pen-named Oldtimer. Oldtimer is a global player within the wind turbine industry and has produced a great many offshore as well as onshore wind turbines. Both types of wind turbines are considered to be very reliable resulting in Oldtimer enjoying a good reputation in the wind turbine market. The collaboration is (was) characterised by an intensive virtual and face-to-face communication – system supplier relationship, as Supplier has created technical applications to Oldtimer ever since the industry was in its infancy in the late seventies. Thus, there exists a well-developed mutual understanding and communication among the employees. The reason why “was” is placed in brackets is due to the fact that Oldtimer has cancelled an exclusive agreement between the two companies. In addition, Oldtimer has decided to develop all SW in-house implying that Supplier now only creates the HW part of the technical control applications; in other words, their role as being system supplier to Oldtimer is cancelled.

The cancellation of the above system supplier agreement resulted in that Supplier initiated a search for developing a new BM. In such an attempt to gain an understanding of potential customers’ requirement and thereby attract new customers, Supplier established a department termed New Business Department (NBD) to innovate and implement the new BM. The common denominator of this group of customers is that it consists of organisations lacking in wind turbine experience. So far (the time for data collection), Supplier has interacted with a number of new customer and the first deliveries has been executed. However, in contrast to Oldtimer, the communication between Supplier and the new customer call for exhaustive collaboration both inter-organisationally and cross-organisationally as neither social bonds nor technical pre-understanding exist.

### B. Communication in the well-established BM

Despite the fact that the aforementioned cancellation of the system supplier agreement and thus Oldtimer’s insourcing of SW activities entailed some changes in the development approach, the three decades of intense collaboration still influences the interorganisational communication processes. For instance, after meetings at Supplier, the employees from Oldtimer are passing by the desk of the Supplier employees, who did not participate in the meeting, in order to clarify technical issues. On the one hand, these informal communication processes are occasionally regarded as being problematic, on the other hand, however, this mutual communication and thus understanding is the reason why supplier is able to produce a WTC to Oldtimer within the current BM. As one of Supplier’s technical employees express it *“The quality of this document [specifications received from Oldtimer via an email, author] is inadequate. If I did not have a good understanding and knowledge of Oldtimer’s requirements and wishes, I would not be able to build a breaker panel from this document. It is necessary for me to read the text between the lines. Another supplier would not be able to produce a breaker panel based on this document”*.

### C. Communication in the new BM

As for the new BM, only a few innovation projects have been conducted in accordance with the established guide-lines for the new BM for which reason the experience in this regard is

still in its infancy. Supplier is working determinedly to improve the BM procedures for guiding the communication processes and collaboration with potential new customers. Despite this increasing formalisation of the communication processes, employees from NBD highlight two hurdles when signing an agreement with a new customer: Cost and delivery time are the core value pro-position in the new BM. However, referring to these employees, it is problematic that the engineers always start from scratch when developing technological solutions. Seen from the engineers’ perspective, the cross-functional communication and collaboration in the new BM involving sales and the technical employees give rise to misunderstandings downstream as well as upstream. According to this group of employees, the sales employees do not establish appropriate communication media and thus do not attempt to understand a potential new customers’ needs/wishes before presenting solutions to potential new customers; instead, the communication and collaboration are regarded as a “feel one’s way process”. It causes vague technical pre-specifications, which often have to be modified later on during the innovation processes.

### V. Analysis of the two BMs

The situational practices cross-functionally Supplier are characterised by a craft-based history in which the development of the technical application to Oldtimer forms the basis of the organisation. In this regard, the top management has implemented some changes in both BMs; for instance, it has reversed the orientation of the production from craft-based to industrialised and reorganised all departments involved in the responsibilities in terms of the new BM activities.

Irrespective of whether the innovation activities take place in collaboration with a new customer or Oldtimer, the approach to conducting the communication and collaboration draw on three decades of intense system supplier relationship with Oldtimer and existing technical solutions (drawings, diagrams; from now on artefacts). This way of doing business in both BMs forms the backbone of the organisation; the internal Supplier jargon is to term it the “Oldtimer mind-set”. Anyhow, when Supplier employees communicate, collaborate and present technical solutions (artefacts) to a potential new customer, or for that matter to Oldtimer, the underlying basis is the “Oldtimer mind-set and existing technical solutions”. One might say that it is a kind of Oldtimer logic that influences both the virtual and face-to-face communication processes – i.e. the practices in which the BMs function.

Addressing the communication and collaboration involving Supplier and Oldtimer, the communication about artefacts (specifications, subsequent production preparation and physical production of the technical applications) clearly demonstrate the constitutive effect of three decades of close communication and collaboration. The technical employees draw on their technical experience and existing technical solutions gradually developed in close communication and collaboration with Oldtimer. The combination of technological awareness and well-established communicative social bonds among the technical employees enables them to “read between lines” in terms of the applied artefacts and thus produce a technical applications despite vague specifications. That is, the underlying basis for doing PD activities is the accumulated experience (endogenous and exogenous matters) – social communicative bonds as well as

technical communications – for instance, several times during meeting both Supplier’s and Oldtimer’s employees use their mobile-phones and/or laptops to virtual communication with employees, who do not participate in the meeting.

As for the new BM, the situational practice being established regarding the communication processes and collaboration with potential new customers is characterised by the fact that both companies navigate through uncharted territory. The agenda for evolving the new BM is to develop a wind turbine producing electricity, and accordingly, the

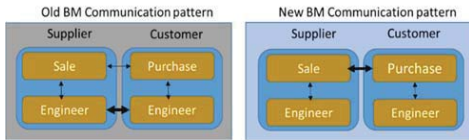


Figure 1: Communication pattern

technical matters seem to be at the top of the list of priorities; however, an ongoing process is to improve the mutual understanding and thus the communication between the sales employees and the potential new customer in question. In the new BM it seems to be crucial to nurse and develop the relationship to prove Supplier’s trustworthiness. Hence, Supplier seems to attach great importance to continuously improving the virtual and face-to-face communication and thus the mutual understanding between the organisation and the new customers. Figure 1 illustrate the change in communication patterns. The arrows illustrate who is communication with who and the thickness of the arrows illustrate the degree of the communication.

## VI. Discussion and conclusion

At the outset, the purpose of this research is to shed light on BM innovation in a supplier-customer collaboration developing wind turbine control applications and “to which extent the employees’ communication process influence the implementation of a new BM”.

The analyses indicate that the starting point for nursing and improving the mutual understanding varies from customer to customer. For instance, the pattern of communication and relationship building activities initiated in connection with a new customer from the Far East turned out to be rather complex and unpredictable. In contrast, improving the communication and mutual understanding with a new customer located in the North part of European proved to be more predictable and effortless. Just after visiting the last-mentioned new potential customer, the technical salesman comments “we are at the same wavelength”. Thus, despite the company had a “recipe for the new BM” and the employees applied this as suggested by [4]; [5]; [14]; [15], the communication processes constrained the implementation of the new BM. Our findings illustrate that the application of a recipe are influenced by endogenous matters (like [6]) as for instance a “prevailing mind-set in the company” and the engineers’ biological faculties in a reciprocal interaction with the exogenous matters as for instance technical artefacts.

Directing attention to the potential new customer from the Far East, the technological gap is very wide as this new customer only has limited experience with wind turbines. The customer has previously developed two wind turbines, which both suffer from quality problems. Hence, the findings in our

analysis indicates that the technological content of experience becomes important for the communication processes, as this new customer really needs to be convinced that Supplier’s technological artefacts are reliable. This indicates that the communication processes influence the evolution of and thereby the implementation of new BMs: an intense face-to-face communication paves the way for improving the understanding in relation to the technical artefacts including the interfaces to all other sub-systems in the wind turbine. When communicating and creating the technical artefacts in collaboration with the aforementioned European new customer, the experience gap is considered to be narrower and therefore more manageable. It turns up, that the project manager from the European new customer has previously been employed at Oldtimer and thus is familiar the Supplier’s existing technological artefacts, communication processes and especially “the mind-set” when doing the development in the new BM. Thus, the degree of complexity in relation to virtual and face-to-face communication in term of necessary learning processes to establish the necessary foundation for signing an agreement (artefact) as well as to accomplish the innovation activities diverges. Our findings reveals that communication processes do not take place per se and that human faculties [6] influence the communication processes (like [6]). However, the analysis clearly illustrates that a human’s faculties depends on and evolves in a reciprocal interaction with the applied technical solutions/artefacts.

Oldtimer is exceptionally well-versed in wind turbines, for which reason this customer is very much aware of its value propositions - needs and requirements - in terms of the technical artefacts to be created in the existing BM. In contrast, a potential new customer’s experience in relation to wind turbines in general and Supplier’s technical artefacts in particular is rather limited. Customers in this new BM do not have a clear sense of their value proposition for the technical artefacts. Typically, in the Oldtimer BM, the customer takes the reins in terms of communication and collaboration when drawing up technical artefacts, while Supplier leads the communication and collaborative way in relation to the BM for a new customer. As mentioned in the theoretical chapter the Cube approach [2] ranks alongside the BM elements, while the Canvas approach [3]; [13] brings the clarification of the value proposition to the fore when innovating BMs. Our findings support the last-mentioned approach.

The BM innovation activities performed with a new customer are challenging; especially the learning processes and the technical clarifications of the artefacts are complex and require an ability to communicate with and last but not least to understand the new customer’s value proposition and perspective of the collaboration. In addition to the fact that different languages/cultures may give rise to communicative misinterpretations, there seems to exist a technology gap between a new customer’s expectations to the technical artefacts and the one being offered as a solution. The technical solution proposed to a new customer draws on three decades of relationship with Oldtimer and this “Oldtimer mind-set” in combination with the application of existing technical solutions influences the communication with a potential newcomer. An automobile metaphor might illustrate the tension field within the new BM. “The creation of technical application in collaboration with Oldtimer is like a Mercedes; it is a high-end car, which is regarded as being very reliable. However, a new customer does not necessarily

need a Mercedes; it has too many dispensable features available, making it excessively expensive. It might be that the new customer need a Tata car! Accordingly, facilitating appropriate communication bonds to enable a proper learning process of the customer's real requirement is central for evolving a BM.

Facilitating appropriate communication bonds to enable proper learning processes of the customer's real requirement is central for evolving a BM. As illustrated in Figure 2 the degree of communication at the engineer level develops over time, but the success of the new BM depends on finding a new way to evolve this communication bond faster (from curve A to curve B).

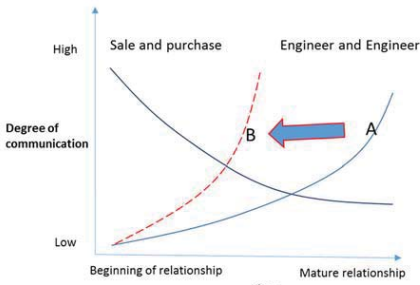


Figure 2, Change in who communicate and to what degree over time

The practical implication from this case study highlights two factors that practitioners should consider as means to improve the communication bonds. First practitioners should ensure that usable artefacts are accessible as this improve the communication processes. Second, if information is missing to ensure a continuation of the situational communication processes practitioners should proactively make use of new communication technology to gain access to the missing information; in the Oldtimer case we experienced the positive effect of an adaptable use of new communication technology to gain access to missing information.

References

[1] H. Chesbrough, and K. Schwartz, Innovating Business Models with Co-development Partnerships. *Research Technology Management*, Vol.50, No.1, pp.55-59, 2007

[2] P. Lindgren, and O.H. Rasmussen, The Business Model Cube, *Journal of Multi Business Model Innovation and Technology*, pp.135-182, 2013

[3] A. Osterwalder, and Y. Pigneur, *Business model generation: a handbook for visionaries, game changers, and challengers*: Wiley, 2010

[4] C. Baden-Fuller, and V. Mangematin, Business models: A challenging agenda. *Strategic Organization*, Vol. 11, No. 4, pp. 418–427, 2013

[5] C. Baden-Fuller, and S. Haefliger, Business Models and Technological Innovation. *Long Range Planning*, Vol. 46, pp.419–426, 2013

[6] R. Prasad, Human Bond Communication, *Wireless Personal Communications*, Vol. 87, No. 3, pp 619–627, 2016

[7] J. B. Mathiasen, and C. Koch, Product development as reading and writing doings within sociotechnical practices: the reciprocity between engineers and artefacts, *Technology Analysis & Strategic Management*, Vol. 27, No. 5, pp.604-620, 2015

[8] J. Dewey, *Essays and How We Think. The Later Works Volume 8*, Southern Illinois University Press, 1933

[9] J. Dewey, *Logic: The Theory of Inquiry. The Later Works Volume 12*. Southern Illinois University Press, 1938

[10] D. Nicolini, *Practice Theory, Work, and Organization*. Oxford: Oxford University Press, 2012

[11] J. Magretta, Why business models matter. *Harv Bus Rev*, Vol. 80, No. 5, pp. 86-92 2002

[12] C. Zott, and R. Amit, The fit between product market strategy and business model: implications for firm performance. *Strategic Management Journal*, Vol. 29, No. 1, pp. 1-26, 2008

[13] A. Osterwalder, Y. Pigneur, and C.L. Tucci, Clarifying business models: Origins, present, and future of the concept. *Communications of the association for Information Systems*, Vol. 16, No. 1, pp.1-25, 2005

[14] M. Perkmann, and A. Spicer, What are business models? Developing a theory of performative representations In *Technology and Organization: Essays in Honour of Joan Woodward*. Vol. 29: pp.265-275, 2010

[15] C. Baden-Fuller and M.S. Morgan, Business Models as Models, *Long Range Planning* Vol. 43, pp.156-171, 2010

[16] B. Elkjær, Organizational Learning: The 'Third Way'. *Management Learning* Vol. 35, No. 4, pp.419-434, 2004

[17] M. Crotty, *The Foundations of Social Research: Meaning and Perspectives in the Research Process*. Sage Publications Ltd., 2012

[18] A. Dubois, and L.E. Gadde, Systematic combining: an abductive approach to case re-search, *Journal of Business Research*, Vol. 55, pp.553– 560, 2002

[19] A. Bryman, and E. Bell, *Business Research Methods*, Second Edition, Oxford University Press, 2007

## Global Standards, IPR and Patents Business model Ecosystems

### “An Outlook to future Global SIP BMES Business Models”

**Peter Lindgren**

Aarhus University, Business Development and Technology, Denmark

peterli@btech.au.dk

**Krishna Kumar Sirohi,**

Chair, Gsfi Standards Committee - Founder & President, I2tb Research Foundation

president@i2tb.in

Aarhus University, Business Development and Technology, Denmark

**Jesper Bandsholm**

Bandsholm@tech.au.dk

### ABSTRACT

During the last decade sensors, wireless and IOT technologies in general have influenced our everyday life and have increased many-fold the potentials of our Businesses and Society by shaping old business model ecosystems (BMES) and creating new. In this “world” exponential growth of standards, IPRs, Patents (SIP) are constantly being developed in the favour of having more devices that can meet the benefits of standards, IPR’s and Patents. Different SIP BMES in ICT and related application areas, such as energy, telemedicine, wireless robotics, biotechnology put emphasis on participating and being important player in the Global SIP BMES potentials – focused on Standards, IPR and patents .

How can the Global SIP BMES be understood. How can SIP BMES innovated and operated to ensure the full achievement of their vision. How can SIP BMES at the same time secure sustainable business in a rapid and disruptive game of technology and business model innovation (BMI) of these SIP BMES.

Keywords: Business Model Ecosystems, Standards, IPR, Patents, Business Model Innovation

#### 1. Introduction - What is the Agenda in today’s SIP BMES today?

Worldwide there are several SIP BMES –that deals with the SIP’s of ICT and related areas.



Figure 1. Overview of the SIP BMES – specific Standards - layers inspired by Wang Ping (April 2011)

Some are international SIP BMES that develops international standards. The three largest and most well-established are the International Organization for Standardization (IOS), the International Electrotechnical Commission (IEC), and the International Telecommunication Union (ITU), which have each existed for more than 50 years (founded in 1947, 1906, and 1865, respectively) have established thousands of standards covering almost every conceivable topic. Many of these have been and are adopted worldwide replacing various incompatible "homegrown" standards. Many of these standards are naturally evolved from those designed in-house within an industry, or by a particular country, while others have been built from scratch by groups of experts who sit on various technical committees (TCs). These three organizations together comprise the World Standards Cooperation (WSC) alliance – in our BMES terminology - a part of a Global BMES for SIP.

A large variety of independent international SIP BMES such as the ASME, the ASTM International, the IEEE, the Internet Engineering Task Force (IETF), SAE International, TAPPI, the World Wide Web Consortium (W3C), and the Universal Postal Union (UPU) develop and publish standards for a variety of international uses. In addition to these, a large variety of independent international SIP BMES such as (ITU, GSC, ARIB, ETSI, TTC, TIA ....) run their own BMES independent but also collaborating together. Some are national standards or more regional standards NAT and REG SIP BMES. REG SIP BMES such as the European Committee for Standardization (CEN), the European Committee for Electro technical Standardization (CENELEC), the European Telecommunications Standards Institute (ETSI), and the Institute for Reference Materials and Measurements (IRMM) in Europe, the Pacific Area Standards Congress (PASC), the Pan American Standards Commission (COPANT), the African Organization for Standardization (ARSO), the Arabic industrial development and mining organization (AIDMO), and others. No matter what - standardization, IPR, Patens and the SIP BMES - have done until today – they had and still have a large impact and influence on our existing world and also future world. Just imagine the development of GPS and how this has developed over the years. Standards provides according to ETSI SIP BMES (ETSI 2016)

<b>Safety and reliability</b>	Adherence to standards helps ensure safety, reliability and environmental care. As a result, users perceive standardized products and services as more dependable – this in turn raises user confidence, increasing sales and the take-up of new technologies.
<b>Support of government policies and legislation</b>	Standards, IPR and Patens are frequently referenced by regulators and legislators for protecting user and business interests, and to support government policies. Standards, IPR and Patens play e.g. a central role in the European Union's policy for a Single Market.
<b>Interoperability.</b>	Ability of devices to work together relies on products and services complying with standards – but IPR and Patens also with business and protection of businesses competences – even core competences
<b>Business benefits</b>	Standardization provides a solid foundation upon which to develop new technologies and to enhance existing practices. Specifically standards: Open up market access
<b>Provide economies of scale</b>	<b>Standards provides business security for being able to produce large bats and invest in mass production</b>
<b>Encourage BMI</b>	<b>Standards provides business with increasing BMI further on behalf of standards</b>
<b>In cre ase awareness of technical developments and initiatives</b>	<b>Standards provides platform for increasing awareness</b>
<b>C onsumer choice</b>	Standards provide the foundation for new features and options, thus contributing to the enhancement of our daily lives. - Mass production based on standards provides a greater variety of accessible products to consumers.

**Table 1 Benefits of Standards according to ETSI SIP BMES**



Businesses and developers of technical standards are generally concerned with SIP, which e.g. detail how products interconnect with each other, and safety standards, which established characteristics ensure that a product or process is safe for humans, animals, and the environment. Their interest is also in defining how the behavior and performance of products and processes is measured and described in data sheets. To illustrate the impact in just a part of the ICT area of a SIP BMES figure 1 shows a technical understanding of these IPR and Standardizations issues and expected impact areas.

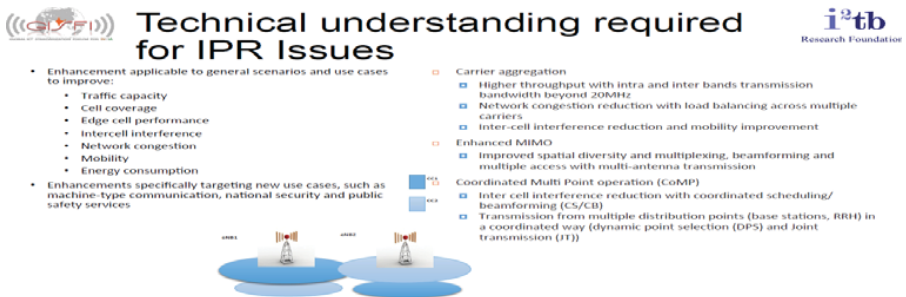


Figure 2: Overview of IPR and standardization issues related to 5G and a SIP BMES (Sirohi, 2016)

Overlapping or competing SIP BMES – that “rules the different SIP BMES in Global SIP BMES” - tend to cooperate although purposefully, by seeking to define boundaries – entry and exit barriers (Porter 1985) to their SIP BMES (Lindgren 2016) - between their scope of business, and by operating in a hierarchical fashion in terms of national, regional and international scope – SIP BMES. Unless adopted by a government - SIP’s carry no force in law. In other word SIP BMES have no power or business if governments do not support and adapt the SIP’s and support SIP BMES’s. However, most jurisdictions have truth in advertising laws, and ambiguities can be reduced if a business offers a product that is "compliant" with SIP’s. In the resent years the word “compliant” has be set under pressure, been defined differently and innovated disruptively in and outside many SIP BMES – in e.g. Health BMES, Furniture and Textile BMES, and also ICT and related SIP BMES.

## 2. The Business Model Ecosystem of SIP’s

Any SIP BMES is representing more business models BM’s from more businesses (Lindgren 2016). Equal to this each SIP BMES has more BM’s from more business included and as seen in figure 4 as for one of the young SIP BMES – the GISFI BMS were in 2016 involved in 7 working groups across different businesses and academia’s.





However as BMES, businesses increasingly begin integrating and sharing their BM's, BM dimensions and BM components and their BM begin to act accordingly and to certain behaviour patterns (Lindgren & Aagaard, 2014) the preview for the role of SIP BMES we claimed will increasingly be attach to anything, anybody, anywhere and anytime (Lindgren et al., 2013). This will happen both for BM alone but more important together with other BM's in networks. Soon everyone and everything at any time and in any place will have the possibility to act and be in contact with a multitude of BM's SIP's even that act persuasively. Then the question raise *"How to secure SIP's automatically through a world of SIP BMES that all wants to secure their survival of their BMES and their businesses BM's?"*. In effect, the next question also raises *"How to enable and secure SIP and SIP BMES development in a future world of dynamic, intangible, persuasive BMES BM's and technologies?"*

Tomorrow will be a world of network based on persuasive business model and technologies (Lindgren 2015) consisting of BMES and BM's with all kinds of SIP, SIP structures, SIP strategies/objectives and SIP offers, including virtual, digital and physical value propositions. Thus SIP business model providers and developers are facing a SIP context in which BMI will take place in BM's within a multitude of BM's (Mucchi et al., 2011). Those SIP BMES and SIP Businesses that are based on advanced SIP technology and BM's that can take SIP fast from creation, capturing, delivering, receiving and consumption will provide businesses with key competitive advantages – and thereby form the winning SIP BMES of the future. Those SIP BMES that can change fast to this context will have the opportunity to survive. SIP BMES solutions, that technologically and business-wise can meet different stakeholders demands for increasing persuasiveness, agility, flexibility, individualization, and privacy will stand a change to change competence of existing SIP BMES

An interesting synergy and spinoff of the above mentioned challenge to SIP BMES –is expected to be a development to a completely The new and changed understanding of IP's and BM's of such will be disruptive in character and context. This development we proposed would take SIP BMES from a static, physical, digital and proactive business model context to a relational interactive, sharing and persuasive business model context. This could create an optimum for individuals and businesses valuing "privacy, freedom, flexibility, agility, security – and SIP's at the same time.

#### 4. Conclusions

Different SIP BMES wants to stay major players in the Global SIP BMES world. A manifold of SIP BMES exists and future conditions for SIP BMES seems to be influenced by new disruptive technologies embedded in BM's that are and will be increasingly integrated in more and more advanced persuasive BM's. As persuasive BM's can be operating physically, digitally and virtually – integrated, connected and dynamic– and deliver value propositions, wherever and whenever the user, customer, network partner, employees, objects and businesses demand it – the SIP BMES BM's will be challenge. Future SIP BM's will not only be a matter of the 'surface' of things, places, people and time, but also as part of the BM's related to inside and outside things, bodies and time (SW2015)(SW2016)(NJIT 2016). How SIP BMES BM's can change to this context is still very open.–No matter what – "the rules of BM's" for SIP BMES is expected to change. There are proposals to solutions but they are still not covering all aspects of SIP. There are still questions to be explored before reaching the next step of SIP's BMES BM's.

#### References

- Afuah, Allan, and Christopher L. Tucci. (2001) *Internet BM's and Strategies: Text and Cases*. Boston: Irwin/McGraw-Hill, 2001.
- Chesbrough, H. (2002) "The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-off Companies." *Industrial and Corporate Change* 11, no. 3 (June 1, 2002): 529–55.
- Chesbrough, Henry William. *Open BM's: How to Thrive in the New Innovation Landscape*. Boston, Mass: Harvard Business School Press, 2006.
- Cyber Creative Institute (2016)
- ETSI 2016: <http://www.etsi.org/standards/why-we-need-standards>
- Fogg, B. J. *Persuasive Technology: Using Computers to Change What We Think and Do*. The Morgan Kaufmann Series in Interactive Technologies. Amsterdam; Boston: Morgan Kaufmann Publishers, 2003.
- Fogg, B. J., ed. *Mobile Persuasion: 20 Perspectives on the Future of Behavior Change*. Stanford, Calif: Stanford Captology Media, 2007.
- Gassmann, O, Ellen Enkel and Henry Chesbrough (2010) The future of open innovation R&D Management 40, 3, 2010 Blackwell Publishing Ltd
- Sirohi Kirshna 2016: IEEE 2016 (11/05/16) Krishna Sirohi's Talk-2 in IEEE 5G Symposium @ IIT Patna 2
- Lindgren, Peter, Morten Karmøe Sondergaard, Mark Nelson, and B.J. Fogg. "Persuasive BM's." *Journal of Multi BMI and Technology* 1 (2013).
- Lindgren, Peter, Yariv Taran, and Harry Boer. "From Single Firm to Network-Based BMI." *International Journal of Entrepreneurship and Innovation Management* 12, no. 2 (2010): 122.
- Lindgren, Peter, Yariv Taran, and Kristin Margrethe Saghaug. "A Futuristic Outlook on BM's and BMI in a Future Green Society." *Wireless Personal Multimedia Communications Symposia Proceedings*, August 10, 2010, 23.
- Lindgren, Peter; Aagaard, Annabeth "The Sensing Business Model"  
In: "Wireless Personal Communications", Vol. 76, No. 2, 2014, p. 291-309
- Mucchi, Lorenzo, Luca Simone Ronga, and Enrico Del Re. "Physical Layer Cryptography and Cognitive Networks." In *Trustworthy Internet*, edited by Luca Salgarelli, Giuseppe Bianchi, and Nicola Blefari-Melazzi, 75–91. Milano: Springer Milan, 2011.
- NJIT 2016 - Kickoff Operation of CTIF Global Capsule (CGC) jointly with the Round Table Discussions on Knowledge Home New Jersey Institute of Technology (NJIT) September 8, 2016 sponsored by IEEE
- Onur, Ertan. "TU Delft Presentation SW2010 Held at the Strategic Workshop on Sensor Networks in Florence, July 2010."
- Prasad Ramjee (2014) 5G: 2020 and Beyond. River Publishers, 2014. 126 p. (The River Publishers' Series in Communications). Publication: Book Pruthi, S. (2012). Wireless robotics: A history, an overview, and the need for standardization. *Wireless Personal Communication*, 64, 597–609 Ruggieri, M., Madsen, O. B., & Prasad, R. (2012). Editorial: Special issue on wireless robotics research and standardization. *Wireless Personal Communication*, 64, 457–460.
- Seiter, John S. and Cass, Robert H. *Persuasion, Social Influence, and Compliance Gaining*. Fifth edition. Boston: Pearson, 2014. Schøler, H., & Toftegaard, T. S. (2012). Wireless communication in mobile robotics a case for standardization. *Wireless Personal Communication*, 64, 583–596. SW2010 Strategic Workshop 2010 Held at Hotel Relais Certosa. Florence, Italy. Twelfth Strategic Workshop 2010. Distributed and Secure Cloud Clustering (DISC). May 26-28, 2010.
- SW2015. Strategic Workshop 2015 Held at Villa Mondragone Via Frascati, Monte Porzio Catone, Napoli, Italy." Seventeenth Strategic Workshop 2015. Human Bond Communications (HBC). May 18-20, 2015.
- SW 2016. Eigtheens Strategic Workshop 2016 Held at Glymour Hotel, Rejkavik, Island. Future Teleinfrastructure for Multi Sensory Devices (FIND) May 23-25, 2016.
- Taran, Yariv, Harry Boer, and Peter Lindgren. "Managing Risks in BMI Processes. Proceedings of the 11th International CINet Conference Practicing Innovation in Times of Discontinuity." Zürich, Switzerland: Continuous Innovation Network, 2010.
- Wang Ping (April 2011). A Brief History of Standards and Standardization Organizations: A Chinese Perspective East West Center
- Whinston, Andrew B., Dale O. Stahl, and Soon-Yong Choi. *The Economics of Electronic Commerce*. Indianapolis, IN: Macmillan Technical Pub, 1997.
- Wikipedia 2016 The LTE standardization case
- Zott, C., R. Amit, and L. Massa. "The Business Model: Recent Developments and Future Research." *Journal of Management* 37, no. 4 (July 1, 2011): 1019–42.

# *Public Engagement in a Research Project: designing citizen science activities as part of its business model*

Nello Colella

University of Cassino and Southern Lazio  
OrgLab – Laboratory of Organization Studies  
Cassino, Italy  
n.colella@unicas.it

Francesco Bolici

University of Cassino and Southern Lazio  
OrgLab – Laboratory of Organization Studies  
Cassino, Italy  
f.bolici@unicas.it

**Abstract**— The pervasive use of Information and Communication Technologies (ICT) has enabled the possibility to open some scientific projects to non-academic researchers. The potentialities of distributed research systems are evident (broad geographical coverage, increased human computational power, engagement in science), while the process and the problematics of designing the research activities so to be inclusive and facilitate external participation, are still not exhaustively defined. This short paper, choosing to follow an Action Research method, aims to present the experience of designing a set of activities for public engagement to be integrated in a robotics research project (HeritageBot). Our aim is to discuss the dynamics and the problems emerging while planning citizen science activities through empirical episodes, in order to start understanding both limitations and working heuristics of this approach. We will also discuss how public engagement activities could be integrated into the business model of a research project aiming to bring its results to the market.

**Keywords**—*Citizen science; Open Science; User Engagement; Action Research; Experiments; Research Design; Heritage Bot*

## I. INTRODUCTION

Many scholars are testing innovative research approaches, directly involving nonprofessional scientists into scientific projects. These attempts (known as “citizen science”) have often provided positive results, nevertheless they are still not structured enough to represent a solid and shared base for future initiatives. The lack of well-established theoretical frameworks and detailed operative guidelines was evident in designing citizen science activities for a robotic research project named HeritageBot (HB). Thus, the focus of this short paper is to open and contribute to the debate on the theoretical bases and the methodological alternatives that a scholar designing citizen science activities will face. We will illustrate the advantages and the limitations that we encountered adopting Action Research methodology in our project. Moreover, we will discuss the potentialities of considering participatory activities as integrated part of the project business model. Thanks to the evolution of ICTs, Scientific tasks can now be digitally designed so that virtually everyone can actively participate “*through changes to the processes and scale of creative collaboration itself, enabled by social software such as wikis, online forums and their descendants*”

[1]. However, designing effective open-citizen science activities still represents a challenge in many fields. Identifying the target audience, the common knowledge on a specific research topic and the drivers that could stimulate an active participation is often a lengthy and costly process. At the same time there are several potential risks of integrating public engagement activities in a research project (e.g. uncertainty about intellectual property rights, sensitivity of research issues, complex nature of research activities). Even so, public engagement can provide concrete value to research projects (especially those aiming for the market) as long as the researchers are able to successfully integrate them into the framework of the projects’ BM. Some well-known benefits are: i. the access to a large human computational power (e.g. through the so called *crowdsourcing* initiatives); ii. the possibility of raising a broader awareness and knowledge on specific scientific topics; iii. an early understanding of the market needs and expectations; iv. a large number of users testing early versions of the research outcome. In this short article we introduce the case of a multidisciplinary robotics research project: HeritageBot (HB). Among the goals of the project there is the integration of public’s participation and engagement in its business model. In this short paper, we will first describe the main characteristics of HB project, in order to contextualize our case. Next, we will focus on Action Research methodology: through this framework the researchers can both participate into the scientific activities of the project and at the mean time observe their development. We will explain the reason for choosing such methodology and the advantages that it could bring to this study. In the following section will present a review of the fundamental studies regarding open and citizen science principles, potentialities and limitations. In the discussion and conclusion sections, we will focus on the episodes and experiences matured during the HB project, addressing the processes and challenges of developing public engagement initiatives in scientific research and the dynamics of integrating them into the project’s business model.

## II. THE RESEARCH DOMAIN : HERITAGEBOT

HB<sup>1</sup> is a multidisciplinary research project, currently developing a hybrid multifunctional modular, remotely controlled robotic device. The device is specifically conceived (but not limited) to be used in cultural patrimony preservation and valorization, with particular reference to archeological and architectural applications. The hybrid nature is given by its quadruped-quadcopter locomotion strategy. In particular, the robot is provided with a patented four legs structure along with a quadrotor unit, integrated in the chassis. This configuration provides the device with the capability of facing relatively rough terrain conditions, as well as of flying its way over otherwise unavoidable obstacles. Moreover, the quadcopter unit, provides improved stability and balance when walking on highly irregular ground, thanks to the included magnetometer and gyroscope, and allows low altitude continuous flight. This configuration is particularly useful when inspection of sites that are too dangerous for human operation, or otherwise inaccessible, is needed. In fact, the device is designed to remotely obtain an optimal quantity of visual and metrics data from desired locations. The modularity is given by its upgradable multifunctional structure. The machine is being designed to allow the installation of a multitude of different sensors, tools and multiple locomotion strategies, granting a potentially wide range of applications. The research project is multidisciplinary: together with the engineering team (in charge of designing and assembling the robot) there are researchers from economics, business, architecture, finance, legal and organizational studies. This integration of multiple research fields is motivated by the aim of producing a research outcome with a clear and immediate value for the market. Our research laboratory leads the activities focused on studying and developing an open-citizen science framework to facilitate the integration of external actors in the project business model. While the idea of “opening” the business model to actors not traditionally considered embedded (employed) in a specific organization is well known [2], it is still open the challenge to understand if and how the business model of a research project can be enriched by introducing public engagement initiatives. While the rich literature on BM (starting from well-known works as [3, 4]) has been affected by the potentialities of new technologies (e.g. [5]), there is still a lack of a formal framework for the introduction of public participation strategies in technology intensive, market aimed projects.

## III. THEORETICAL BACKGROUND: ACTION RESEARCH METHODOLOGY

The coordination and alignment of the described open-citizen science principles, with the long term objectives of HB’s BM represents a challenging task. Being our activity mainly focused on studying the introduction of coherent public inclusive solutions in the project’s research workflow, our first

<sup>1</sup> Further details about HeritageBot project can be retrieved at [heritagebot.unicas.it](http://heritagebot.unicas.it)

and crucial effort was to define an appropriate methodology. We started from the assumption that research is a complex activity from both a technical (e.g. amount and variety of needed knowledge, skills required to use research instruments) and a social perspective (e.g. ability to collaborate in team, the process of connecting with the actors needed to plan, fund and execute a project). Our double role as individuals was also taken into account: we act as part of the HB project members, but also as researchers attempting to introduce a change in the internal process of doing research (embedding a set of open-citizen science principles in a traditional research activity). Thus, we have chosen to adopt an Action Research (AR) methodology because, among its main foundations, there is that “a complex social process can be studied best by introducing changes into that process and observing the effects of these changes. Thus, the action research method approaches scientific research from an interventionist’s viewpoint. Researchers both observe and participate in the phenomena under study” [6]. As Blum [7] explains, AR method is funded on two main phases: the *diagnostic stage*, in which the researchers and the subjects of the research collaboratively analyze the problem, and the *therapeutic stage*, where experiments are performed in order to trigger and understand the changes. Susman and Evered [8] provided a more detailed structure of AR based on five iterated phases: diagnosing, action planning, action taking, evaluating and specifying learning. To proceed with the design of our research we, therefore, adopted a hybrid methodological approach, translating the traditional citizen science process [9] into the steps proposed by the action research methodology. AR provides a well-structured and easy to follow framework, fitting the vast open-citizen science landscape, as explained in the following sections.

## IV. CITIZEN SCIENCE LANDSCAPE & BM

The increasing pervasiveness of internet, provides a richer set of communication and collaboration possibilities that academic, governmental and private institutions are exploring as opportunities to build a collective intelligence [10, 11]. In many cases, they are seeking to develop models and tools to trigger and support active public participation in scientific projects<sup>2</sup>. A starting point to better comprehend this phenomenon is the definition of *open science* as the “*umbrella term encompassing the multitude of assumptions about the future of knowledge creation and dissemination*” [12]. Open science focuses on proposing shared principles and offering tools and workflows to promote transparency, reproducibility, dissemination and transfer of new knowledge [13]. However, the connotations of the term have generated debates regarding its benefits, limits, applicability and goals. Part of the

<sup>2</sup> The European research program Horizon 2020 has a section regarding “science with and for society” to allow “all societal actors to work together during the whole research and innovation process”. Another example is the inclusion of Openness goals in the funding convention of CERN. Mozilla has also started its very own open science initiative with Mozilla Science Lab. Other notorious examples are Arduino and Raspberry, having at the core of their competitive advantage the intrinsically open nature of their products.

literature perceives openness as an opportunity to dramatically increase the range of available expertise, thus expanding the spectrum of solvable problems by facilitating creative collaboration [1, 14]. In other instances, openness means democratization of knowledge, assessing that, often, scientific data, methodologies and findings are unequally distributed. This picture raises barriers for: first, generating trust, legitimacy and funding for research projects [15, 16]; second, promoting communication and collaboration between researchers [17, 18, 19], and third, engaging citizens and disseminating knowledge and benefits, derived from research, to the society as a whole [20, 21]. Open science principles are not without criticism, some scholars argue it may take a level of effort and time, to release data, that may surpass the potential benefits, and that the incentives and control over the quality of the produced science may be reduced [22, 23, 24]. The derived concept of Citizen Science encompasses the idea of engaging a wider audience [25]. Regardless of individual expertise on the research, the proposition, in this case, is to successfully combine public engagement, interests and objectives with the scientific objectives of researchers themselves [26]. Moreover, integrating citizens, facilitates awareness and diffusion of the research-produced knowledge, thus promoting scientific literacy [27, 28, 29]. While most of the citizen science literature focuses on environmental studies, in the fields of technology science and robotics, exploiting the interests and curiosity of the people<sup>3</sup>, may represent a promising opportunity to crowdsource data, ideas and solutions [30]. However, as reciprocal benefits for researchers and public are well documented, so do the implied boundaries and practical challenges. Scientific information has to be diluted to overcome the motivational and knowledge gaps between researchers and public, and within the public itself [31, 32]; and goals, methodologies and protocols have to be clearly defined, making the whole process often time consuming and costly [30, 33, 34]. Starting from the works of [3, 4], the previously cited principles, with their benefits and limitations, can be reformulated in a BM perspective. While in the past most BM relied on a closed approach, the increasing costs and complexity of R&D, and the latent potentialities of distributed knowledge are determining the need of a shift towards more openness [35]. But the transition to integrating the open-citizen science formula into market oriented BMs is complicated. By showing pre-production prototypes, or releasing sensible data about the development of innovative products to the public, the risks of information spillovers to competitors may increase. However, the process shares common characteristics with [36] open innovation concept as it may “accelerate internal innovation and expand the market for external use of innovation”. To avoid the risks, while benefiting from distributed knowledge as a long term strategic resource, BMs need to be innovated, in a way that these dangers transform into crucial parts of their value proposition.

<sup>3</sup> A great example of the interest of public in robotics is the success of the Lego Mindstorms, with numerous websites, many instructional books, a large community of amateurs devoted to the product and even parallel projects aiming at increasing the potentialities of these “toys”.

This process coincides with [37] BMI definition as adding “*something new that changes the basis of the BM ... balancing time, costs, value, and learning*” but also raises the strategic dilemma of how to integrate the innovation while maximizing security when opening up [38].

## V. PUBLIC ENGAGEMENT – DESIGNING THE CITIZEN SCIENCE APPROACH

Cutting-edge robotic technologies like HB are often hidden from the common citizen eye [39]. In these conditions, the public faces barriers to, first, correctly identify robots as need fulfilling devices; second, to participate both passively, experiencing first-hand and learning about the benefits of these devices, and actively, by providing suggestions and new ideas, on the development of innovative devices. To overcome said barriers, so to incentivize the collective innovation discovery process, we first need to design a methodological framework that is compatible with HB’s business model. The difficulties of engaging the public in robotics technologies are a product of their complicated and often secretive nature. However, some projects proved that is possible to integrate the participation of external actors in the innovation processes. Inmoov<sup>4</sup> and iCub<sup>5</sup> are examples of successful initiatives in this context. Both projects’ business models are very similar, even if target different audiences. Both fully release their technology and know-how for public fruition, and both generate revenue by selling specific components or pre built devices. Yet, these two examples, while being open in nature, following an open innovation approach, are still expert-oriented, keeping up the barriers for a wider participation. In some other, less operative, cases<sup>6</sup>, participatory engagement initiatives have proved to be an effective scientific literacy driver, promoting bidirectional communication and the divulgation of the research outcomes [30]. HB’s business model is, however, substantially different from the previous examples. Its core value proposition relies on the novelty of the technologies under development, making it not easily compatible with an open approach. The lack of previous analogous research cases imposes the need of designing appropriate public inclusive initiatives from the ground up. In this perspective, we will be testing, as previously stated, a hybrid methodological approach. We start by taking the general citizen science process proposed by [27], and translate them in a series of HB contextualized operative steps, applying action research methodologies as proposed by [8] (table I). In our case, the first step in the process cycle of creating user engagement strategy, is coincident with the *diagnostic* phase of the AR methodology. This phase is instrumental “*to the identification of the primary problems that are the underlying causes of the organization’s desire for change*” [40]. We started by assessing the problem through repeated meetings with all the team members. This initial stage was particularly difficult, mainly because of terminology

<sup>4</sup> Project’s website: <http://inmoov.fr/>

<sup>5</sup> Project’s website: <http://www.icub.org/>

<sup>6</sup> A collection of which can be found in <http://www.roboticslearning.com/>

TABLE I. CITIZEN SCIENCE – ACTION RESEARCH

Methodology		
Citizen Science Bonney et al.	Action Research Susman & Evered	HeritageBot Research Project
Choose a scientific question.	Diagnosis	Meetings – Interviews with research team
Form scientist /evaluator team.	Action Planning	Identification of compatible initiatives
Develop, test, and refine protocols		Pilot Testing
Recruit participants.	Action Taking	Launch initiatives & experiments
Train participants.		
Accept, edit, and display data.	Evaluating	Analysis of the metrics and KPIs
Analyze and interpret data.		
Disseminate results.	Specifying Learning	Tweaking
Measure outcomes.		

and language differences, product of the multidisciplinary of the research and of the complex nature of the device. Nevertheless, thanks to valuable contributions from the other members, we were able to precisely identify and translate the problem into a set of initial goals: I. Promote the awareness on the project. II. Facilitate knowledge dissemination. III. Attract and valorize the distributed knowledge of extra-organizational actors. With the gathered information, transitioning to the *action planning* phase, we started designing and discussing a series of potential initiatives, coherently with the framework emerged in the previous step. Being these activities characterized by different levels of openness, target expertise level and participation in the research process, we summarized them in our *activity matrix* (table II). On the high openness, expertise and participation quadrant we proposed to create an open hardware/software section of HBs website. We received contrasting reactions from the rest of the team, mainly because of the risks of potential undesired information spillovers. A BMI process would be necessary in order to integrate such activity, as the current BM lacks compatibility, and this innovation is not feasible in the short run. On the medium expertise quadrants (high openness – participation) we proposed to partially release blueprints and source of the device, exclusively referred to those sections with unsolved problems. By doing so, we expect to facilitate the crowdsourcing of solutions from the technology enthusiasts’ communities. This carries similar problems as the previous proposed activity. At the medium openness level, the idea was to develop, thanks to web 2.0 tools, a virtual environment where anyone can interact with a 3D replica of the robot, visualizing its capabilities. The application would be also accompanied with a comments section, in order to allow feedback (medium expertise – participation). The problem with such activity is that the development would be extremely costly in terms of budget and time. The last quadrant (low expertise, low openness, low participation) encompasses the widest audience. The initiative regards the publication of project related content through micro blogging and social media platforms [10, 14], to create a cross promotional informative, knowledge dissemination and, to a certain degree, participative infrastructure. The published content will cover all the facets of the team’s research activities, plus news and

TABLE II. EXPERTISE – OPENNESS – PARTICIPATION ACTIVITY MATRIX

Openness Level	Expertise Level			Participation Level
	High	Medium	Low	
High	Open Hardware – Software	Partial Open Hardware – Software	Robotics Workshops	High
Medium	Open Access	Virtual Demo	Demonstrations	Medium
Low	—	—	Micro Blogging – Social Media	Low

related articles from the world of robotics, cultural heritage, startupper and hi-tech industry. The micro blogging - social media approach would be compatible with the previously proposed ideas (that could be progressively implemented in the future), would allow to reach our goals, and has the added value of providing, ex-ante, a huge base of potential users<sup>7</sup>. Moreover, it is fully compatible with the project’s current BM and will provide essential metrics to evaluate our initiatives’ performance, allowing us to continuously improve the strategy (*analysis of the metrics and tweaking* steps). The remaining challenge, in this context, is the creation of publishable content that is not only informative, but also participative. Following the identification of potential activities, the next step will be the *pilot testing* phase. We will be setting up a small scale experiment of the proposed activities with students of a startupper course. This will allow a preliminary evaluation of the public response and engagement, and to gather additional suggestions and ideas. Together with the mentioned online initiatives, we will promote open access to HB-related publications, and testing a set of connected offline workshops and demonstrations to allow the public to experiment in real life the capabilities of HB.

VI. CONCLUSIONS

We addressed the case of introducing public engagement initiative in the business model of a market aimed robotics research project (HB). We found that the current business model of HB, together with the complex nature of the device, challenges the introduction of canonical open – citizen science propositions. We started by analyzing the literature about open and citizen science, and by exploring successful initiatives based on their principles. The lack of well-established theoretical frameworks and detailed operative guidelines, together with the fact that HB’s business model can’t be changed towards the open innovation direction in the short run, we had to formulate a contextualized hybrid methodological approach to identify a series of compatible initiatives. Our initial goals are of generating awareness, disseminating scientific knowledge and of facilitating the public participation in the research, without harming the project’s value proposition. By combining the traditional citizen science process with AR methodology we are being able to address some of the challenges that arose during our

<sup>7</sup> Social media user base <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>



research. Among others, of adequately calibrating the degree of openness, determining the amount of information to be disclosed and identifying appropriate ways to engage the widest audience. By following the proposed methodology, we explored all the facets of the project, analyzed our constraints and developed a series of potential activities. We then formalized these initiatives in an activity matrix, summarizing the different degrees of openness, target degree of user expertise and participation level. As for today we are still at the very early stages of our research, but we find that the identified methodology has the potential of providing satisfactory results for HB.

#### VII. ACKNOWLEDGMENT

This research was supported by a *Lazio Innova* grant to the University of Cassino and Lazio Meridionale for the research project **HeritageBot**.

#### REFERENCES

[1] M. Nielsen, "Doing science in the open", in *Physics World*, 22(05), 30, 2009.

[2] H. Chesborough, "Open business models: How to thrive in the new innovation landscape", Boston: Harvard Business School, 2007.

[3] A. Afuah, & L. C. Tucci, "Internet Business Models and Strategies", Boston: McGraw Hill Irwin, 2003.

[4] A. Osterwalder, Y. Pigneur, & L. C. Tucci, "Clarifying business models: Origins, present, and future of the concept", *Communications of AIS*, 16, 2004, pp.1-25.

[5] P. Lindgren, "NEW Global ICT-based Business Models", The River Publishers Series in Standardisation, 2011.

[6] R. L. Baskerville, "Investigating information systems with action research", *Communications of the AIS*, 2(3es), 4, 1999.

[7] F. Blum, "Action Research – A scientific approach?", *Philosophy of science*, 22, 1955, pp. 1-7.

[8] G. Susman and R. Evered, "An assessment of the scientific merits of action research", *Administrative Science Quarterly*, 23, 582-603, 1978.

[9] R. Bonney et al., "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy," *BioScience*, vol. 59, no. 11, pp. 977–984, Dec. 2009.

[10] J. Surowiecki, "The Wisdom of the Crowds", Anchor Books, New York, 2004.

[11] A. Weiss, The power of collective intelligence, *NetWorker*, 9(3), 16–23, 2005.

[12] B. Fecher & S. Friesike, "Open Science: One Term, Five Schools of Thought" In S. Bartling & S. Friesike (Eds.), *Opening Science*, pp. 17–47, Cham: Springer International Publishing, 2014.

[13] I. Grigorov et al. "FOSTER Open Science Learning Objectives", Zenodo, 2015.

[14] M. Nielsen, "Reinventing discovery: The new era of networked science", Princeton University Press, Princeton, N.J., 2012.

[15] A. Grand, C. Wilkinson, K. Bultitude, & A.F. Winfield, "Open Science A New "Trust Technology"?", *Science Communication*, 34, 2012 .

[16] C. Puschmann, "(Micro)Blogging Science? Notes on Potentials and Constraints of New Forms of Scholarly Communication" *Opening Science*, Part II, 89-107, Springer International Publishing, 2012.

[17] O. Tacke, "Open Science 2.0: how research and education can benefit from open innovation and Web 2.0", In T. J. Bastiaens, 2010.

[18] T. J. Vision, "Open data and the social contract of scientific publishing" *BioScience*, 60(5), 330–331, 2010.

[19] T. Gowers & M. Nielsen, "Massively collaborative mathematics" *Nature*, 461(7266), 879–881, 2009.

[20] A. Whyte and G. Pryor, "Open Science in Practice: Researcher Perspectives and Participation." *International Journal of Digital Curation*, 2011.

[21] Royal Society (Great Britain) and Policy Studies Unit, "Science as an open enterprise", 2012.

[22] J. D. Gezelter, "Open Science and Verifiability", *The Open Science Project*, 5, 2011.

[23] S. Friesike, & T. Schildhauer, "Open science: many good resolutions, very few incentives, yet", In *Incentives and Performance*, Springer, pp. 277–289, 2015.

[24] G. Goldman, "Cautiously Open to Open Science - The Equation", 2013, retrieved from <http://blog.ucscusa.org/gretchen-goldman/cautiously-open-to-open-science-138>

[25] R. Bonney, H. Ballard, R. Jordan, E. McCallie, T. Phillips, J. Shirk, & C. C. Wilderman, "Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education", A CAISE Inquiry Group Report, 2009.

[26] J. Silvertown, "A new dawn for citizen science", *Trends in Ecology & Evolution*, 24(9), 467–471, 2009.

[27] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, & J. Shirk, "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy" *BioScience*, 59(11), 977–984, 2009.

[28] D. Brossard, B. Lewenstein, & R. Bonney, "Scientific knowledge and attitude change: The impact of a citizen science project", *International Journal of Science Education*, 27(9), 1099–1121, 2005.

[29] M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, J. Vandenberg, (2009) "Galaxy zoo: Exploring the motivations of citizen science volunteers", arXiv Preprint, 2009.

[30] C. Wilkinson, K. Bultitude, & E. Dawson, "Oh Yes, Robots! People Like Robots; the Robot People Should do Something": Perspectives and Prospects in Public Engagement With Robotics, *Science Communication*, 33(3), 367–397, 2011.

[31] M. Ebner, & H. Maurer, "Can Weblogs and Microblogs Change Traditional Scientific Writing?", *Future Internet*, 1(1), 47–58, 2009

[32] M. C. Powell, & M. Colin, "Participatory Paradoxes: Facilitating Citizen Engagement in Science and Technology From the Top-Down?" *Bulletin of Science, Technology & Society*, 29(4), 325–342, 2009.

[33] M. J. Raddick, G. Bracey, K. Carney, G. Gyuk, K. Borne, J. Wallin, A. Planetarium, "Citizen science: status and research directions for the coming decade", *AGB Stars and Related Phenomena 2010: The Astronomy and Astrophysics Decadal Survey*, 2010, 46P.

[34] R. Bonney, H. Ballard, R. Jordan, E. McCallie, T. Phillips, T., J. Shirk, & C. C. Wilderman, "Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education", A CAISE Inquiry Group Report, Online Submission, 2009.

[35] H. Chesbrough, "The Era of Open Innovation", *MIT Sloan Management Review*, 35-41, 2003.

[36] H. Chesbrough, "Open Innovation: A New Paradigm for Understanding Industrial Innovation", in Chesbrough, Vanhaverbeke and West, *Open Innovation: Researching a New Paradigm*, 2006

[37] P. Lindgren P. and M. A. Abdullah, "Conceptualizing strategic business model innovation leadership for business survival and business model innovation excellence", *Journal of multi business model innovation and technology*, vol 1, pp. 115-134, River Publishers, 2013.

[38] P. Lindgren and R. Jorgensen, "Towards a Multi Business Model Innovation Model. Journal of multi business model innovation and technology, 1-22. River Publishers, 2012.

[39] Special Eurobarometer 382 (2012). Retrieved on [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_382\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_382_en.pdf)

[40] R. L. Baskerville, "Investigating information systems with action research," *Communications of the AIS*, vol. 2, no. 3es, p. 4, 1999.

# Sustainable Business Models for the Adoption of Energy Management Cloud Platforms within Enterprises

George Suciuc, Octavian Fratu, Carmen Voicu  
Telecommunication Department  
University POLITEHNICA of Bucharest  
Bucharest, Romania  
george@beia.ro

Lucian Necula, Laura Ghenciu, Victor Suciuc  
R&D Department  
BEIA Consult International  
Bucharest, Romania

**Abstract**—Nowadays, in order to adapt to increasingly dynamic market changes and ensure market competitiveness while enhancing energy efficiency, comfort and security in working environments, enterprises concern themselves with the adoption of novel business models, technologies and methodologies. The main purpose of this paper is to present author's conceptual model of a Cloud energy management system and to propose adaptations of EPC business models which can accelerate the market penetration of similar solutions. This paper provides an overview on the benefits and downsides that EPC (energy performance contracting) can provide to the business sector and financial institutions. It also presents marketing aspects associated with energy management solutions that aim to enhance energetic performance, comfort and security within enterprises from various domains.

**Keywords**—*Business Model, Energy Performance Contracting, Energy Management Cloud Platform, ESCO*

## I. INTRODUCTION

Within enterprises of all sizes and activity fields, monitoring the energy demand became one of the most important decision-making tools since companies use consumption data to stay competitive without neglecting the environment [1].

In the context of highly demanding environmental policies and growing energy costs, many enterprises tried to find more convenient financing sources for their investments and to adapt their business models in order gain competitive advantages [2].

Based on this, this paper briefly presents authors' conceptual model of an energy management Cloud platform and proposes business models based on "Shared Savings", "Guaranteed Savings" and „Chauffage" energy performance contracting [3]. These business models come to overcome existing barriers caused by the lack of financing options and technical expertise in adopting energy efficiency measures.

The rest of the paper is organized as follows: Section II presents the principles of energy performance contracting, analyzing functional and marketing aspects of energy management solutions, Section III describes authors' conceptual model of a Cloud energy management platform,

Section IV proposes the business models which can be used to enhance the market penetration of solutions similar to the proposed one, while Section V concludes the paper.

## II. RELATED WORK

In this section, we present funding methods used within EPCs, business models associated with energy performance contracting and energy management systems available on the global market.

### A) *Funding and models of energy performance contracting*

When energy performance contracting first came to the market, ESCO (Energy Service Company) entities were performing both technical and financing activities since financial institutions were reticent when it came to EPC projects. Nowadays, ESCOs do no need to struggle with project financing, because there is a competitive marketplace with financial institutions willing to provide support for new investments.

Financing can either be ensured by the ESCO itself or by the customer, ESCO being responsible for providing a savings guarantee. To finance the investment, ESCOs or their customers can either use personal capital or involve a third party as a credit source [4].

Energy performance contracting can rely on various models, "Shared Savings" and "Guaranteed Savings" being the most common [5].

A "Shared Savings" model implies that the energy savings are shared between the ESCO and the customer in accordance with a pre-arranged percentage over the term of the contract. The ESCO uses the savings to repay the loan to the financing institution and, in return, guarantees a certain level of energy efficiency. Since the client takes over some of the performance risk, it is likely that it won't be willing to assume any credit risk. This contract model usually involves financing from a third-party institution, ESCO being responsible for repaying the loan and taking over the credit risk [6]. This model provides a great advantage to the customer in developing markets since the customer doesn't support any financial risk. According to Bertoldi et al. [7], the model tends

to create barriers for small ESCO companies in obtaining financing.

Within a “Shared Savings” business model, the compensation is achieved through fixed fees, unit pricing and/or gainsharing.

Under the “Guaranteed Savings” performance contract, the ESCO guarantees a certain level of energy savings in a way that exempts the customer of any performance risk. Even if the ESCO arranges the funding, the client is directly financed by a financial institution and is responsible for the repayment of the loan. According to Parviainen [8], this model is less appropriate for economies where the ESCO concept is relatively new, due to the fact that customers are usually not willing to support the financing and the associated risks and due to the lack of familiarity with project financing for energy efficiency measures. If the savings are lower than guaranteed, the ESCO supports the payment of the difference between what was promised and what was achieved. In some cases, such a scenario would lead to additional penalties for the ESCO.

In Europe, most of the ESCO projects were undertaken in the public sector since public institutions are usually perceived as a client who is unlikely to go out of business. Most implemented ESCO projects targeted areas such as energy management, co-generation, public lighting, and heating, cooling and ventilation in public industrial facilities, commercial centers and hospitals.

The main instruments and methodologies used for energy performance contracting are the regulation for certification of the solution provider that implements energy performance contracting and the European Code of Conduct for Energy Performance Contracting (CC – CPE).

### *B) Energy Management Systems*

At a global level, many companies turned their focus on the on-growing demand for energy management to reduce costs, pollutant emissions, and to improve security and comfort within home, public and work buildings. Several energy management solutions are already available on the market for both home users and utilities. In this section, we briefly analyze three of the main energy management solutions, focusing on their business model and market strategy.

Cisco Energy Management Suite [9] is a Cloud solution that provides energy consumption measurement, monitoring and management features for each IP device connected in a network, independent of the vendor. The solution is available in four different modules, each adapted to specific needs:

- Cisco Energy Management for Distributed Offices – this module provides energy usage data visualization and analytics for physical IP connected devices;
- Cisco Energy Management for Data Center – in addition to the module dedicated to distributed offices, this module provides network-based monitoring of all connected physical and virtual devices and systems;

- Cisco Energy Management Optimization Service – this module is based on an annual subscription and provides optimization of energy usage within client's IT infrastructure. By the means of a proprietary software, namely Cisco Energy Management Software, it provides centralized access for network devices;
- Cisco Energy Management as a Service – this is a SaaS (software-as-a-service) application which allows the management of energy consumption without requiring any hosting of the proprietary application in client's data center. Basically, since the solution is not hosted by the customer, it does not imply any costly server administration. Clients can have access to this solution by the means of one-, three- or five-year subscription.

Additionally, Cisco offers a free start trial with a 45-day license for access to trial versions of the software and limited-function monitoring and reporting features for up to several hundred of connected devices.

DEXCell Energy Manager [10] is a web-based energy management software developed by DEXMA, which provides real time monitoring and analysis, utility bill tracking, user friendly dashboards, energy patterns and energy grader based on Big Data algorithms, benchmarks. The platform supports integration of different protocols and devices from several manufacturers.

Within their current business model, DEXMA offers the DEXCell Energy Manager as SaaS (Software-as-a-Service), on an annual subscription basis, with two standard packets – “Starter” and “Professional”, and a custom packet dedicated to more complex needs – “Enterprise”. Additionally, they also provide a demo account with 30-day free access to three examples of fully functioning projects, including consulting by phone with an energy specialist. Some details for each subscription packet are presented as follows:

- Starter – this is the basic packet, it provides features for single site projects, including dashboards, simple analysis, notifications, reports. This offer features one location and ten data points and it also includes access to a collection of external energy efficiency applications built on DEXMA APIs, collection known as “Energy Apps Market”;
- Professional – this packet provides all the features from the Starter packet, plus more advanced data analysis, alerts and reports. Also, this offer features unlimited locations and 20 data points;
- Enterprise – this offer features unlimited locations and data points and provides all features from the Professional packet, addressing to customers with large and complex projects. The subscription is fully customizable, implying that the product's features and price offer are set upon request.

All subscriptions, except for the Enterprise offer, are provided on an annual fixed cost and allow for the addition of more data points upon request, at a fixed monthly cost.

Engage [11] is an energy management solution developed by Efergy, which provides tools for real-time visualization and statistics of data related to energy usage and energy costs, aiming to improve customers' energy usage habits. Integration of Internet of Things (IoT) and time-critical cloud applications represent the main challenge [12, 13]. The solution is available in the form of software (Engage platform and app) and hardware kits with different pricing schemes, that include access to the web platform and the additional devices needed, according to the scope and features of each kit. Some of the main offers in the actual business model are described below:

- Engage Hub Kit – provides features and tools for monitoring home energy using devices such as computer, smartphone, tablet. This kit includes the web-based software and a hub device which is to be installed in the home to gather energy data;
- Engage E2 Hub Kit – provides the standard Engage Hub Kit features with an additional software – elink, which enables users to download the data from the energy management platform, features more advanced data visualization and simulation of different consumption patterns and tariff schemes. The kit includes the web-based software platform, the elink software, one sensor, one data transmitter, one hub and one wireless home data visualization device;
- Engage Solar Kit – this solution provides the standard Engage Hub Kit monitoring features for both energy generation and consumption. The kit includes the software solution, two electricity sensors with two data transmitters and a data hub device.

### III. CLOUD ENERGY MANAGEMENT PLATFORM

This section aims to present authors' conceptual model of a Cloud Energy Management platform that aims to provide enterprises with the means to manage energy consumption while enhancing comfort and security in the working environment.

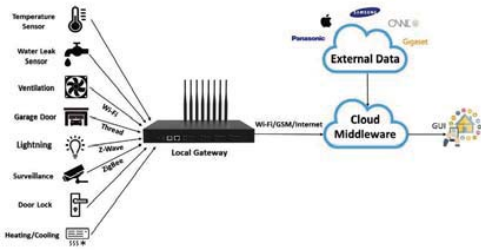


Fig. 1. Energy-aware resource monitoring and management platform

As depicted in Fig. 1, the proposed platform aims to:

- Monitor air quality (temperature, relative humidity, air pressure, pollution, particles, etc.), visual comfort, acoustic and energy demand parameters in real time

by using a wide variety of sensors deployed in the working environment;

- Generate energy consumption awareness by providing data regarding energy consumption in various forms and by pointing out the sources of inefficient energy usage (technologies, processes, habits, etc.);
- Use machine learning to detect and learn movement patterns and typical interactions, thus combining human with software based control to ensure automation of decisions taken to enhance comfort, security and energy efficiency;
- Import data from third party Clouds of commercial energy management solution providers, such as Apple, OWL, Gigaset, Panasonic, Samsung, etc., by the means of a Cloud Middleware;
- Provide support for a wide variety of sensors and appliances from various Smart Home solution providers by the integration and convergence of different communication systems and protocols through a Gateway;
- Allow integration within the existing electric network infrastructure;
- Provide the user with a web application to remotely control and survey the building.

### IV. PROPOSED SUSTAINABLE BUSINESS MODELS BASED ON ENERGY PERFORMANCE CONTRACTING

The development of adaptation mechanisms based on new business models and concepts is an important item on the agenda of companies that want to play a major role on the market.

The proposed business models for the platform market deployment are based on EPC (Energy Performance Contracting) and imply that the customers are provided with the solution (Cloud energy management platform) as an energy management measure and also with the guarantee that the savings generated by its implementation will be sufficient to finance the full cost of the investment.

Fig. 2 illustrates the principles Energy Performance Contracting relies on.

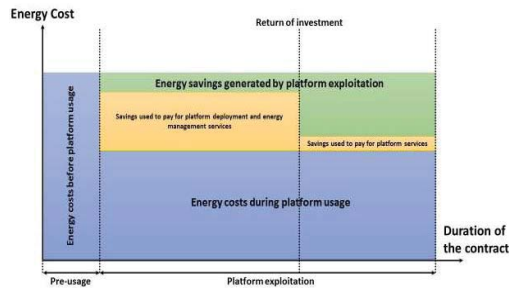


Fig. 2. Energy Performance Contracting

Therefore, the key points of EPCs are:

- The energy management platform provider ensures a proper deployment and exploitation of the platform during the whole contract duration;
- The energy management platform provider needs to ensure energy cost savings compared to the pre-implementation energy cost baseline;
- The cash flow generated by the means of savings achieved during the exploitation of the platform are used to pay back the investment in the platform implementation;
- Until the full return of investment for the platform deployment within customer’s facility, the energy savings are split between the platform provider and the beneficiary;
- After the full return of investment, the customer only pays the platform provider a fee associated with the provided platform services;
- The relationship between the entity providing the platform and the customer is a long-term, fair and transparent one;
- All the steps within the EPC contract are done in a legal matter.

Even though most of the energy performance contracts are financed by long term loans, some customers would be able to pay a part of the platform implementation cost with capital budget allocations.

The amount of energy saved can be verified using one of these methods:

- Deemed or stipulated savings – The customer pays the energy management platform provider estimates which were agreed upon before signing the contract;
- Savings based on utility bills – This method is based on a comparison between the baseline consumption determined by past energy bills and the energy consumption measured before adopting any energy efficiency actions;
- Measured savings – This method is the most expensive to use yet the most exact in determining savings. It involves consumption measurements both before and after platform implementation in customer’s facility. The method also relies on adjustments based on weather conditions (specific to a certain time of the year or climate), changes in the use of customer’s facility, equipment loads which might vary (as a response to the market demand), etc.

A) „Guaranteed Savings” performance contracts

As depicted in Fig. 3, a “Guaranteed Savings” performance contract between the platform provider and the customer implies that contractor (platform provider) supports payment cuts or contract-stipulated penalties if the savings are lower than guaranteed.

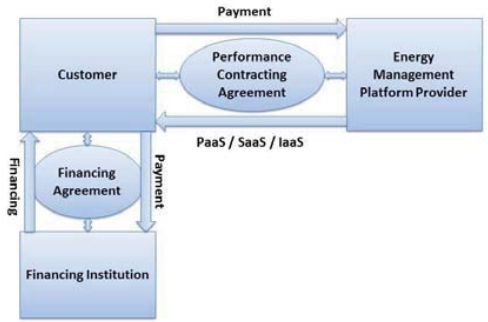


Fig. 3. Energy Performance Contract based on “Guaranteed Savings”

Under such a performance contracting agreement, the Energy Management Platform provider guarantees a certain level of energy savings, thus the client is usually absolved of any performance risk. Following this agreement, the energy management platform provider usually takes over the entire performance risk. Customers are financed directly by banks or other financing entities.

If the savings are not enough to ensure a full return of the investment for the customer, the energy management platform provider has to cover the difference. If the savings exceed the guaranteed level, the customer pays a percent of the energy savings generated by platform’s exploitation to the provider.

B) „Shared Savings” performance contracts

As depicted in Fig. 4, a “Shared Savings” performance contract between the platform provider and the customer implies that the customer makes pre-fixed value or percentage payments from the savings to the platform provider until a predetermined amount or time period is achieved.

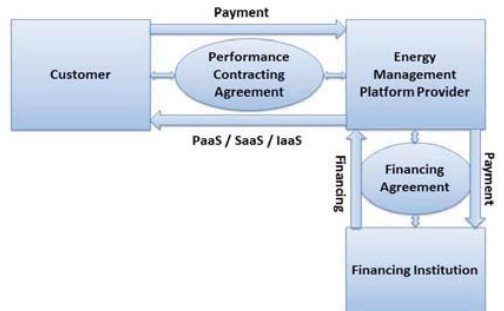


Fig. 4. Energy Performance Contract based on “Shared Savings”

Under a shared savings performance agreement, the customer takes over a pre-determined part of the performance risk. Thus, within this business model, the energy management platform provider assumes both the performance and the customer credit risk.

C) „Chauffage” performance contracts

As presented in Fig. 5, a “Chauffage” contract implies that the energy management platform provider would take responsibility for the provision of platform services and for the purchasing of energy. The energy management platform provider also takes responsibility for providing energy resources at a lower price or for providing a higher level of energy service for the same price. As in the case of “Guaranteed savings” and “Shared savings” performance contracts, the “Chauffage” contract relies on savings which are used by the customer to repay platform deployment and usage costs and also to generate a return of capital. In this scenario, the “Chauffage” contract allows the customer to outsource facility services and investment.

and services, energy performance contracts are to become an important mean to support the implementation of energy management platforms within enterprises.

As a future work, we intend to present insights on the achieved results after a full development of the conceptual platform, followed by its commercialization in accordance to the proposed business models based on energy performance contracting.

ACKNOWLEDGMENT

The work has been supported in part by UEFISCDI Romania under grants no. 20/2012 “Scalable Radio Transceiver for Instrumental Wireless Sensor Networks - SaRaT-IWSN”, Power2SME, grant no. 262EU/2013 „eWALL” support project, grant no. 337E/2014 “Accelerate” project, through the project Power2SME and by European Commission by FP7 IP project no. 610658/2013 “eWALL for Active Long Living - eWALL”.

REFERENCES

- [1] M. Schulze, et al., “Energy management in industry—a systematic review of previous findings and an integrative conceptual framework”, *Journal of Cleaner Production*, vol. 112, pp. 3692-3708, 2016.
- [2] P. Kumar, C. Martani, L. Morawska, L. Norford, R. Choudhary, M. Bell, and M. Leach, “Indoor air quality and energy management through real-time sensing in commercial buildings”, *Energy and Buildings*, vol. 111, pp. 145-153, 2016.
- [3] P. Bertoldi, S. Rezessy, and E. Vine, “Energy service companies in European countries: Current status and a strategy to foster their development.” *Energy Policy* 34.14, pp. 1818-1832, 2006.
- [4] E. Stuart, P. H. Larsen, J. P. Carvallo, C. A. Goldman, and D. Gilligan, “US Energy Service Company (ESCO) Industry: Recent Market Trends.” 2016.
- [5] M. Halme, Minna, et al. “Business models for material efficiency services: Conceptualization and application.” *Ecological Economics* vol. 63, no. 1, pp. 126-137, 2007.
- [6] P. Lee, P. T. I. Lam, and W. L. Lee, “Risks in Energy Performance Contracting (EPC) projects.” *Energy and Buildings*, vol. 92, pp. 116-127, 2015.
- [7] D. Kindström, and M. Ottosson, “Local and regional energy companies offering energy services: Key activities and implications for the business model.” *Applied Energy*, vol. 171, pp. 491-500, 2016.
- [8] J. Parviainen, “Finance solutions for energy efficiency services and their applicability in materials efficiency business” Master’s thesis at Helsinki School of Economics, 2004.
- [9] S. H. Kim, “Energy profiling for demand side management using the CISCO EnergyWise product.”, UBC Social Ecological Economic Development Studies (SEEDS), pp. 1-23, 2015.
- [10] Dexmatechcom, DEXMA. Retrieved 12 September, 2016, from <http://www.dexmatech.com/software/>
- [11] Efergycom. Retrieved 13 September, 2016, from <https://engage.efergy.com/>
- [12] G. Suciú, S. Halunga, A. Vulpe, and V. Suciú, “Generic platform for IoT and cloud computing interoperability study.” *IEEE Signals, Circuits and Systems (ISSCS)*, International Symposium on, pp. 1-4, 2013.
- [13] Z. Zhao, et al., “Developing and operating time critical applications in clouds: the state of the art and the SWITCH approach.” *Procedia Computer Science*, vol. 31, no. 68, pp. 17-28, 2015.

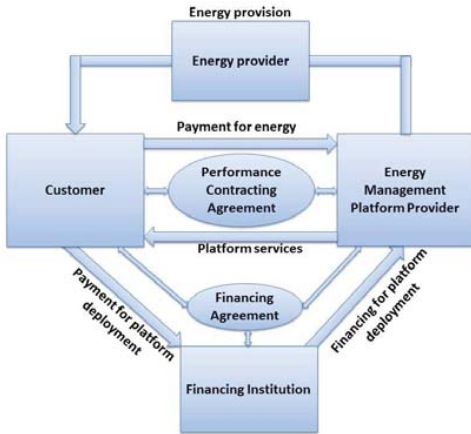


Fig. 5. “Chauffage” Energy Performance Contract

In terms of risks, the energy management platform provider is responsible both for the energy supply and demand efficiencies and for ensuring at least the guaranteed level of savings generated by the usage of the energy management platform. The credit risk is supported by the customer who benefits both from platform services and energy resources.

V. CONCLUSIONS

This paper focuses on the go-to-market strategy of energy management platform providers by proposing adaptations of EPC schemes that present a great potential for enhancing market penetration of these solutions. In the context of an increasingly competitive business sector, energy management systems became an important decision-making tool for companies of all sizes and activity domains.

Given a wide range of financing opportunities that address energy usage optimization and the provision of green products

**Data Driven Business Models definitions and barriers in a European Electronics Manufacturing Company.**

**- A Preliminary Case study**

---

Torben Caesar Bisgaard Bjerrum

Department of Business and Technology  
Aarhus University,  
Birk Centerpark 15,  
DK-7400 Herning  
Denmark  
E-mail:torbenb@btech.au.dk

**Introduction**

The next industrial revolution is in the making, with data as the hot topic (cPP, 2014). This indirectly also affect the tangible products, that are being supplemented or changed with intangible services, and smart tangible products, that rely on data. This development, has with new technology such as big data, Internet of things (IOT) accelerated and is called the fourth industrial revolution (K. Schwap,2016). This is the new reality to many businesses, and the theory regarding service development as an added value proposition is written, almost a decade ago (Bains et al, 2007). This is a challenge because the customers are increasing their demands and creating a pull effect, that the businesses not are ready for. These technologies need to be adopted by all business types, in order to maintain competitive advantages and avoid the commodity trap. However, the DDBM will not reach mass impact on industries before the digitalization of the different businesses, because the value potential cannot be extracted in some ecosystems or internally in the business without DDBM capabilities being present at all levels (Neely et al, 2014). The challenge is also that these new business models has shorter life span and needs continuously monitoring and updates along with the technology (Lindgren & Rasmussen, 2013).

In relation to the business model innovation (BMI), the impact of these new DDBM will relate directly on businesses' ability to adopt to the digitalization of the current and new business models, not only for the business themselves, but also for network partners and the business customers (K. Schwap, 2016 ).

The ability of the businesses model ecosystem (BMES), being able to adopt these new business model, are also dependant on the technology used in the ecosystem and the collaboration that comes into play in the BMI process internally in the businesses. Therefore, the strategic coordination of these interdisciplinary capabilities, needs to evolve with the adoption of the digitalization of the business, the dataflow and the innovation related to the new business models (Brownlow et al, 2015).

**Purpose**

This paper describes the most commonly used DDBM from the literature, and addresses the barriers that exist when traditional businesses moves from traditional product oriented logic to new service oriented logic. This preliminary case study aims also describe the most common barriers and the initial work within the organisation with DDBM and the attempt to develop these with the BMI process in a business. The barriers that are addressed are of both organisational- and technical character in relation to the development of the DDBM.

The paper also relates to the BMI processes in terms of competences, management and describing the knowledge gathering and sharing being an interdisciplinary discipline within an organization, as an attempt to make faster and more resilient BM.

This study will use the term data, with the intent to cover all the aspects of data, used in the literature (Big Data, Small data, Smart Data etc.), in an attempt to cover the data issue in a broader sense, in to-be business models, in relation to both innovation and DDBM to show, where the DDBM has eligibility in future new business models and how business are capable of developing DDBM.

**Methodology**

The research method in this article is a combination of literature studies and participatory action research (PAR). The eight months in the EEM is considered a cross-sectional study with primarily qualitative data collected. The data has been collected as primary, secondary and tertiary sources. The interview has been both unstructured, and semi-structured interviews.

The definition of action research in this article is related to the definition by (Bryman and Bell, 2011), where the researcher and the client collaborate on a solution based in the observation and findings they agree on, where the researcher was involved on management level on a daily basis at the business.

The results of the study relate to the idealist approach with a shared reality among participants of the environment. This shared reality is almost instantly tested and verified. This creates a bottom up approach that forms an abductive strategy, where underlying issues are not seen by first glance, but revealed through dialog and shared understanding, as the data thickens and insight are revealed. This approach is in line with the AR method.

**Business models**

Business models describes how the business creates, delivers and capture value in the business (Osterwalder et al, 2005). The business model is a construct around the concept of value proposition and how this value proposition is created, delivered and captured in the overall strategy of the business. The Business model is not the strategy itself (Gahzzi, 2014).

Despite more than a decade of research in the business model topic, researcher have different names and views on the different dimensions in the business model. The most popular models used are Osterwalder and Pygneur (2010) “business model canvas”; Zott and Amit (2010) “activity system perspective”; Johnson et al’ (2008) “reinvented business model”; and Morris et al’ (2005) “entrepreneur’s business model”. But others less known frameworks for business models exists.

The businesses have more than one business model (Lindgren et al, 2013), and the business model framework that adopts this mind set is Lindgren and Rasmussen (2013) “business model cube”. The multi business model framework is used in the PAR at the EEM in this study.

In the BM Cube framework, a BM consists of seven dimensions. Each of the dimensions are based on research that are described in the literature. A short description is presented in the table below

Dimension	Description
Value Proposition	Value proposition is the value the BM is offering. These consist of three elements <ul style="list-style-type: none"> <li>- Products</li> <li>- Services</li> <li>- Processes</li> </ul>
Customers and Users	The users and receivers of the BM. These are B2B , B2C or B2G. <ul style="list-style-type: none"> <li>- Customers pay monetary for the value</li> <li>- Users <i>do not</i> pay monetary for the value</li> </ul>
Value Chain	Value chain are the functions that are needed to make the BM work. <ul style="list-style-type: none"> <li>- Primary functions describe the value creating processes and functions from inbound logistics to outbound logistics and service.</li> <li>- Secondary functions describes the value creating functions needed to support the primary functions</li> </ul>
Competences	Describes the resources in the business divided in four main groups <ul style="list-style-type: none"> <li>- Technology</li> <li>- Human Resources</li> <li>- Organization</li> <li>- Culture</li> </ul>
Networks	All businesses consist of a network that can be of the three following <ul style="list-style-type: none"> <li>- Physical</li> <li>- Digital</li> <li>- Virtual</li> </ul>
Value formula	Calculates the value of the BM. This is done traditionally in relation to monetary values <ul style="list-style-type: none"> <li>- Value = Revenue-Cost</li> </ul> and/or as <ul style="list-style-type: none"> <li>- Non-monetary values / other values</li> </ul>



Relations	<p>Describes the relation between the other six dimensions in the BM or in relation to other BMs and energizes the BMs. The purpose is to carry the values and enables the BM to capture, deliver and receive values. Relations is essential for a BM to work.</p> <p>Relations can be either</p> <ul style="list-style-type: none"> <li>- Tangible</li> </ul> <p>and/or</p> <ul style="list-style-type: none"> <li>- Intangible</li> </ul>
-----------	---

**Data Driven Business Models and data**

Data is gathered in businesses, with the purpose to gain business intelligence from systematic analysis and statistics (Hartmann et al 2014; Neely et al 2015) This offering of value extracted from data is used both internally and externally in the businesses (Porter & Heppelmann, 2014).

The description and characteristics of big data being initially the three V’s, which over time has expanded to even more V’s, in order to describe the main aspects of the data science as the field are being developed, in symbiosis with the technology development in general. This early state description when first described, indicates continuous development of the technology and the business models.

Patterns in capitalizing on data are many and increasing with the advances in technology. Recent literature has proposals for the most stereotypic business models and usage of data according to the current technology and ecosystems (Parmar et al, 2014).

The summarized patterns from (Parmar et al, 2014) presented in short form:

<b>Pattern/BM</b>	<b>Technology</b>	<b>Utilization</b>
<ul style="list-style-type: none"> <li>- Augmenting Products to Generate Data</li> </ul>	<ul style="list-style-type: none"> <li>- Sensors (IOT)</li> <li>- Wireless Communication</li> </ul>	<p>B2B, B2C markets. Improves service, repair and design. Develops new service business models</p> <p>Ex.</p> <ul style="list-style-type: none"> <li>- Engines</li> <li>- Wind Turbines</li> <li>- Kitchen appliances</li> <li>- Intelligent bearings</li> <li>- Progressive insurance</li> </ul>
<ul style="list-style-type: none"> <li>- Digitizing Assets</li> </ul>	<ul style="list-style-type: none"> <li>- Digitization</li> </ul>	<p>B2B, B2C markets. New platforms for manuals, entertainment, education. Enhances existing business models or create new business models.</p> <p>Ex.</p> <ul style="list-style-type: none"> <li>- Music</li> <li>- Movies</li> <li>- Manuals</li> <li>- Books</li> <li>- Museums</li> <li>- Health records</li> <li>- E-mail</li> </ul>
<ul style="list-style-type: none"> <li>- Combining Data Within and Across Industries</li> </ul>	<ul style="list-style-type: none"> <li>- ICT standards</li> <li>- Sensors (IOT)</li> <li>- Wireless Communication</li> <li>- Data integration</li> <li>- Big Data Capabilities</li> <li>- Cloud- / fog- computing</li> </ul>	<p>B2B, B2G. Coordination across information across Industries, Ecosystems, Sectors. Create new business models</p> <p>Ex.</p> <ul style="list-style-type: none"> <li>- Social services</li> <li>- Health</li> <li>- Infrastructure optimization</li> </ul>

		<ul style="list-style-type: none"> <li>- Property Maintenance</li> <li>- Supply chain mgmt.</li> <li>- Fraud detection</li> </ul>
<ul style="list-style-type: none"> <li>- Trading Data</li> </ul>	<ul style="list-style-type: none"> <li>- ICT</li> <li>- Big Data</li> <li>- Analytics</li> </ul>	<p>B2B, B2C. New business models in new eco systems to the business</p> <p>Ex.</p> <ul style="list-style-type: none"> <li>- Satellite navigation</li> <li>- Weather data in insurance</li> <li>- Internet usage</li> <li>- User behavior</li> <li>-</li> </ul>
<ul style="list-style-type: none"> <li>- Codifying a Distinctive Service Capability</li> </ul>	<ul style="list-style-type: none"> <li>- ICT</li> <li>- Cloud computing</li> <li>- Software</li> </ul>	<p>B2B. New Business models in new ecosystems to the business.</p> <p>Ex.</p> <ul style="list-style-type: none"> <li>- Automated Best in class primary value chain functions</li> <li>- Automated Best in Class business processes</li> </ul> <p>For:</p> <ul style="list-style-type: none"> <li>- Travel processes</li> <li>- Internal audit processes</li> <li>- Modeling internal resources flow</li> </ul>

**Brief introduction to the EEM**

The business is an old electronics manufacturer that has evolved from the agricultural business, into modern electronics in the B2B market, within several different industries. Turnover 2014 in the range of 75mill € The business has 600 employees, where approximately 250 are in the HQ and the rest are located in 11 subsidiaries around the world, functioning as own business units.

The EEM serves a conservative market, with little or no interest in changing technology. This is partly due to the rules and regulations of the market. New technology requires for both the EEM and the customer certifications being developed on a local or national basis. The customer base is approx. 6000, with the Pareto rule in force, where 20 percent of the customers generates 80 percent of the revenue, and 20 percent of the customers are using 80 percent of the business available resources. The business is well known in the market and branded on the high quality and long durability of its products. The last decade the electronics the EEM delivers has been moving from a niche market towards a commodity market, pressed on price and quality.

The challenge for the business is that through a project, the EEM by accident stumbled across a delivery that involved service. The service term in the business has previously been repairs on warranty and service in the traditional sense. The transformation from a traditional product oriented logic to a service oriented logic is the main issue for the business. This new business model (BM) also challenges some of its customers that will change to competitors.

**Discussion:**

**The EEM Business models and development barriers**

The EEM have experienced new business opportunities, that in order to be realized, needs to be implemented into a well proven and old organization. This induces challenges that are already known in the literature.

Development of new business models that utilizes new technology in the form of data and differentiates the business from the competition has not been the main focus for neither the R&D- or the Innovation department. The focus for these departments has been on developing new business models with internal focus. This have involved implementing technology that would reduce the fixed costs, performance and to some extend quality but also to test trending technologies.

The physical product development platform and understanding of customers, relies heavily on the persons in the business, that do not have customer contact. This silo thinking, in a business where 'over the wall engineering' is practiced, needs in own words, a clear strategy regarding the DDBM initialization. This is an issue in the organization,

with the new intangible services that relies heavily on the end user analysis, and knowledge sharing in the development process.

The new BM(s) with service and data changes the known, and well proven business model, into a new business model, where EEM has to take on a new role as a service and software supplier, that delivers new value to the customer. This changes the both the competences required in the organization. This could also be moved into the networks, which again relates to new competences both in organization and culture, hence communication.

The current business models for the EEM, is that it supplies to systems builders, that uses the business products as the whole end-product or as part of the end-product. The new service business model that the business wants to focus on, has this moved from delivering a value proposition, that was contained in a piece of electronic hardware, to a value propositions that relates to a deeper insight into the customer usage of the hardware and the entire lifecycle of the hardware. The analysis done previously on the product delivered had been limited to complaints on durability of the physical product, or size and cost. In the new DDBM, this also involve the changing requirements of the end customer in both the hardware, software and analysis, with shorter respond times in all aspects of the production.

In the EEM there is currently technology in the products that allows for data gathering. More technology in form of sensors, needs to be incorporated to reach the goal of producing the required service models. The existing technology is not being utilized and in the organization and there is no overview of what the current platform is capable of handling. Issues regarding the willingness of customers to deliver data has been an issue because the customers do not know what the technology comprise of, the value of introducing the technology and the quality of the data extracted. This in turn will mean a massive investment in a data platform cable of both gather data and doing the analysis of this data hence securing the correct data quality.

#### **The EEM DDBM innovation barriers**

The EEM have within the last two years before the study, seen an issue regarding the ability to spot trend and test technology. Therefore it made a small innovation department, with direct relations to the marketing and sales department. The innovation department is manned with engineers that have been with the business several years. This department deals with both technical innovation and trends, but also has business understanding and development in broader terms.

Due to the fact that the innovation department, was new in the EEM, and the new business opportunity initially was confronted only from the business side, and not from the traditionally engineering perspective, gave transparency to the barriers in the organization. The EEM discovered that different initiatives related towards data collection and usage, was being tried in other departments in the organization.

The innovation department saw good usage of the analytics used in the marketing department. SWOT analysis, sensitivity analysis and market share estimations came from this part of the EEM, due to the close relations with management. But a structured approach in the EEM relating to data, data handling, analytics and service design was not present.

The data driven business model selected was a two-step plan with 1) the 'Augmenting Products to Generate Data' and then combining it with 2) 'Combining Data Within and Across Industries'. This capability would lead to a combination of a product and service platform, that would deliver predictive maintenance for the EEM costumers. The combination of data handling and analytics, is a new competence required for the initialization of the business model, and also barrier in the understanding of requirements in the development phase. Understanding the handling of data in relation to data handling and legislation regarding ownership and the right to use data, and the requirement of customers' discretion was also a barrier. Some of the data contains the core competencies of the customers in the EEM.

Another barrier in the DDBM in the EEM business, was the lack of understanding of the setup up and maintenance of the business model, both internally in the business and in the customer segment. The board was not used to deal with these intangible platforms. The service oriented strategy in combination with products demanded a completely new mindset for the EEM.

#### **Conclusion**

The initial case study indicates there is a need to study the DDBM because they are closely related to the next level of evolution in the world of business for both products and services. This relates to trends like IOT, Big Data. There is also a need in the EEM to develop an innovation framework that combines the old business models with new DDBM, and structure the technological, competences and value propositions of these new BM. Depending on the nature of the BM this might also lead to the development of new dimensions being utilized in these DDBM.

The EEM are an example of how, one challenge leads to a chain reaction of interventions within the business, if the current BM are not maintained. This being a single case study the researchers' opinion is that the EEM is not behind the business and businesses in usual. On the contrary, the EEM is due to the business which is knowledge intensive in forerunner in the business and the organizations of the same size.

The EEM are by starting this journey, in the process of moving their business away from a commodity to a completely new way of doing business, but also to highlight the issues that are within the organization and that needs to be fixed. This new culture with focus on business models, will be in better alignment externally and develop a shared understanding and faster development in this transformational journey.

ANTONIO GHEZZI, (2014), "The dark side of business models: the risks of strategizing through business models alone", *Strategic Direction*, Vol. 30 Iss 6 pp. 1 – 4

BAINES, T. S., LIGHTFOOT, H. W., EVANS, S., NEELY, A., GREENOUGH, R., PEPPARD, J., ROY, R., SHEHAB, E., BRAGANZA, A., TIWARI, A., ALCOCK, J. R., ANGUS, J. P., BASTL, M., COUSENS, A., IRVING, P., JOHNSON, M., KINGSTON, J., LOCKETT, H., MARTINEZ, V., MICHELE, P., TRANFIELD, D., WALTON, I. M. & WILSON, H. 2007. State-of-the-art in product-service systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 221, 1543-1552

BROWNLOW, J., ZAKI, M., NEELY, A. & URMETZER, F. 2015. Data and Analytics - Data-Driven Business Models: A Blueprint for Innovation. *Cambridge Service Alliances*. University Of Cambridge: Institution for Manufacturing.

BRYMAN, A. & BELL, E. 2011. *Business research methods*, Oxford, Oxford University Press.

CHESBROUGH, H. 2007. Business model innovation: it's not just about technology anymore. *Strategy & Leadership*, 35, 12.

CPPP, E. B. D. V. 2014. Strategic Research and Innovation Agenda. 49.

HARTMANN, P., ZAKI, M., FELDMANN, N. & NEELY, A. 2014. Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms. *Cambridge Service Alliances*. University Of Cambridge: Institution for Manufacturing.

K. SCHWAB, "The Fourth Industrial Revolution," World Economic Forum, 2016.

LINDGREN, P. & RASMUSSEN, O. H. 2013. The business model cube. *Journal of Multi Business Model Innovation and Technology*, 1, 135-182.

NEELY, A., BENEDETTINI, O. & SWINK, M. 2014. Impact of firm characteristics on survival: an empirical analysis in the context of service strategies. *Cambridge Service Alliances*. University Of Cambridge: Institution for Manufacturing.

OSTERWALDER, A., PIGNEUR, Y. & TUCCI, C. L. 2005. Clarifying business models: Origins, present, and future of the concept. *Communications of the association for Information Systems*, 16, 1.

Porter, Michael E., and James E. Heppelmann. "How Smart, Connected Products Are Transforming Competition." *Harvard Business Review* 92, no. 11 (November 2014): 64–88.

R. Parmar; I. Mackenzie; D. Cohn; D. Gann, "The New Patterns of Innovation: How to Use Data to Drive Growth.

# 5G, an approach towards future telemedicine

Sadia Anwar  
Center for TeleInfrastruktur,  
Aalborg University,  
Aalborg, Denmark,  
saa@es.aau.dk

Ambuj Kumar  
Center for TeleInfrastruktur,  
Aalborg University,  
Aalborg, Denmark  
in\_kumar@es.aau.dk

Bhawani S. Chowdhry  
Mehran University of  
Engineering and  
Technology, Jamshoro  
c.bhawani@ieec.org

Ramjee Prasad  
Center for TeleInfrastruktur,  
Aalborg University,  
Aalborg, Denmark  
prasad@es.aau.dk

**Abstract**—The use of smart phones has been increasing rapidly and it is expected that in future most people will have a smart phone capable of high speed internet connection. The capability of smart phones with high definition display, computation power and multitude of sensors made it an excellent candidate for telemedicine application. Telemedicine's applications and high data medical information generally require high definition visuals and lower latency connection in addition mobility and reliability. The next generation of wireless communication standard known as 5G will provide data speed in gigabit per second with lower latency and higher reliability connection and can be better approach for future telemedicine. In this paper we survey the current state of telemedicine along with examining the characteristics of 5G technology. We also present research challenges concerning 5G and telemedicine.

**Keywords**—5G telemedicine, wireless communications

## I. INTRODUCTION

According to recent survey [1] 90% of medical practitioners are utilizing smartphones for clinical applications. By using smartphones, communication or interaction to patients becomes very easy and more effective. The estimated number of clinical applications available on smartphone is approximately 95,000. Furthermore, Internet of things (IoT) is changing the way of our everyday life, as it is going to allow many things feasible and accessible e.g., smarter health care services, smart town and cities, smart automobiles and smarter retailing and shopping.

Wireless mobile communication is not only changing the way of our life but also becoming an integral part of our lives. Smartphones contribute a lot in keeping us updated in every aspect of life, as they can access a large number of information through Internet anytime, anywhere, and at a high speed. In early days, cellular technology was only focused on messaging and making phone calls but now it is possible to perform much more complex communication and computation tasks, in particular being connected with broadband mobile Internet connection.

As smartphones or smart devices are becoming more and more computationally powerful, along with high resolution displays, multiple ways of interaction, and various kind of embedded sensors that need for faster data rate, the low latency and higher reliability become more and more crucial and thus important. Therefore, 5G will be very important in

this scenario, having capacity and high data rates, and is able to transmit data at Gb/s speed in 2020, which is about 200 times faster than 4G.

Telemedicine or telehealth is an inter-disciplinary area, which uses telecommunication technology to deliver medical information, medical help and services to remote areas at distant places, especially for the elderly and disabled persons. Telemedicine can be utilized as a first line strategy in case of emergency situations and disaster management. It was first applied in Mexico disaster by NASA in 1985. Voice communication was possible within 24 hours. With the passage of time and advancement in technology, this latency is now become small as in milliseconds by continuous modification and now by the new technology concept 5th generation mobile wireless communication (5G) which is becoming reality in 2020 [2].

We predict 5G along with telemedicine is going to revolutionize healthcare, and can be imagined in case of availability and time factors. Whereas, availability factor is basically introducing more specialists and medical physician to come and practice the telemedicine and time factor is handled by the faster wireless technology with high data rates such as 5G effectively and precisely for the delivery of virtual authentic medical care to patients.

The main lineaments of telemedicine are that it should be climbable, transparent or crystalline, provide geographical or global coverage, and fault tolerance with security, and authentication. Hospital digital networking concept came into being because of new faster wireless communication technologies, which are enhancing the approach of specialists and patients with each other to communicate visually and talk and share their condition with suggestive approach and strategies.

The rest of the paper is structured as follows: Section II describes important attributes of 5G in relation to telemedicine. Section III presents what are the research challenges in connection to 5G and telemedicine. And lastly, conclude the paper in Section IV.

## II. IMPORTANT ATTRIBUTES OF 5G IN RELATION TO TELEMEDICINE

In 2020, new and more revolutionized telehealth services will need wireless technology which can support high

definition (HD) video quality, faster speed, low latency with no interruption in signaling pathway, more authentic in providing security and supportive for subscribers to use and subscribe applications especially in telemedicine, which is a concept truly based on the delivery of medical information to remote areas by utilizing wireless communication technology. 5G is going to address M2M (machine to machine) communication. Which is further divided into:

- Massive machine type communication, it includes low cost and low energy devices e.g. sensors involved in monitoring or diagnosis of vital signs. These devices need faster speed and real time communication [3].
- Mission critical type communication, it is real time controlled and has automated field function and processing such as robotically performed remote tele-surgery. In tele-surgery doctors can perform a surgery without their presence in a particular location, virtually through wireless technology. Tele-surgery needs more immense coordination between different surgeons which are remotely connected and this also requires that information of each performed task should be available with minimum latency in real-time scenario. 5G, in this scenario, will be more reliable with faster speed and low latency [4].

Some of the important attributes concerning 5G are as follows:

*A. Some of the important attributes concerning 5G are as follows:*

- Connectivity and coverage in Gb/s [8], will be available even in natural disasters or emergency situations, thus having this technology, more lives can be saved.
- Strong authentication with secured patient profile would be imagined with 5G. However, research is still under process to avoid jamming attacks and similar complications.
- Remote monitoring of patients in distant areas would be easy, because 5G will be able to provide higher bandwidth.
- Power consumption of batteries is viewed for 2020, and charging is available for at least 10 years in M2M scenario, [8].
- 5G also enables various sensors and implant to connect with different service configurations to see patient condition and to provide them best regime or therapy.
- 5G has mobility with faster speed and low latency [8], which is less than one millisecond. Virtual reality and HD videos should be as detailed as human retina can detect. For this 300 Mb/s is generally required and this technology is 100 times above with low data rate and ultra-reliability, to carry out the High definition videos. It will be very helpful in dealing

patients in video conferences to avoid breach in communication [5].

- Medical image transmission and videos require larger bandwidth. These images and videos are compressed during transmission and it is expected on the other end that there would be no data or image loss after receiving. 5G can handle HD videos and high data transmission in case of remote monitoring of patients, video conferencing, robotic equipment's, smart pharmaceuticals such as ventilators, fluid rate and drug delivery systems attached to the patients in remote areas. As an example, in diabetic patients, insulin reservoir are controlled and connected with wireless technology. Sensors provide the signals on daily basis for body glucose level. Insulin is then injected accordingly after taking the values.
- With 5G, first-aid could be delivered in small duration of time and communication without interruption available with the blink of eye.
- More and more patients are motivated to consult with doctors online. It is not only saving time but its cost effective features making it more reliable. 5G can play a more vital role to access these consultations frequently and remotely with ultra-low cost and low-end data rate reliability.
- Cloud service robotics for supportive therapy will be easy as offloading and uploading of information for recognition of language, object and cognitive skills will be predicted on real time with no interruption. Cost will also be reduced as less number of robots is involved [4].
- Remote monitoring of elderly patients will be optimized through continuous updates about life style modification and treatment for patients with chronic diseases and related disabilities. Their rehabilitation care plan and strategies can be reviewed by wireless communication, for this wireless technology has to follow the same standard for urban and rural areas where there is higher density of followers and users.

III. RESEARCH CHALLENGES

A new and advance form of existing technology always has inherent challenges to deal with new problems. First generation mobile communication (1G) came in 1980s while fifth generation (5G) is going to be commercialized in 2020. It is a journey of approximately 40 years. During the modification and transformation of these technologies, we are still lacking proper infrastructure, standards, security and legislations concerning wireless communication technology. Fig.1 shows and elaborate concept of telemedicine services running on 5G network.

The challenges that we are facing in wireless communication concerning telemedicine are:

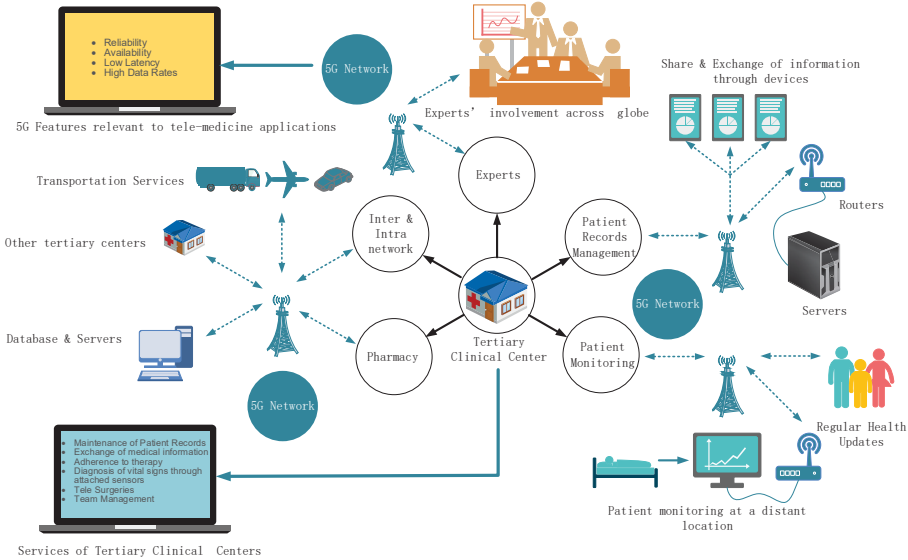


Fig 1: 5G for telemedicine applications

A. Bandwidth requirements for offloading medical services and applications

Clinical services can be divided for telemedicine into two main categories within a clinical organization.

- Inter telemedicine services
- Intra telemedicine services

Tertiary clinical center implicate two or three medical professionals to deal with a common patient case especially in pediatric cardiology, where data which is in the form of images or medical records are transferred to get a common solution of a given problem e.g., ECG, chest radiology and murmurs of neonates is called inter telemedicine services.

On the other hand, different tele conferences and meetings would be arranged to join different medical professionals from different departments called intra telemedicine services. This is mostly happened by communication through wireless technology. For Medical images transmission and videos require larger bandwidth and high resolution for 3D (three dimensional) images and other body scans so that after compression during transmission, there would be no data loss or interruption in the other receiving end because service is shared by different users at one time in a common place.

On the contrary, this band- width requirement provoke the situation to more critical level, when there is a need to see or monitor a patient on a distant place like patient with

disabilities or elderly people with degenerative pathologies (e.g.: Alzheimer, or the insurgence of other pathologies such as stroke, cardio circulatory and muscular dysfunction etc. Latest clinical advice is to perform physical activity regularly during the day and to maintain a social activity. For this there is a need of comprehensive health care plan. So, constant engagement with physician is needed.

The biggest socioeconomic challenge that has been arising in Europe is aging. According to EU commission public health policy, by 2025 more than 20% of Europeans will be 65 or over, with a particularly rapid increase in numbers of over-80s. (EU Commission) [6].

There is a need to decrease their abundance in hospital because it will not only create burden for medical staff but also made the environment more congested to deal with. Latest GPRS (General packet data rate service) system is not able to provide this real time communication precisely especially for the recording of patient movements during exercise, for this patient has to be monitored in hospital to get the values accurately. Apparatus utilized in hospitals are very outdated and based on the radiometry with limited area for coverage [7].

There is a need of high definition breakage free wireless communication that can fill up the gap of space (which means less occupancy of patients) and time factor (response is quick). Patient and physician can communicate data from different attached sensors, exoskeletons to see patient movements and

smart pharmaceutical devices which are providing dose tailored on suggestive and maintained therapy on a signaling pathway by the physician on real time with no delay especially for remote areas. 5G can change this whole scenario and can help to fill this gap created by space and time factor. Data could be easily analyzed with suggestive therapy on a mobile anywhere, anytime with a few milliseconds delay and can be handled with fastest communication.

Less costly cellular displays, medical applications and software's can easily be accessed and managed. Medication and dose adjustment by sensors and smart pharmaceutical devices for cardiac and diabetic patients can easily be done by physician with positive feedback mechanism.

*B. Reliability and availability requirements*

Recent increase in emergency situation and disasters, forecasting or their prediction is not easy. Sometimes, it is very time consuming, but recently happened events, can give us a knowledge that how we can deal with them in future and how help can be provided based on a real time with no communication loss. Billions of people are affected by these disasters. These disasters could be technical, natural or human generated which can cause deaths, disabilities and psychological stress for a long time.

Hospitals are continuously active in these crises and the biggest challenge in disaster management is communication. Proper system and methodology is not fully available in handling, processing and transmission of critical data on real time especially at a distant place properly. In disasters, regional communication services are severely affected and communication becomes constrained and distorted.

There is a need of secure architecture for wireless communication system and 5G will be beneficiary in such scenario as 5G has a high band-width and low latency, and this system can be integrated with cloud computing system by this health care services will be more reliable and available.

*C. Standards and security*

5G, being a new emerging technology is still lacking standards and this might take many months to years, to be defined as well implemented. Many of the universities and government organizations are working in this part to make 5G concept into a reality based technology. It is said that set standards will released in 2020, to make this target date into reality many of the public and government organization like IEEE, 3GPP, universities, ITU are making efforts to make this possible [8].

Prevailing Technological and operational guidelines which are in practice concerning telemedicine are [9]:

- Audio-video and data transmission should be of high quality and fulfill telehealth existing practice. Equipment's must have latest security software's as

recommended by the device manufacturer. There must be a backup plan in case of communication or data loss.

- Multiple authentications and no activity timeout function can be utilized and in case of mobile or device loss or get stole, provider must be able to disable the connection and data.
- Real-time connectivity or synchronization requires bandwidth of at least 384kbps for down and uplink assignments and such kind of services must also provide a resolution of 640\*480 at 30 frame/seconds. According to some health practitioner that high speed standard quality of video is not similar on a same bandwidth provided by different technologies.
- Video conference must utilize a link that in case of band width loss, stabilize and connection must not be lost.
- Whole disk encryption must be done for storage of a synchronized intentional data if cloud services are unable to give higher level of safety for health professionals and patient records, then it should be streamed to avoid unauthorized users and hacking attacks.
- Patient and physician should pretest the connection before communication is going to be happen. Tele or video conference software's must open one session to be conducted for discussion if there is an attempt or a hack to open second session, it must be automatically log off or access to second session must be denied.

These guidelines are helpful to set the standards. Here are some of the areas, where standardization is still needed.

*1) Minimum delay require for delay application services*

Telephonic education about medical issues require rate of data transfer from few 100 Mb/s to many 100 Gb/s The speed of 4G is approximately 10 to 20 Mb/s. Synchronized applications needs a minimum delay of 400 milliseconds [10].

International telecommunication service has given us the value of transmission of a data packet in IP networks for 3G which is 100 to 400 milliseconds for end to end delay and 1 per 1000 packets loss for synchronized services. In telemedicine for routine patient checkup QoS (quality of service) for 3G offers 150 to 400 millisecond delay which is less important as compared to emergency services where it offers a delay of 0 to 150 millisecond [11].

*2) Bandwidth requirement for video based application*

5G in such perspective is effective because it can transfer the images and data for telemedicine without interruption with very small delay or latency of about one millisecond. It will be more effective and reliable especially in remote patient monitoring with effective real time consultation [12].



Information technology utilization especially in health care sector has been very slow and the gatekeepers involved in this sector, making it difficult to develop in case of standards for data delivery and content [4].

In case of any network like Wi-Fi, 3G, 4G wireless network or wire system for an effective video call transmission, require a minimum band width of 230kbps (standard resolution) for uploading while for downloading of an additional video of ongoing call per video window require 178MB more. If we add more people in ongoing communication 128kbps are more needed for additional downloading per video window [13].

File size and type for transfer in medical facility is different for telediabetic retinopathy, MRIs scans, digital chest films, electrocardiogram studies and telepathology [14].

By adopting 5G which has data rate of 10 Gbit/s is more appropriate in transmission of these files with ease, higher speed and data rate because it has ability to cover all the challenges arises due to speed and audio video quality by providing higher bandwidth with low latency rates.

### 3) Security requirement and data encryption

Standards for data transmission and policies are not fully grown, still on basic or initial phase which needs reevaluation. There are not safety standards available for new innovative telemedicine services and for these cost is very high for their transmission on a broad band services [15].

Security makes a linkage between three main parts. These are self-sustainability which means, data is fully available not fragmented. Second is confidentiality which denotes that data is shared between authorized persons or users and third part of this linkage is availability which is the utilization of different services of networking. These three parts are interconnected and their balance actually is the success of provision of medical services. The first two parts are very important in telemedicine while, third is not as critical as others because there is always a backup plan available for data access. Self-sustainability is also important in this scenario, because in case of data loss it will directly affect the patient treatment or regime; the reason for this is that sound information is lost during transmission.

Data can also be encrypted, as a secret key is shared between one particular health unit and related staff dealing with a particular patient. They can access the data for treatment anytime. Data can be hacked or misused iso security is much more important within the hospital. Food and drug authority is only concerned to regulate medical instrument or devices and does not set any standards for consumer devices and application related to privacy concerns. So basically it is for patient safety but not concerned with patient privacy.

## IV. CONCLUSION

In our survey we can conclude that work is still needed to define standards, and in particular security framework to make telemedicine services secure and also to make these services

more applicable globally. A comprehensive framework is required to address future telemedicine using 5G technologies. There are, however, many uncertain areas what 5G is promising such as if a consistent, low latency and reliable connection will be available globally. More future research is needed to explore what kind of telemedicine application can emerge in connection to 5G technology, and more advanced smart devices having various sensors, and artificial intelligence.

## REFERENCES

- [1] Mobile devices in healthcare come with pros and cons,? SearchHealthIT. [Online]. Available: <http://searchhealthit.techtarget.com/tip/Mobile-devices-in-healthcare-come-with-pros-and-cons>. [Accessed: 03-Jun-2016].
  - [2] Y. Smith, ?A Brief History of NASA's Contributions to Telemedicine,? NASA, 16-Aug-2013. [Online]. Available: <http://www.nasa.gov/content/a-brief-history-of-nasa-s-contributions-to-telemedicine>. [Accessed: 03-Jun-2016].
  - [3] 5G Radio Access for Ultra-Reliable and Low-Latency Communications,? Ericsson Research Blog, 11-May-2015. .
  - [4] 5G-PPP: 5G and e-Health.?
  - [5] 5G Vision: 100 Billion Connections, 1 ms Latency, and 10 Gbps Throughput.? [Online]. Available: <http://www.huawei.com/minisite/5g/en/defining-5g.html>. [Accessed: 03-Jun-2016].
  - [6] Policy - European Commission.? [Online]. Available: [http://ec.europa.eu/health/ageing/policy/index\\_en.htm](http://ec.europa.eu/health/ageing/policy/index_en.htm). [Accessed: 03-Jun-2016].
  - [7] V. Oleshchuk and R. Fensli, ?Remote Patient Monitoring Within a Future 5G Infrastructure,? Wirel. Pers. Commun., vol. 57, no. 3, pp. 431-439, Jul. 2010.
  - [8] J. Best, ?The race to 5G: Inside the fight for the future of mobile as we know it,? TechRepublic. [Online]. Available: <http://www.techrepublic.com/article/does-the-world-really-need-5g/>. [Accessed: 03-Jun-2016].
  - [9] J. Bernard and J. D. Linkous, ?Core Operational Guidelines for Telehealth Services Involving Provider-Patient Interactions,? [Online]. Available: <http://www.americantelemed.org/docs/default-source/standards/core-operational-guidelines-for-telehealth-services.pdf?sfvrsn=6>. [Accessed: 03-Jun-2016].
  - [10] [S. L. Kota, K. Pahlavan, and P. A. Lepp, nen, Broadband Satellite Communications for Internet Access. Springer Science & Business Media, 2011.
  - [11] Y.1541:ýNetwork performance objectives for IP-based services.? [Online]. Available: <https://www.itu.int/rec/T-REC-Y.1541-201112-1/en>. [Accessed: 03-Jun-2016].
  - [12] 5G Radio Access for Ultra-Reliable and Low-Latency Communications,? Ericsson Research Blog, 11-May-2015. .
  - [13] What is the minimum amount of bandwidth required for a video call? Do you have high resolution video?? [Online]. Available: <http://support.oovoo.com/link/portal/3908/4244/Article/1503/What-is-the-minimum-amount-of-bandwidth-required-for-a-video-call-Do-you-have-high-resolution-video>. [Accessed: 03-Jun-2016].
  - [14] D. S. Kayange, ?Telemedicine Available Bandwidth Estimation Simulation Model for Effective E-Health Services: Categories, Requirements and Network Application,? Acad. Res. Int., vol. 5, no. 5, pp. 11220, Sep. 2014.
- A. Tedesco, D. Di Lieto, L. Angrisani, M. Campanile, M. De Falco, and A. Di Lieto, ?Telemedicine & Broadband,? Intechopen, 16-Aug-2013. [Online]. Available: [http://cdn.intechopen.com/pdfs/14281/InTech-Telemedicine\\_broadband.pdf](http://cdn.intechopen.com/pdfs/14281/InTech-Telemedicine_broadband.pdf). [Accessed: 03-Jun-2016].

# Sparse Modeling Methods for Misbehavior Detection in LSA Networks

George Papageorgiou, Konstantinos Voulgaris, Constantinos Papadias  
Athens Information Technology,  
Athens, Greece,  
Email: gepa.kovo,cpap@ait.gr

**Abstract**—In this paper we consider the task of misbehavior detection in License Shared Access (LSA) networks. Since the functionality of such wireless communication networks depends to a great extent on the compliance of the Licensee Users (LUs) with the agreed policy, the task of successfully detecting unauthorized activity is of major importance. The transmitted signals of the misbehaving LUs are modeled by using sparsity arguments. Since prompt detection is required, we employ the recently established Soft-Thresholding with simplified Exact Line search Algorithm (STELA) for solving the related sparse optimization task and use its solution to estimate the transmitted power for each LU. Moreover, we exploit the STELA and modify it accordingly in order to solve various reweighted  $\ell_1$ -norm tasks. Finally, our simulation results clearly demonstrate the advantages of the proposed techniques for the detection process.

## I. INTRODUCTION

Licensed Shared Access (LSA) is passing from its initial, static stage to a more dynamic one. Instead of allocating resources to a few licensee operators under long term sharing agreements LSA proposes that multiple licensee and incumbent users share radio resources with resource allocations lasting for as little as a few seconds. While such improved dynamicity increases the efficiency of spectrum sharing, it also requires stricter compliance with the agreed sharing rules, so that Licensee Users (LUs) do not exceed their allocated resources in the time, space, and frequency domains. To this end, the problem of accurately and timely identifying such misbehaving users is raised.

In order to ensure the functionality of such an architecture we consider that the entire network is under the supervision of the Incumbent Users (IUs). The IU network is composed of access points, which are sparingly used as sensors. The goal is the detection of possible LU transmissions in the geographic area that is under the control of the IUs according to the LSA agreement. The task of detecting unauthorized transmissions reflects to the ability of a method to successfully locate the source(s) of transmissions, and therefore identify the LUs that violate the agreement policy, i.e., misbehave. Thus, the detection aims at: a) locating the LU that violates the agreement policy and b) estimating the severity of the violation (e.g., transmit power) for each LU over a specific period and at a specific band. Basic sensing techniques have already been used for the localization of the transmission source and are based on: a) the Angle of Arrival (AoA) information or, b) the Received Signal Strength Indicator (RSSI) estimation [1],

[2]. However, such methods are prone to errors due to radio propagation effects, which often lead to the increase of the probabilities of misdetection or false alarm.

The focus of this work is to provide a source localization framework. The misbehaving LU detection process is performed by exploiting sparse modeling and optimization methods, based on the knowledge (or estimates) of the current channel state information (CSI). Sparsity-aware learning and related optimization techniques have been at the forefront of the research in signal processing, encompassing a wide range of topics, such as compressed sensing, denoising and signal approximation techniques, [3]. In the current manuscript, we employ STELA and modify it accordingly in order to solve advanced sparse optimization tasks, based on the reweighting of the  $\ell_1$ -norm we perform a comparison between these methods in terms of their estimation performance and their computational requirements.

**Notation:** Throughout this work, small letters, e.g.,  $t$ , denote scalars; bold capital letters, e.g.,  $\mathbf{H}$ , denote matrices; and, bold lowercase letters, e.g.,  $\mathbf{x}$ , are reserved for vectors (each vector is regarded as a column vector), while the symbol  $\mathbf{x}^H$  denotes the Hermitian of the respective matrix/vector. Also,  $\text{diag}(\mathbf{A})$ , where  $\mathbf{A}$  is a square matrix, denotes the respective vector with entries the diagonal elements of  $\mathbf{A}$ . The  $j$ -th coordinate of the vector  $\mathbf{x}$  is denoted by  $x_j$  and  $h_{ij}$  is reserved for the entry of the matrix  $\mathbf{H}$ . Finally,  $\mathbf{a} \circ \mathbf{b}$  denotes the Hadamard product between the two vectors  $\mathbf{a}$  and  $\mathbf{b}$ ;  $\mathcal{S}_a(b) := [b-a]^+ - [-b-a]^+$  is the soft-thresholding operator; and,  $[s]_0^1 := \min(\max(s, 0), 1)$  denotes the projection of  $s$  onto  $[0, 1]$ .

## II. SYSTEM MODEL

Consider a Cognitive Radio (CR) network composed of a known number  $N_I$  of IU sensors and  $N_L$  LUs. All users are located in a square geographic area  $\mathcal{A}$  that is divided into smaller cells. Without loss of generality we assume a square grid of size  $M \times M$ , as shown in Figure 1. Furthermore, it is assumed that: a) the number of the transmitting LUs is much smaller than the size of the grid ( $N_L \ll M^2$ ) and that only a single LU is located at each cell (this can always be accomplished by a finer segmentation of the grid) and b) the channels  $g_{n,k} \in \mathbb{C}$  between the  $k$ -th LU (transmitter) and the  $n$ -th IU sensor are block fading and frequency flat, see [4]. Finally, in the following work, we consider that the channels

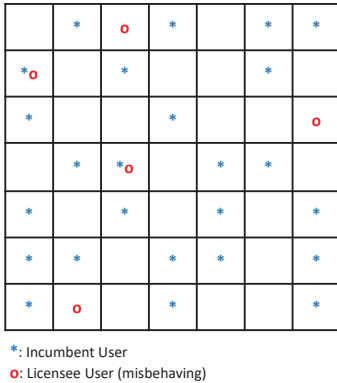


Fig. 1: Illustration of a  $7 \times 7$  grid of the geographical area  $\mathcal{A}$ .

are known (or estimated) within a small symbol interval  $T_S$ , which was also considered in [5], [6], [7].

Let the transmitted signal at time instance  $t$  from the  $k$ -th LU be  $x_k(t) \in \mathbb{C}$ ,  $t = 1, \dots, T_S$ . Since the number of the LUs is much smaller than the number of all the possible locations over the area  $\mathcal{A}$  ( $N_L \ll M^2$ ) the transmitted signal from all the LUs is modeled as a sparse vector denoted by  $\mathbf{x}_*(t) \in \mathbb{C}^{M^2}$ . We assume that the number  $N_L$ , the location of the LUs, which is the support set  $\mathcal{S}$  of the sparse vector  $\mathbf{x}_*(t)$ , as well as the channels remain unchanged within the sampling time period  $T_S$ . The signal sample received at each (IU's)  $n$ -th sensor is assumed to be of the linear form:

$$y_n(t) = \mathbf{g}_n^H \mathbf{x}_*(t) + v_n(t), \quad n = 1, \dots, N_I, \quad (1)$$

where  $\mathbf{g}_n := [g_{n,1}, \dots, g_{n,M^2}]^H$  characterizes the channels between the  $n$ -th sensor and all possible locations and  $v_n(t)$  is the additive noise assumed i.i.d. with zero mean and bounded variance, uncorrelated with  $g_{n,k}$ . Equation (1) can be written in a more compact form as

$$\mathbf{y}(t) = \mathbf{G}\mathbf{x}_*(t) + \mathbf{v}(t), \quad (2)$$

where  $\mathbf{y}(t) = [y_1(t), \dots, y_{N_I}(t)]^T$ ,  $\mathbf{G}^H = [\mathbf{g}_1 \dots \mathbf{g}_{N_I}]$  and  $\mathbf{v}(t) = [v_1(t), \dots, v_{N_I}(t)]^T$ .

In order to successfully detect a possible unauthorized transmission of a LU, our goal is to recover both the support of the sparse vector  $\mathbf{x}_*(t)$  (which reveals the location of the transmitters), but also estimate its values (transmitted signal) or the transmitting power over the fixed sampling time period  $T_S$ . In addition, in order to classify the activity of a LU (legitimate- $\mathcal{H}_0$  or illegitimate- $\mathcal{H}_1$ ), a standard energy detector (other detectors are also applicable) is used on the estimated signal  $\hat{\mathbf{x}}(t)$ :

$$E_k = \frac{1}{T_S} \sum_{t=1}^{T_S} |\hat{x}_k(t)|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \tau_{ed}(k), \quad (3)$$

where  $\tau_{ed}(k)$  is a model-defined threshold which depends on the agreed power transmission levels for each LU. In the simplest scenario, where no transmissions are authorized by the IU network,  $\tau_{ed}(k)$  is set at the noise level for all LUs. Selecting an appropriate estimator is of major importance, since for non-sparse estimates there is an increased probability of false alarm resulting in an unfair classification of the activity for the respective LU (misbehaving). Since the Ordinary Least-Squares (OLS) method produces a non-sparse solution (thus is prone to false alarm errors), we resort to sparse optimization methods.

### III. DETECTION VIA SPARSE OPTIMIZATION METHODS

There are two major paths towards modeling sparse vectors/signals. The first one, focuses on minimizing the  $\ell_0$  (pseudo)-norm of a vector, which equals the number of non-zero coordinates of a vector (this is a non-convex function). Thus, according to (2), the resulting minimization task is:

$$\min_{\mathbf{x}(t)} \|\mathbf{x}(t)\|_0 \text{ s. t. } \|\mathbf{y}(t) - \mathbf{G}\mathbf{x}(t)\|_2 \leq \varepsilon, \quad t = 1, \dots, T_S. \quad (4)$$

However, the cost function in (4) is non-convex and the optimization task is known to be NP-hard (combinatorial).

The second path, which is nowadays considered as the standard alternative technique, is to minimize the  $\ell_1$ -norm of the unknown sparse vector, which also generates sparse representations and is the closest convex relaxation to the  $\ell_0$ -norm. Moreover, instead of using the formulation similar to (4), its unconstrained formulation is often considered, which is known as the *Least Absolute Shrinkage and Selection Operator* (LASSO) [8], [9], [10], [11], i.e.,

$$\min_{\mathbf{x}(t)} \left\{ \frac{1}{2} \|\mathbf{y}(t) - \mathbf{G}\mathbf{x}(t)\|_2^2 + \lambda \|\mathbf{x}(t)\|_1 \right\}, \quad t = 1, \dots, \quad (5)$$

where  $\lambda$  is a user defined regularization constant that controls the vector's sparsity level.

The optimization task in (5) is solvable with a variety of methods such as the Alternating Direction Method of Multipliers (ADMM) [12], [8], the Homotopy method [13], the FISTA [10] and the FLEXA [14]. Although the cost function that is minimized is convex, the non-differentiable term ( $\ell_1$ -norm) is the reason for the relatively slow convergence for the majority of the aforementioned methods.

#### A. LASSO with STELA

More recently, a simplified exact line search method has been proposed for the LASSO task in (5). The so-called Soft-Thresholding with simplified Exact Line search Algorithm (STELA) [5], [6], which is also given under the framework that handles a more general class of non-differentiable optimization tasks, offers good approximation properties whilst offering computational advantages against its competitors. This is very important in many Cognitive Radio (CR) environments, since transmission operations are extremely fast, thus imposing time limitations for successful detection.

**Algorithm 1** Weighted STELA

---

```

1: procedure WSTELA( $\mathbf{G}$ ,  $\mathbf{y}$ ,  $\lambda$ ,  $\tau_{max}$ ,  $\mathbf{W}$ )
2:    $\tau \leftarrow 0$ ,  $\mathbf{x}^{(\tau)} = \mathbf{0}$ 
3:   while  $\tau < \tau_{max}$  do
4:      $\mathbf{r}(\mathbf{x}^{(\tau)}) = \text{diag}(\mathbf{G}^H \mathbf{G}) \circ \mathbf{x}^{(\tau)} - \mathbf{G}^H(\mathbf{G}\mathbf{x}^{(\tau)} - \mathbf{y})$ 
5:      $\mathbb{B}\mathbf{x}^{(\tau)} = (\text{diag}(\mathbf{G}^H \mathbf{G}))^{-1} \circ \mathcal{S}_{\lambda} \text{diag}(\mathbf{W})(\mathbf{r}(\mathbf{x}^{(\tau)}))$ 
6:      $\mathbf{z} = \mathbf{G}(\mathbb{B}\mathbf{x}^{(\tau)} - \mathbf{x}^{(\tau)})$ 
7:      $\gamma^{(\tau)} = \left[ -\frac{(\mathbf{G}\mathbf{x}^{(\tau)} - \mathbf{y})^H \mathbf{z} + \lambda(\|\mathbf{W}\mathbb{B}\mathbf{x}^{(\tau)}\|_1 - \|\mathbf{W}\mathbf{x}^{(\tau)}\|_1)}{\mathbf{z}^H \mathbf{z}} \right]_0^1$ 
8:      $\mathbf{x}^{(\tau+1)} = \mathbf{x}^{(\tau)} + \gamma^{(\tau)}(\mathbb{B}\mathbf{x}^{(\tau)} - \mathbf{x}^{(\tau)})$ 
9:      $\tau \leftarrow \tau + 1$ 
10:  Output:  $\hat{\mathbf{x}} := \mathbf{x}^{(\tau_{max})}$ 

```

---

### B. One-step Approach Weighted LASSO with STELA

Despite the fact that the optimization task in (5) designates sparse representations, it does not satisfy the ‘‘oracle properties’’<sup>1</sup> as shown by Fan and Li in [15]. To this end, Zou proposed a more advanced model, the so-called Adaptive LASSO task [16], which assigns different weights to the sparse vector’s coefficients:

$$\hat{\mathbf{x}}(t) := \arg \min_{\mathbf{x}(t)} \left\{ \frac{1}{2} \|\mathbf{y}(t) - \mathbf{G}\mathbf{x}(t)\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}(t)\|_1 \right\}, \quad t = 1, \dots, T_S, \quad (6)$$

where  $\mathbf{W}$  is a diagonal matrix with weights

$$w_{jj} = \frac{1}{|x_j(t)|^\gamma}, \quad j = 1, \dots, M^2, \quad (7)$$

where  $\gamma > 0$  and  $\mathbf{x}(t)$  is the OLS solution. However, other choices are also applicable in (7), such as the LASSO solution, although the oracle properties may no longer be satisfied. Although the task in (6) enjoys better statistical properties for the appropriate selection of  $\mathbf{x}(t)$  and  $\lambda$ , it still suffers from slow convergence if solved with a standard technique such as the ADMM. To this end, we introduce the STELA method for solving the respective task.

### C. Iteratively Reweighted LASSO with STELA

Another approach is the use of the log-penalized estimator term in (5) for every  $t$ , i.e.,

$$\min_{\mathbf{x}(t)} \left\{ \frac{1}{2} \|\mathbf{y}(t) - \mathbf{G}\mathbf{x}(t)\|_2^2 + \lambda \sum_{j=1}^{M^2} \log(|x_j(t)| + \epsilon) \right\}, \quad (8)$$

for  $\epsilon > 0$  and by taking a first-order Taylor-series approximation of the logarithm penalty about the current value, the

<sup>1</sup>An oracle procedure asymptotically satisfies the following properties: a) identifies the correct subset model and b) has the optimal estimation rate. Continuous shrinkage is also desired.

over-approximation of (8) is obtained:

$$\hat{\mathbf{x}}^{(i)}(t) := \arg \min_{\mathbf{x}(t)} \left\{ \frac{1}{2} \|\mathbf{y}(t) - \mathbf{G}\mathbf{x}(t)\|_2^2 + \lambda \sum_{j=1}^{M^2} w_{jj}^{(i-1)} |x_j(t)| \right\}, \quad (9)$$

$$w_{jj}^{(i-1)} = \frac{1}{|\hat{x}_j^{(i-1)}(t)| + \epsilon},$$

where  $\hat{\mathbf{x}}^{(0)}(t)$  is the LASSO solution. The task in (9) is known as the Iteratively Reweighted LASSO (IRWL). The linear approximation in (9), which is a convex task and was also proposed in [17], is analogous to the one proposed in [18], except that the authors in [18] did not make use of the perturbation  $\epsilon$  and they argued strongly in favor of the on-step approximation in (6). In order to solve the task in (9) efficiently and with minimal computational cost, we propose the STELA method for the IRWL task, directly applicable with a fixed weight matrix  $\mathbf{W}$ , which is updated after each ( $i$ -th) step in which a weighted LASSO task is solved.

The general weighting STELA scheme is given in Algorithm 1 for the weights given in III-B and III-C. In the next section, we compare the various STELA implementations for the LASSO and the weighted LASSO tasks for detecting misbehaving LUs. Various simulation results demonstrate each method’s performance and highlight the advantages introduced by the implementation of the weighted STELA methods in order to improve the estimation.

## IV. EXPERIMENTAL EVALUATION

For the experimental evaluation of the method we have considered the following setup. Let  $h_{n,k} \in \mathbb{C}$  denote the block fading, Rayleigh distributed component from the  $k$ -th LU to the  $n$ -th IU receiver/sensor. The respective path loss component  $r_{n,k} := d_{n,k}^{-\alpha}$  where  $d_{n,k}$  is the distance between the source and the receiver and  $\alpha$  corresponds to a propagation constant, is also computed over the grid. A typical choice for the selection of the exponent in an urban environment is between 2.5 and 3.7 (here we let  $\alpha = 3$ ). Hence, the channel between the  $k$ -th LU and the  $n$ -th IU sensor is modeled as  $g_{n,k} = r_{n,k} h_{n,k}$ .

We have evaluated each one of the following algorithms in terms of detection, (power) estimation and convergence speed.

- STELA: solves the LASSO task given in (5). The regularization parameter has been optimized and set at  $\lambda = 1$ .
- AD-LASSO<sub>OLS</sub>: this one-step-reweighting scheme solves (6) with weights given in (7) from the LS solution and  $\gamma = 3$ . The regularization parameter is set here to  $\lambda = 10$ .
- AD-LASSO<sub>LASSO</sub>: one-step-reweighting scheme which solves (6) with weights given in (7) from the LASSO solution and  $\gamma = 1$ . The LASSO task is solved by STELA for  $\lambda = 1$ ; however, since the solution is a sparse one, a small  $\epsilon$  parameter ( $10^{-3}$ ) is also considered as in (9) in

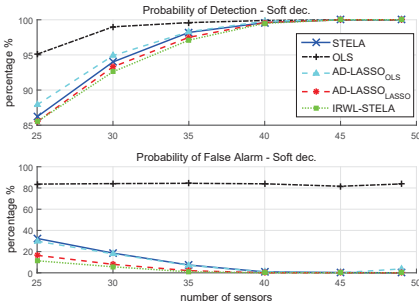


Fig. 2: Probabilities of detection and false alarm while varying the number of sensors, according to the *soft decision rule*.

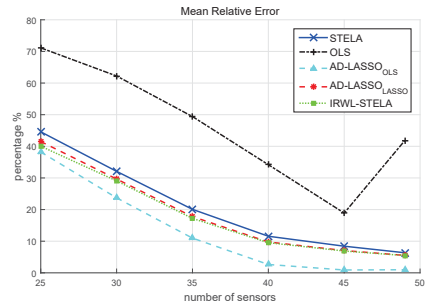


Fig. 4: The relative error between the transmitted and the estimated energy.

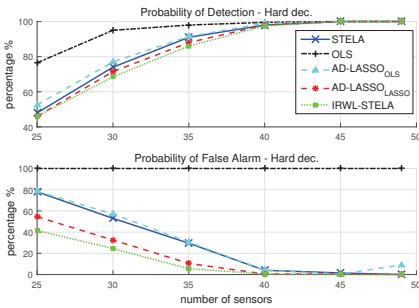


Fig. 3: Probabilities of detection and false alarm while varying the number of sensors, according to the *hard decision rule*.

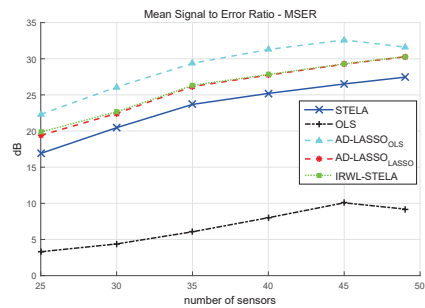


Fig. 5: The SER for each method, while varying the number of sensors over the grid.

order to guarantee the numerical stability of the method. The  $\lambda$  value for the AD-LASSO task is set at  $\lambda = 10$ .

- IRWL-STELA: solves the task given in (9) for  $10^{-3}$ . The regularization parameter at the initial LASSO step is set at  $\lambda = 0.8$  and for the iterations of (9) at  $\lambda = 10$ .
- OLS: the Ordinary Least-Squares method is only given as a reference point.

It should be noted that all lambda values are optimized according to the MSE criterion. We have also terminated each method once the accuracy between subsequent iterations dropped below  $10^{-12}$ . All experiments were performed in a conventional PC operating at 3.4 GHz on a 64-bit system.

We consider a topology of a  $7 \times 7$  square grid, where 5 LUs (corresponding to approx. 10% of the possible locations) are distributed uniformly at random over the entire grid. Assuming that the band is occupied by the IU's network, any possible transmission by a LU (above the noise level) is considered illegitimate. Each one of the LUs transmits with a fixed power (energy) selected uniformly at random over the interval [60,100]. We also assume that  $\mathbf{h}_n \in \mathcal{CN}(\mathbf{0}, \mathbf{I})$  and  $v_n(t) \in \mathcal{CN}(0, 0.5)$ . The distances from each cell to

the next are considered  $d = 1$  between two subsequent cells (on the same row or column),  $\sqrt{2}$  for the first diagonal cell and 0.8 for transmission within the cell. Once computing the estimated signal  $\hat{\mathbf{x}}(t)$ ,  $t = 1, \dots, T_S$ , the energy at each cell is evaluated according to (3) and the user is classified as legitimate or illegitimate. We let the number of IUs vary between 25 and 49 users chosen at random over the grid, while we measure the probability of detection, i.e.,  $P_d = \Pr\{E_k > \tau_{ed}(k)|\mathcal{H}_1\}$ , and the probability of false alarm, i.e.,  $P_f = \Pr\{E_k > \tau_{ed}(k)|\mathcal{H}_0\}$ , in two different ways: a) a *soft*; and, b) a *hard* decision rule. In the soft rule, the detection and false alarm probabilities are computed by taking into account each one estimated location. In the hard rule, a qualitative measure for the entire architecture is used; that is, the detection is considered successful only if the entire set of misbehaving users is identified and the false alarm probability is set at zero only if no extra user is incorrectly classified as a misbehaving one. For the experiment we have chosen  $T_S = 30$  and performed 200 independent Monte Carlo runs.

In Figure 2 we have evaluated the probabilities of detection and false alarm according to the soft decision criterion, for

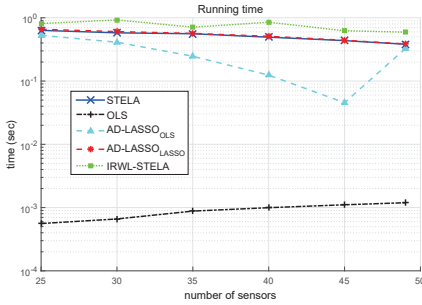


Fig. 6: Log-scale of the time (sec) required for each method to converge.

each method. The detection probabilities are almost equally good (except for OLS). However, with the IRWL-STELA method we have improved (reduced) the probability of false alarm up to 20% with respect to the STELA. It is also worth noting the performance of the AD-LASSO<sub>LASSO</sub> with a slightly greater probability of false alarm.

Next, we present the probabilities obtained according to the hard decision rule. As can be readily seen, the performance of each method is similar; however, the probability of detection has decreased, while the probability of false alarm has increased for all of the methods. Figure 3 depicts the respective values. By investigating these results we conclude that the difference in performance for each method reduces as the number of sensors increases (the respective probabilities tend to 100% for the detection and 0% for the false alarm). The hard detection rule is stricter as it requires that the complete sets are recovered (no single mistake is forgiven). Good probabilities result in an overall good functionality of the LSA network. Moreover, in Figure 4, each method's performance towards power estimation is shown. It can be seen that the AD-LASSO<sub>OLS</sub> manages to perform a better estimation towards the transmitting power. The estimation is also evaluated in dB scale by the Mean Signal to Error Ratio (SER), where the SER is defined as:

$$\text{SER} = 10 \log_{10} \left( \frac{\|\mathbf{x}_*(t)\|_2^2}{\|\mathbf{x}_*(t) - \hat{\mathbf{x}}(t)\|_2^2} \right),$$

as shown in Figure 5. It is observed that the reweighting methods manage to decrease the estimation error significantly; however the Adaptive LASSO with the OLS weights excels.

Finally, in Figure 6 we have evaluated the time running required for each method to converge. It should be noticed that the total computational cost is only indicative of the required time, since convergence also depends on the selection of the regularization parameter  $\lambda$ .

## V. CONCLUSIONS

In the current work we propose a variation of the STELA method for solving sparse optimization tasks based on

reweighting techniques. Although these methods are considered within the framework of misbehavior detection in LSA networks, they are readily applicable to any similar sparsity-aware learning optimization tasks. The reweighting techniques have not only shown to improve the sensing process by reducing the probability of false alarm but they have also improved the estimation of the transmitted power, resulting in a better usage of the available spectrum.

## ACKNOWLEDGMENT

This work has been supported by the European Commission FP 7 Research Project ADEL (grant agreement No: 619647).

## REFERENCES

- [1] R. Peng and M. L. Sichițiu, "Angle of arrival localization for wireless sensor networks," in *Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on*, vol. 1. IEEE, 2006, pp. 374–382.
- [2] C. Papamathou, F. P. Preparata, and R. Tamassia, "Algorithms for location estimation based on rssi sampling," in *Algorithmic Aspects of Wireless Sensor Networks*. Springer, 2008, pp. 72–86.
- [3] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015.
- [4] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.
- [5] Y. Yang, M. Zhang, M. Pesavento, and D. P. Palomar, "An online parallel algorithm for spectrum sensing in cognitive radio networks," in *Signals, Systems and Computers, 2014 48th Asilomar Conference on*. IEEE, 2014, pp. 1801–1805.
- [6] Y. Yang and M. Pesavento, "A novel line search method for nonsmooth optimization problems," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1726–1730.
- [7] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for mimo ad hoc cognitive radio networks," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3117–3131, 2011.
- [8] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, 2011.
- [9] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale 1-regularized least squares," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 606–617, 2007.
- [10] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [12] J. Eckstein, D. P. Bertsekas *et al.*, "An alternating direction method for linear programming," 1990.
- [13] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA journal of numerical analysis*, vol. 20, no. 3, pp. 389–403, 2000.
- [14] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *Signal Processing, IEEE Transactions on*, vol. 63, no. 7, pp. 1874–1889, 2015.
- [15] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [16] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [17] B. A. Johnson, Q. Long, Y. Huang, K. Chansky, and M. Redman, "Model selection and inference for censored lifetime medical expenditures," *Biometrics*, 2015.
- [18] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of statistics*, vol. 36, no. 4, p. 1509, 2008.
- [19] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the cramér-rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, 2009.

# DOA Estimation Based Bayesian Compressive Spectrum Sensing Using Dual Polarization Antenna Receiving

Kun Su, Xuekang Sun  
School of Network Education  
Beijing University of Posts and Telecommunications  
Beijing, China 100876  
Email: sukun@bupt.edu.cn

Caili Guo, Shiyu Ren  
School of Information and Communication Engineering  
Beijing University of Posts and Telecommunications  
Beijing, China 100876  
Email: rsy@bupt.edu.cn

**Abstract**—The application of Bayesian compressive sensing framework to spectrum sensing has successfully attracted more attention, as it can sample sparse signals at sub-Nyquist rates in wideband cognitive radio network (CRN) to alleviate the bandwidth requirements on the hardware of most receivers. However, in the presence of arriving angle, the energy of the signal cannot be fully received, which decreases the performance of a receiving system using the orthogonal polarization antenna. To deal with above problem, we propose direction of arrival (DOA) estimation based Bayesian compressive spectrum sensing (DBCS) for dual polarization antenna (DPA) receiving system. In this scheme, we establish a Bayesian compressive sensing model for DPA receiving system and introduce DOA estimation into sensing model to increase spectrum sensing accuracy. Simulation results show that the proposed scheme can effectively improve the accuracy of spectrum sensing compared to no considering DOA.

**Index Terms**—Bayesian compressed sensing; dual polarization antenna; spectrum sensing; DOA

## I. INTRODUCTION

With the development of wireless communication technology, more and more frequency resources are used. But due to the fixed spectrum allocation policy, the utilization of spectrum resources becomes very low. As an intelligent spectrum sharing technology, cognitive radio (CR) can be used to find “spectrum holes” and take advantage of these holes without harmful interference to licensed users. It can be seen that the spectrum sensing is a crucial technology in CR. However, Nyquist sampling for wideband spectrum needs to use plentiful bandwidth resources of receivers. While Compressive Sensing (CS) can provide a solution to relieve the pressure from the high sampling rate [1]. There are several compressive sensing algorithms which are widely used, such as Orthogonal Matching Pursuit (OMP) [2,3], Basis Pursuit (BP) [4] and Bayesian compressive sensing (BCS) [5]. As Bayesian compressive sensing has better robustness and faster convergence rate than other algorithms [6], it can be adopted for spectrum sensing.

At present, the research on wideband spectrum sensing is mostly focused on the single polarization antenna system and assumes that the transmitted signal can be fully received in order to simplify the structure. But the fact is that when

the receiving antenna does not match with the polarization direction of the incoming wave [7], it makes the performance of spectrum sensing unstable. Array antenna, especially large scale multi-antenna, is not suitable for second users (unlicensed users) because of the antenna space, size and other factors. However, as the orthogonal dual polarization antenna receiving system can receive all the information of the incoming wave, its performance is more stable in contrast with single polarization antenna. Compared with array antenna, it needs less space and is simpler.

In a real case, the impact of DOA cannot be ignored. The energy of received signals may be reduced as a result of the presence of DOA. The performance of polarization detectors may go bad. For the studies of DOA, most are focused on array antenna. And a large number of methods, such as multiple signal classification (MUSIC) [8], and rotation invariant subspace algorithm (ESPRIT) [9], maximum likelihood (ML) [10], subspace fitting (SF) [11] and so on, are also emerging.

In this paper, we propose DOA estimation based Bayesian compressive spectrum sensing (DBCS) using dual polarization antenna receiving. In this scheme, we firstly use the established Bayesian compressive sensing model in DPA receiving system to obtain the preliminary sensing spectrum without considering the presence of DOA. And then we choose a frequency form it as the occupied signal frequency, on which the amplitude of the sensing signal is maximum. Finally, we introduce DOA estimation based on above frequency signal to spectrum compressive sensing model for the improved sensing accuracy.

The remainder of the paper is organized as follows. In section II, we introduce BCS and the signal model. In section III, DOA estimation based Bayesian compressive spectrum sensing using dual polarization antenna receiving is proposed. The simulation and analysis are shown in section IV. Section V provides a summary of the full text.

## II. SIGNAL MODEL

### A. Polarized signal

The polarized signal  $\mathbf{X}$  can be decomposed into two parts.  $\mathbf{X}_h$  and  $\mathbf{X}_v$  denote the horizontally (h) and vertically (v) polarization components respectively. We use Jones vector to represent the polarized signal  $\mathbf{X}$ .

\*This work is supported by Chinese National Nature Science Foundation (61372116).

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_h \\ \mathbf{X}_v \end{bmatrix} = \begin{bmatrix} \mathbf{X}_h^i + j\mathbf{X}_h^q \\ \mathbf{X}_v^i + j\mathbf{X}_v^q \end{bmatrix} \quad (1)$$

where  $i$  is the in-phase component and  $q$  is the quadrature component. We can get the Jones vector  $\mathbf{X}_p$  [12].

$$\mathbf{X}_p = \begin{bmatrix} \cos \theta_p \\ \sin \theta_p e^{j\varphi_p} \end{bmatrix} \quad (2)$$

$\theta_p$  is polarization angle and  $\varphi_p$  is phase difference. Amplitude-phase descriptor  $(\theta_p, \varphi_p)$  can be used to represent polarization state.

$$\theta_p = \arctan \frac{|\mathbf{X}_v^i|}{|\mathbf{X}_h^i|}, \theta_p \in \left(0, \frac{\pi}{2}\right) \quad (3)$$

$$\varphi_p = \arctan \frac{|\mathbf{X}_v^q|}{|\mathbf{X}_v^i|} - \arctan \frac{|\mathbf{X}_h^q|}{|\mathbf{X}_h^i|}, \varphi_p \in (0, 2\pi) \quad (4)$$

### B. BCS

According to the theory of compressive sensing, the input wideband spectrum signal vector is denoted by vector  $\mathbf{X}$ . Its sparsity is  $\mathbf{K}$ , which represents the number of non-zero elements in  $\mathbf{X}$ .  $\Phi$  denotes the  $M \times N$  Gaussian random measurement matrix. The length of compressive measurements  $\mathbf{Y}$  is  $M$ , where  $M < N$ . Thus,  $\mathbf{Y}$  can be represented as

$$\mathbf{Y} = \Phi \mathbf{X} + \mathbf{e} \quad (5)$$

where  $\mathbf{e}$  is zero-mean Gaussian noise with variance  $\delta^2$ . Therefore, we can get the Gaussian likelihood model

$$p(\mathbf{Y}|\mathbf{X}, \delta^2) = (2\pi\delta^2)^{-\frac{M}{2}} e^{-\frac{1}{2\delta^2} \|\mathbf{Y} - \Phi \mathbf{X}\|^2} \quad (6)$$

As analyzed above, the problem of compressive sensing has been converted to a linear regression problem with the sparse limit. Supposing  $\Phi$  is known, what we need to do is to estimate the sparse vector  $\mathbf{X}$  and noise variance  $\delta^2$ . Therefore we use prior information to estimate posterior probability of distributions of interested parameters by the Relative Vector Machine (RVM). Firstly, we assume every element of  $\mathbf{X}$  follows non-zero Gaussian distribution.

$$p(\mathbf{X}|\alpha) = \prod_{i=1}^N N(x_i|0, \alpha_i^{-1}) \quad (7)$$

And  $\alpha_i$  is inverse-variance for the accuracy of the Gauss density function. A Gamma prior is considered over  $\alpha$ . We can define a zero-mean Gaussian prior on each element of  $\mathbf{X}$ .

$$p(\alpha|a, b) = \prod_{i=1}^N \Gamma(\alpha_i|a, b) \quad (8)$$

After calculating marginal distribution for the hyper parameters  $\alpha$ , the overall prior on  $\mathbf{X}$  is represented as

$$p(\mathbf{X}|a, b) = \prod_{i=1}^N \int_0^\infty N(x_i|0, \alpha_i^{-1}) \cdot \Gamma(\alpha_i|a, b) d\alpha_i \quad (9)$$

In the above formula,  $\prod_{i=1}^N \int_0^\infty N(x_i|0, \alpha_i^{-1}) \cdot \Gamma(\alpha_i|a, b) d\alpha_i$  follows the Student-t distribution. When we choose suitable  $a$  and  $b$ , the Student-t distribution will get the peak about  $x_i = 0$ . As a result, the sparse prior in (9) promotes  $x_i$  to be zero. In the same way, the inverse

of the noise variance  $\alpha_0 = 1/\delta^2$  may have a Gamma prior  $\Gamma(\alpha_0|c, d)$ . Under the hypothesis that hyper parameters  $\alpha$  and  $\alpha_0$  are known, given the CS measurements  $\mathbf{Y}$  and the projection matrix  $\Phi$ , the posterior for  $\mathbf{X}$  can be introduced as a multivariate Gaussian distribution whose mean and covariance are  $\mu$  and  $\Sigma$

$$\mu = \alpha_0 \overline{\Sigma} \Phi^T \mathbf{Y} \quad (10)$$

$$\Sigma = (\alpha_0 \Phi^T \Phi + \mathbf{A})^{-1} \quad (11)$$

where  $\mathbf{A} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_N)$ ,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ . What we should do is to search for the hyper parameters  $\alpha$  and  $\alpha_0$ . Then we can use RVM to estimate these hyper parameters by performing a type-II maximum likelihood procedure[4]. The logarithm  $L(\alpha, \alpha_0)$  can be written as

$$\begin{aligned} L(\alpha, \alpha_0) &= \log p(\mathbf{Y}|\alpha, \alpha_0) \\ &= \log \int p(\mathbf{Y}|\mathbf{X}, \alpha_0) p(\mathbf{X}, \alpha) d\mathbf{X} \\ &= -\frac{1}{2} [K \log 2\pi + \log |\log \mathbf{C}| + \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y}] \end{aligned} \quad (12)$$

with  $\mathbf{C} = \delta^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T$ . We can use point estimates for  $\alpha$  and  $\alpha_0$  to get the maximum for (8) by the use of a type-II ML approximation. As a consequence, we can obtain:

$$\alpha_i^{new} = \frac{r_i}{\mu_i^2} \quad (13)$$

$$\frac{1}{\alpha_0^{new}} = \frac{\|\mathbf{Y} - \Phi \mathbf{X}\|_2^2}{k - \sum_i r_i} \quad (14)$$

where  $\mu_i$  is the  $i$ -th posterior mean weight from (10) and  $\sum_i$  is the  $i$ -th diagonal element of the posterior weight covariance from (11).  $r_i := 1 - \alpha_i \sum_i$ . As illustrated above,  $\alpha^{new}$  and  $\alpha_0^{new}$  can be obtained as long as  $\mu$  and  $\Sigma$  are known, while  $\mu$  and  $\Sigma$  can also be achieved as long as  $\alpha$  and  $\alpha_0$  are known. So we may use an iterative algorithm in (10)-(14) until a convergence criterion has been satisfied.

### C. DOA

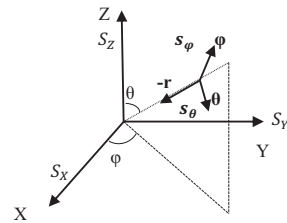


Fig. 1. DOA

We make an assumption that there is a spherical coordinate system for an orthogonal dual-polarization receiving antenna. An incoming signal in three dimensions is shown in Fig. 1. The incident wave arrived along a direction of  $(\theta, \varphi)$ , where  $\theta$



is the elevation angle and  $\varphi$  is the azimuth angle. The signal can be expressed as

$$E = E_\theta + E_\varphi \quad (15)$$

$E_\theta$  and  $E_\varphi$  are the horizontal component in  $\theta$  and vertical component in  $\varphi$ . Then signal can be expressed in the  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  directions as

$$\begin{bmatrix} E_X \\ E_Y \\ E_Z \end{bmatrix} = \begin{bmatrix} -\sin\varphi & \cos\theta\cos\varphi \\ \cos\varphi & \cos\theta\sin\varphi \\ 0 & -\sin\theta \end{bmatrix} \begin{bmatrix} E_\theta \\ E_\varphi \end{bmatrix} \quad (16)$$

In this paper, we use 2-element array which has two pair of orthogonal dual polarized antenna. Assuming that the signal in the Y-axis is the horizontal component and the signal in the Z-axis is the vertical component. The DOA is the same for the two pair of orthogonal dual polarized antenna. And the received signal  $E$  for each dual polarized antenna can be represented as

$$E = \begin{bmatrix} \cos\varphi & \cos\theta\sin\varphi \\ 0 & -\sin\theta \end{bmatrix} \begin{bmatrix} E_\theta \\ E_\varphi \end{bmatrix} \quad (17)$$

In general,  $\theta$  in most channel models such as 3GPP/3GGP2 and the 802.11n standard is set to  $\pi/2$  [13]. We can get

$$E = \begin{bmatrix} \cos\varphi & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} E_\theta \\ E_\varphi \end{bmatrix} \quad (18)$$

### III. DOA ESTIMATION BASED BAYESIAN COMPRESSIVE SPECTRUM SENSING USING DUAL POLARIZATION ANTENNA RECEIVING

We consider a scenario that there is a receiver with orthogonal dual polarization antennas. In order to alleviate the hardware requirements of the receiver, Compressive Sampling can be used instead of Nyquist Sampling.

We assume that the original wideband signal  $\mathbf{X}$  includes  $N$  subbands, in which  $K$  subbands are occupied randomly.  $\mathbf{X}$  is the  $N \times 1$  matrix with elements  $x_1, x_2, \dots, x_n (n = 1, 2, \dots, N)$  representing the amplitude of  $N$  subbands. The horizontal component of the signal is  $\mathbf{X}_h$  and the vertical component is  $\mathbf{X}_v$ . They are all  $N \times 1$  matrixes. The signal is represented by  $\mathbf{Y}$  after it is compressed.  $\mathbf{Y}_h$  and  $\mathbf{Y}_v$ , which are compressive signals corresponding with  $\mathbf{X}_h$  and  $\mathbf{X}_v$  respectively, are received signals with the channel noise. They are all  $M \times 1$  matrixes.  $\Phi$  is the  $M \times N$  measurement matrix. Then in Gaussian channel, considering the polarization angle  $\theta_p$ , we can get

$$\mathbf{X} = \frac{\mathbf{X}_v}{\sin\theta_p} = \frac{\mathbf{X}_h}{\cos\theta_p} \quad (19)$$

$$\mathbf{Y} = \Phi(\mathbf{X} + \mathbf{e}_1) = \Phi\mathbf{X} + \Phi\mathbf{e}_1 = \Phi\mathbf{X} + \mathbf{e}_2 \quad (20)$$

$$\mathbf{Y}_v = \Phi(\mathbf{X}_v + \mathbf{e}_3) = \Phi\mathbf{X}_v + \Phi\mathbf{e}_3 = \Phi\mathbf{X}_v + \mathbf{e}_4 \quad (21)$$

$$\mathbf{Y}_h = \Phi(\mathbf{X}_h + \mathbf{e}_5) = \Phi\mathbf{X}_h + \Phi\mathbf{e}_5 = \Phi\mathbf{X}_h + \mathbf{e}_6 \quad (22)$$

$\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_5$  are additive white Gaussian noise.  $\mathbf{e}_2, \mathbf{e}_4, \mathbf{e}_6$  can be approximated as a zero-mean Gaussian noise. We assume that they have the same distribution.  $\mathbf{Y}_h, \mathbf{Y}_v$  and  $\Phi$  are known. Thus we may use  $\mathbf{Y}$  to recover  $\mathbf{X}$  by BCS.

Through (19), (21) and (22), we can get

$$\mathbf{Y}_v = \sin\theta_p\Phi\mathbf{X} + \mathbf{e}_4 \quad (23)$$

$$\mathbf{Y}_h = \cos\theta_p\Phi\mathbf{X} + \mathbf{e}_6 \quad (24)$$

Using (23) and (24), we can get

$$\sin\theta_p\mathbf{Y}_v + \cos\theta_p\mathbf{Y}_h = \Phi\mathbf{X} + \sin\theta_p\mathbf{e}_4 + \cos\theta_p\mathbf{e}_6 \quad (25)$$

$\sin\theta_p\mathbf{Y}_v + \cos\theta_p\mathbf{Y}_h$  is approximately equal to  $\mathbf{Y}$  with small noise. In order to obtain  $\theta_p$ , we may use  $\arctan\left[\frac{Y_v}{Y_h}\right]$  to represent it approximately. Using this method, we are able to recover the signal  $\mathbf{X}$  with a low computational complexity

we can select a certain frequency which has a maximum of the amplitude. Using DOA estimation method based on narrow band, we can obtain the DOA at this frequency. And the DOA of the wideband signal is also confirmed.

It is assumed that the interval between the first and second sampling is very short. The DOA of the first sampling signal is the same as the DOA of the second sampling. Thus, the result of the second compressive sampling can be corrected by the DOA which is obtained from the first sampling.

$$\begin{bmatrix} X_h \\ X_v \end{bmatrix} = \begin{bmatrix} \cos\varphi & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X_{h'} \\ X_{v'} \end{bmatrix} \quad (26)$$

$X_{h'}$  and  $X_{v'}$  are the signal where the impact of DOA has been removed. Then we can repeat the above steps to get modified results by the DOA. The system model is as follow

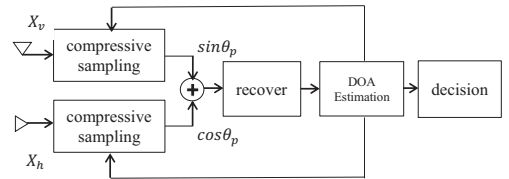


Fig. 2. system model proposed

### IV. EXPERIMENTAL RESULTS

In this section, the performance of the proposed system model which use DOA estimation based the Bayesian compressive spectrum sensing (DBCS) is compared against the traditional system model which only use the Bayesian compressive spectrum sensing and does not consider the DOA (NDBCS) by means of simulations. All our experiments are done using MATLAB R2015a on a 3.00 GHz Intel Pentium PC with 4GB RAM. We do simulation experiments in the Rayleigh fading channel where the cross-polar discrimination (XPD) is 6.5 dB, correlation coefficients are all 0, false alarm probability ( $P_{fa}$ ) is 0.1.

We consider a 500-dimensional vector to represent the wideband spectrum, and choose 10 locations in the spectrum to represent 10 PUs. We use a  $120 \times 500$  Gaussian random matrix to represent the projection matrix. The compressive measurement vector may be compressed from 500-dimensional to 120-dimensional. The variation of measurement vector reflects the variation of compression ratio.

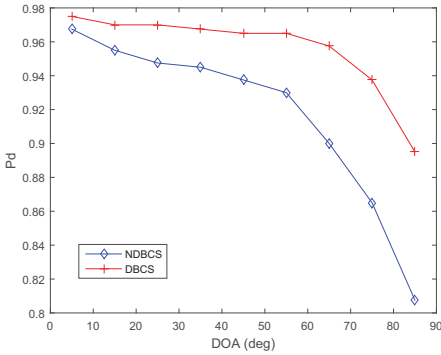


Fig. 3.  $P_d$  at various DOA with  $N=500, M=120, K=20, SNR=10\text{dB}$ .

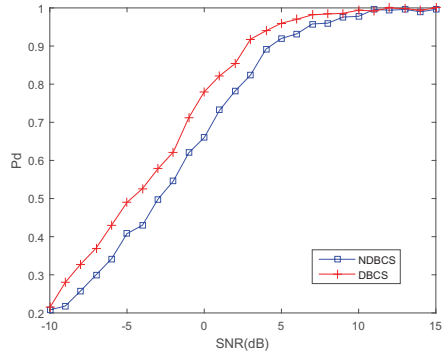


Fig. 4.  $P_d$  at various SNR (dB) with  $N=500, M=120, K=10, \theta = \frac{\pi}{3}$ .

Fig.3 shows the relation between the DOA and detection probability ( $P_d$ ). When DOA changes, the DBCS has better performance than the NDBCS. With the increase of DOA, the  $P_d$  is reduced, but the DBCS is better than the NDBCS.

Fig.4 shows the relation between signal to noise ratio (SNR) and  $P_d$ . In Fig.4, we can see that  $P_d$  enhances correspondingly along with increasing of SNR. It is obvious that the performance of DBCS is better than NDBCSs with the same SNR.

Fig.5 shows the relation between number of measures and  $P_d$ . In Fig.5, we can see that  $P_d$  becomes larger with the increasing number of measures. With the same number of measures, the  $P_d$  of DBCS is higher than NDBCSs. When they have the same performance, the DBCS need less measures.

From the above, we can see that the impact of the DOA for the detection effect is very large. It is indicated that the smaller the DOA, the higher the probability of detection, the better the recovery of the signal with the same SNR and the same number of samples. The performance of DBCS is always better than NDBCS's. Compared with the NDBCS, the proposed method can reduce the influence of DOA.

### V. CONCLUSION

In this paper, a novel approach called DOA estimation based Bayesian compressive spectrum sensing using DPA receiving system is proposed. In this scheme, we build the Bayesian compressive spectrum sensing model based on DPA. At first, we use this model to achieve a preliminary sensing spectrum, and we choose a certain frequency on which the amplitude of the sensing signal is maximum for DOA estimation. Thus, we obtain DOA estimation based Bayesian compressive sensing model considering the affect of DOA. The simulations show that this method has a better detection performance compared with the compressive sensing algorithm that does not consider the impact of DOA.

### REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

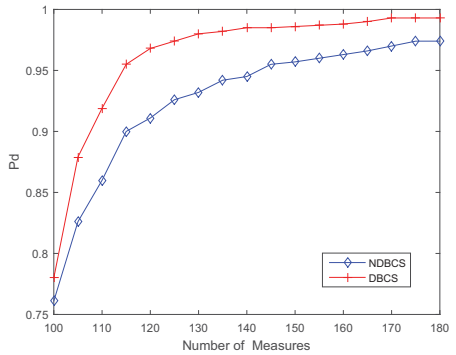


Fig. 5.  $P_d$  at various measures with  $N=500, K=20, \theta = \frac{\pi}{3}, SNR=10\text{dB}$ .

[2] J. A. Tropp and A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[3] Y. Wang, C. Guo, X. Sun and C. Feng, "Time-Efficient Wideband Spectrum Sensing Based on Compressive Sampling," *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, 2015, pp. 1-5.

[4] Chen S S, Donoho D L, Saunders M A. "Atomic decomposition by basis pursuit," *SIAM review*, 2001, 43(1): 129-159.

[5] S. Ji , Y. Xue and L. Carin, "Bayesian compressive sensing," *IEEE Trans on Signal Process*, vol. 56, no. 6, pp. 2346-2356, 2008.

[6] N. Zhang, X. Sun, C. Guo and L. Gao, "PU probability prediction based Bayesian compressive spectrum sensing," *IEEE 2015 IEEE/CIC International Conference on Communications in China (ICCC)*, Shenzhen, 2015, pp. 1-5.

[7] Y. Deng, X. Sun and C. Guo, "Polarization information based spectrum detection algorithm for cognitive radio," *Wireless, Mobile and Multimedia Networks (ICWMMN 2011)*, Beijing, 2011, pp. 155-159.

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276-280, Mar 1986.

[9] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984-995, Jul 1989.

[10] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-

- Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720-741, May 1989.
- [11] M. Viberg, B. Ottersten and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436-2449, Nov 1991.
- [12] F. Liu, C. Feng, C. Guo, Y. Wang and D. Wei, "Virtual Polarization Detection: A Vector Signal Sensing Method for Cognitive Radios," *Vehicular Technology Conference (VTC 2010-Spring)*, 2010 IEEE 71st, Taipei, Taiwan, 2010, pp. 1-5.
- [13] A. Weiss, and B. Friedlander, "Maximum likelihood signal estimation for polarization sensitive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 7, pp. 918-925, Jul 1993.

# Bayesian Cognitive Radio Spectrum Sensing in Multipath Rayleigh Fading Channels

Usama Y. Mohamad and Dirk Dahlhaus

Communications Laboratory, University of Kassel, Germany

Email: mohamad@uni-kassel.de

**Abstract**—A novel Bayesian hypothesis test for sensing a primary user (PU) signal in a cognitive radio single antenna receiver operating in an uncorrelated scattering Rayleigh fading environment is presented, where the optimum threshold for the test is found analytically. The threshold depends on the PU signal autocorrelation function, the average power of the channel impulse response between the PU transmitter and the sensing receiver as well as the noise power spectral density. In opposite to other schemes trying to use cooperative sensing for taking into account the delay dispersion of the channel, the proposed scheme requires no cooperation for sensing and thus reduces the signal processing complexity and the amount of signaling. The performance of the proposed algorithm is studied in channels with an exponential power delay profile. It turns out that the devised sensing approach being applicable for both Gaussian and non-Gaussian PU signals leads to a considerable improvement as compared to the energy detector even if the latter has perfect knowledge of the received signal under the two sensing hypotheses.

## I. INTRODUCTION

Wireless spectrum occupancy measurements evidence spectrum scarcity with a multitude of frequency bands not being fully utilized at a particular time or specific location [1], [2]. The current fixed spectrum allocation is not able to meet the ever increasing demand for flexible data rates. There are several proposals to change today's spectrum access towards dynamic spectrum access [3], among them cognitive radio (CR) firstly proposed in [4]. In a CR system, low-priority secondary users (SUs) have to coexist with high-priority primary users (PUs), where an opportunistic access of the spectrum by the SU guarantees an acceptable interference level at the PU receiver [5].

In [6], possible CR implementations have been categorized into interweave, underlay and overlay systems. Each category is based on different system components and design considerations. Prior to cognitive radio access, spectrum sensing (SSE) has to be carried out in order to detect either spectral holes or the existence of a PU signal [7]. Different survey papers describe possible SSE techniques [8], [9], [10]. The latter differ in terms of computational complexity, the required received signal observation frame length as well as resulting performance represented by the receiver operating characteristics [11].

The standard SSE approach is the energy detector (ED) which is applicable to a large number of different signaling and interference scenarios due to its simple implementation. On the

other hand, its performance depends critically on an accurate estimate of the power spectral density of the noise in the received signal [12]. Waveform based SSE is applied in the presence of a known signal pattern, e. g. a preamble signal, and provides more accurate sensing results at the expense of higher synchronization requirements [13]. Cyclostationarity based SSE exploits the fact that the statistics of the transmitted signals in many communication systems are periodic due to inherent periodicities such as modulation rate and carrier frequency [14]. Matched Filtering SSE correlates the perfectly known PU signal with the received signal. It shows high performance even for a short sensing time, but usually requiring a complex receiver implementation [15]. On top of these classical sensing algorithms, many other novel methods have been introduced such as eigenvalue-based SSE [16], Wishart matrix-based SSE [17] and SSE based on compressive sensing [18].

A particular challenge in SSE arises in multipath fading wireless channels. The presence of the channel delay dispersion makes the SSE more complicated since the intersymbol interference makes the fading signal more difficult to be distinguished from a situation of an absent PU signal. A straightforward solution in such a scenario is to exploit spatial diversity to overcome the fading effects [19]. The resulting cooperative sensing scheme implies that the decisions or metrics coming from cooperating nodes must be combined at a central node provided with the ability of using one of several standard combining techniques [20]. Clearly, the resulting scheme outperforms a non-cooperative approach at the expense of a higher system complexity, an increased signaling between the cooperating nodes as well as higher delays before making a decision [9].

In an attempt to devise a robust low-complexity non-cooperative SSE being able to overcome the problems in fading environments without having to accept the aforementioned drawbacks of the cooperative solution, we introduce a SSE algorithm which is a Bayesian hypothesis test where the received signal under the hypothesis of a PU signal being present is a PU signal at the output of a multipath uncorrelated scattering fading channel superimposed by additive Gaussian noise [21]. The proposed algorithm is generic and can be used for any PU signal distribution.

The paper is organized as follows. Sect. II describes the system model including the fading channel and the received signal. The proposed sensing scheme is derived in Sect. II-C. The

performance is investigated analytically and based on simulations for different types of PU signal autocorrelation functions in Sect. III. Conclusions are finally drawn in Sect. IV. Throughout the paper a boldface quantity like  $\mathbf{Y}$  defines a matrix or vector of suitable dimension. An instance of a random variable  $Y_k$  is denoted by  $y_k$ . The transpose of  $\mathbf{Y}$  is written as  $\mathbf{Y}^T$ , while  $\mathbf{Y}^\dagger$  is the complex conjugate transpose. Furthermore,  $\mathbf{I}_N$ ,  $\delta_k$ ,  $\mathbf{0}$ ,  $\|\cdot\|$ ,  $(\cdot)^*$ ,  $\text{diag}(\mathbf{x})$ ,  $\det(\cdot)$  and  $\mathbf{E}\{\cdot\}$  denote the  $(N \times N)$  identity matrix, the Kronecker delta, an all-zero column vector, the Euclidean vector norm, complex conjugation, a diagonal matrix with main diagonal components containing the coefficients of vector  $\mathbf{x}$ , a determinant and the expectation operator, respectively.

## II. SYSTEM DESCRIPTION

We consider the complex baseband description of a time-discrete signal  $Y_k$  with time index  $k$  at a receiver equipped with a single antenna and operating in an environment which can be described as a linear delay-dispersive uncorrelated scattering (US) Rayleigh fading channel. In the following, a binary hypothesis test is formulated where  $Y_k$  is modeled under hypothesis  $H_0$  characterizing the absence of a PU signal and  $H_1$  characterizing a PU signal being present. The test uses a  $K$ -dimensional vector  $\mathbf{Y} = [Y_1, \dots, Y_K]^T$  with  $K \in \mathbb{N}$ .

### A. PU signal absent hypothesis $H_0$

Without a PU signal,  $Y_k$  is modeled by

$$Y_k = W_k, \quad (1)$$

where  $W_k$  is a circularly-symmetric zero-mean white complex Gaussian random process with average power  $\mathbf{E}\{|W_k|^2\} = \sigma_W^2$ . Therefore,  $\mathbf{Y}$  under  $H_0$  has a covariance matrix  $\Sigma_{\mathbf{Y}|H_0} = \mathbf{E}\{\mathbf{W}\mathbf{W}^\dagger\} = \sigma_W^2 \mathbf{I}_K$  and a probability density function (PDF) given by

$$\mathbf{Y}|H_0 \sim \mathcal{CN}(\mathbf{0}; \Sigma_{\mathbf{Y}|H_0}) \quad (2)$$

with  $\mathcal{CN}(\boldsymbol{\mu}; \Sigma)$  denoting a multivariate complex Gaussian distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ .

### B. PU signal present hypothesis $H_1$

The PU signal  $S_k$  is modeled as a wide-sense stationary process with autocorrelation  $\rho_{SS,\kappa} = \mathbf{E}\{S_{k+\kappa} S_k^*\}$  and average power  $\mathbf{E}\{|S_k|^2\} = \sigma_S^2 = \rho_{SS,0}$ , where  $S_k$  is not necessarily Gaussian distributed. The US fading channel between the PU transmitter and the SU receiver is characterized by a channel impulse response (CIR) whose coefficients are the components of the vector  $\mathbf{C} = [C_0, \dots, C_{N-1}]^T$ . Here,  $N \in \mathbb{N}$  characterizes the excess delay of the channel and  $\mathbf{C} \sim \mathcal{CN}(\mathbf{0}; \Sigma_{\mathbf{C}})$  with  $\Sigma_{\mathbf{C}} = \text{diag}(\mathbf{Q})$  and  $\mathbf{Q} = [q_0, \dots, q_{N-1}]^T$  denoting the channel's power delay profile (PDP) where  $q_n = \mathbf{E}\{|C_n|^2\}$  for  $n = 0, \dots, N-1$ . With these assumptions, the received signal is given by

$$Y_k = \sum_{n=0}^{N-1} C_n S_{k-n} + W_k, \quad (3)$$

where  $W_k$  is defined as in (1) and independent of  $S_k$ . Since the calculation of the PDF of  $Y_k$  under  $H_1$  is analytically difficult and in view of the superposition of  $N$  independent random variables in (3), we adopt the central limit theorem (CLT) to model  $\mathbf{Y}|H_1$  as being Gaussian distributed. From the aforementioned assumptions, it follows directly that  $Y_k$  has zero mean and the autocorrelation function is given by  $\rho_{Y Y, \kappa} = \mathbf{E}\{Y_{k+\kappa} Y_k^*\} = p_{C, \text{tot}} \rho_{SS, \kappa} + \sigma_W^2 \delta_\kappa$  with  $p_{C, \text{tot}} = \mathbf{E}\{\|\mathbf{C}\|^2\} = \sum_{n=0}^{N-1} q_n$  denoting the average total power of the CIR. As a consequence, we obtain

$$\mathbf{Y}|H_1 \sim \mathcal{CN}(\mathbf{0}; \Sigma_{\mathbf{Y}|H_1}) \quad (4)$$

with

$$\Sigma_{\mathbf{Y}|H_1} = p_{C, \text{tot}} \Sigma_{\mathbf{S}} + \sigma_W^2 \mathbf{I}_K \quad (5)$$

and the  $(K \times K)$ -dimensional signal covariance matrix  $\Sigma_{\mathbf{S}}$  containing the value  $\rho_{SS, i-j}$  in the  $i$ th row of the  $j$ th column for  $1 \leq i, j \leq K$ .

### C. Bayesian Sensing Test

Obviously, a Bayesian test based on (2) and (4) does not require any channel estimation, but only an estimate of the channel's PDP that can be obtained in different ways [22], [23]. Furthermore, as pointed out above, the test does not require the PU signal to be Gaussian, but can be applied as soon as estimates of the covariance matrices in (2) and (4) are available. Applying the minimum probability of error approach in [24], the index of the hypothesis to be chosen can be obtained for an observation  $\mathbf{Y} = \mathbf{y}$  from an optimum Bayes test after some algebra in the form

$$\delta(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in \Gamma_1 \\ 0 & \text{if } \mathbf{y} \in \Gamma_0 \end{cases} \quad (6)$$

with the decision regions given by the sets

$$\begin{aligned} \Gamma_0 &= \{\mathbf{y} \in \mathbb{C}^K : L(\mathbf{y}) < \tau_{\text{opt}}\} \\ \Gamma_1 &= \{\mathbf{y} \in \mathbb{C}^K : L(\mathbf{y}) \geq \tau_{\text{opt}}\}. \end{aligned}$$

Here, the log-likelihood ratio is given by

$$L(\mathbf{y}) = -\mathbf{y}^\dagger \Sigma_{\mathbf{Y}|H_1}^{-1} \mathbf{y} + \frac{1}{\sigma_W^2} \|\mathbf{y}\|^2 \quad (7)$$

and the optimum threshold reads

$$\tau_{\text{opt}} = \ln \left( \frac{\pi_0 \det(\Sigma_{\mathbf{Y}|H_1})}{\pi_1 \sigma_W^{2K}} \right), \quad (8)$$

where  $\pi_i$  is the prior probability of hypothesis  $H_i$  for  $i = 0, 1$ . The test statistics in (7) can be reformulated where the result gives a certain insight into the relation of the proposed scheme with respect to well-known sensing approaches like e.g. the ED. Upon definition of the matrix  $\mathbf{B} = \frac{p_{C, \text{tot}}}{\sigma_W^2} \Sigma_{\mathbf{S}}$ , (7) can be written as the complex quadratic form

$$L(\mathbf{y}) = \mathbf{y}^\dagger \mathbf{A} \mathbf{y} \quad (9)$$

with the Hermitian matrix [25]

$$\mathbf{A} = -\frac{1}{\sigma_W^2} \left( (\mathbf{B} + \mathbf{I}_K)^{-1} - \mathbf{I}_K \right) = \mathbf{A}^\dagger. \quad (10)$$

Applying the singular value decomposition (SVD) of  $\mathbf{A}$ , we obtain [26]

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger, \quad (11)$$

where  $\mathbf{V}$  is a  $(K \times K)$ -dimensional unitary matrix with columns representing the eigenvectors of  $\mathbf{A}\mathbf{A}^\dagger$  or  $\mathbf{A}^\dagger\mathbf{A}$ , and  $\mathbf{D} = \text{diag}(\boldsymbol{\lambda})$  contains the sorted non-negative singular values  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_K]^\top$  of  $\mathbf{A}$ . As in [24], we introduce new coordinates  $\tilde{\mathbf{y}} = \mathbf{V}^\dagger\mathbf{y} = [\tilde{y}_1, \dots, \tilde{y}_K]$  and use  $\mathbf{A}$  from (11) in (9), which leads to

$$L(\mathbf{y}) = \mathbf{y}^\dagger \mathbf{V}\mathbf{D}\mathbf{V}^\dagger \mathbf{y} = \tilde{\mathbf{y}}^\dagger \mathbf{D} \tilde{\mathbf{y}} = \sum_{k=1}^K \lambda_k |\tilde{y}_k|^2. \quad (12)$$

With (12), the test in (6) amounts to comparing the weighted energy of the linearly transformed received signal vector to a threshold where both the weights' vector  $\boldsymbol{\lambda}$  and the threshold  $\tau_{\text{opt}}$  depend on the second order moments of the channel and the PU signal processes.

### III. PERFORMANCE ANALYSIS

In this section, the test performance is investigated using both analytic and numerical approaches for a normalized multipath fading channel with  $\sum_{n=0}^{\infty} q_n = 1$  and an exponential PDP given by

$$q_n = (1 - \exp(-1/n_\Delta)) \exp\left(-\frac{n}{n_\Delta}\right)$$

for  $n = 0, \dots, N-1$  with a channel excess delay of  $N = 6$  and a delay spread of  $n_\Delta = 2$ . In view of prior probabilities of both hypotheses assumed to be  $\pi_0 = \pi_1 = 1/2$ , the probability of false alarm (FA)  $P_{\text{FA}}$  and the probability of miss detection (MD)  $P_{\text{MD}}$  result in a probability of error given by  $P_E = \frac{1}{2}(P_{\text{FA}} + P_{\text{MD}})$ . Furthermore, we choose a real-valued signal autocorrelation  $\rho_{SS,\kappa} = \rho_{SS,|\kappa|} = \sigma_S^2 \rho_{|\kappa|}$  and distinguish three cases for  $\boldsymbol{\Sigma}_S$ , namely

- 1)  $\boldsymbol{\Sigma}_S = \boldsymbol{\Sigma}_{S,1}$  for uncorrelated transmission with  $\rho_\kappa = \delta_\kappa$ ,
- 2)  $\boldsymbol{\Sigma}_S = \boldsymbol{\Sigma}_{S,2}$  for correlated transmission with fast decaying autocorrelation function having non-zero elements  $[\rho_0, \rho_1, \rho_2, \rho_3, \rho_4] = [1, 0.5, 0.4, 0.3, 0.2]$  and
- 3)  $\boldsymbol{\Sigma}_S = \boldsymbol{\Sigma}_{S,3}$  for correlated transmission with slowly decaying autocorrelation function having non-zero elements  $[\rho_0, \rho_1, \rho_2, \rho_3, \rho_4] = [1, 0.7, 0.6, 0.3, 0.2]$ .

The average signal-to-noise ratio (SNR) is denoted by  $\bar{\Gamma} = p_{C,\text{tot}} \sigma_S^2 / \sigma_W^2$  and used to characterize the performance of the scheme. Subsequently, we first choose the PU signal  $S_k$  to be Gaussian, but consider the non-Gaussian case in Sect. III-C.

#### A. Probability of false alarm $P_{\text{FA}}$

It is shown in the Appendix that the two matrices  $\mathbf{A}$  in (11) and  $\boldsymbol{\Sigma}_{\mathbf{Y}|H_1}$  in (4) commute, which implies that they are simultaneously diagonalizable [27]. It is shown at the end of the Appendix that the matrices  $\mathbf{A}$  and  $\boldsymbol{\Sigma}_{\mathbf{Y}|H_0}$  in (2) commute as well. As a result, the covariance matrix of the vector  $\tilde{\mathbf{Y}} = \mathbf{V}^\dagger \mathbf{Y}$  will be diagonal under both hypotheses. The covariance matrix of  $\tilde{\mathbf{Y}}$  under  $H_0$  will then be diagonal and given by  $\mathbf{V}\boldsymbol{\Sigma}_{\mathbf{Y}|H_0}\mathbf{V}^\dagger = \sigma_W^2 \mathbf{I}_K$ , which indicates that the

random variables of  $\tilde{\mathbf{Y}}$  are uncorrelated. Consequently,  $L(\mathbf{Y})$  under  $H_0$  can be written as a weighted sum of  $K$  uncorrelated central chi-square distributed random variables  $Z_1, \dots, Z_K$  with two degrees of freedom and variance  $\frac{\sigma_W^2}{2}$  according to

$$L(\mathbf{Y})|_{H_0} = \sum_{k=1}^K \lambda_k Z_k.$$

The characteristic function of  $L(\mathbf{Y})|_{H_0}$  can be written as

$$\begin{aligned} \psi_{L(\mathbf{Y})|H_0}(j\nu) &= \prod_{k=1}^K \psi_{\lambda_k Z_k}(j\nu) \\ &= \prod_{k=1}^K \frac{1}{1 - j\nu\lambda_k\sigma_W^2} \\ &= \prod_{m=1}^M \frac{1}{(1 - j\nu\lambda_m\sigma_W^2)^{r_m}}, \end{aligned} \quad (13)$$

where  $M$  is the number of distinct poles of  $\psi_{L(\mathbf{Y})|H_0}(j\nu)$ . Furthermore,  $r_m$  is the multiplicity of the  $m$ th pole. Based on the partial fraction decomposition, we can write  $\psi_{L(\mathbf{Y})|H_0}(j\nu)$  as

$$\psi_{L(\mathbf{Y})|H_0}(j\nu) = \sum_{m=1}^M \sum_{n=1}^{r_m} \frac{a_{mn}}{(1 - j\nu\lambda_m\sigma_W^2)^n} \quad (14)$$

where  $a_{mn}$  is the coefficient of  $n$ th power of the fraction of  $m$ th pole. For the case of single poles with  $r_m = 1$ , i.e. the poles which appear just once in (13), the coefficient can be calculated as

$$a_{mn} = \prod_{\substack{i=1 \\ i \neq m}}^M \left(1 - \frac{\lambda_i}{\lambda_m}\right)^{-r_i},$$

while for the case of repeated poles with  $r_m > 1$ , i.e. the poles which appear more than once in (13), the coefficient can be written as

$$a_{mn} = \frac{1}{(r_m - n)! (-j\lambda_m\sigma_W^2)^n} \left. \frac{d^{r_m-n}}{d\nu^{r_m-n}} \left[ \prod_{\substack{i=1 \\ i \neq m}}^M (1 - j\nu\lambda_i\sigma_W^2)^{-r_i} \right] \right|_{\nu = \frac{1}{j\lambda_m\sigma_W^2}}$$

Now, the PDF of  $L(\mathbf{Y})|_{H_0}$  can be written as a sum of  $K$  central chi-square distributed random variables  $\Lambda_{mn}$ ,  $m = 1, \dots, M$ ,  $n = 1, \dots, r_m$  with  $2n$  degrees of freedom and variance  $\lambda_m\sigma_W^2/2$  according to

$$p_{L(\mathbf{Y})|H_0}(l) = \sum_{m=1}^M \sum_{n=1}^{r_m} a_{mn} p_{\Lambda_{mn}}(l).$$

Since  $P_{FA} = \int_{\tau_{opt}}^{\infty} p_{L(\mathbf{Y})|H_0}(l) dl$ , we obtain with the cumulative density function (CDF)  $F_{\Lambda_{mn}}(\cdot)$  of  $\Lambda_{mn}$  [28] the sought-for expression

$$\begin{aligned} P_{FA} &= p_{L(\mathbf{Y})|H_0}(l > \tau_{opt}) = \sum_{m=1}^M \sum_{n=1}^{r_m} a_{mn} (1 - F_{\Lambda_{mn}}(\tau_{opt})) \\ &= \sum_{m=1}^M \sum_{n=1}^{r_m} a_{mn} \left( 1 - \frac{2\gamma\left(n, \frac{2\tau_{opt}}{\lambda_m \sigma_W^2}\right)}{\Gamma(n) \lambda_m \sigma_W^2} \right), \end{aligned} \quad (15)$$

where  $\gamma(\cdot, \cdot)$  and  $\Gamma(\cdot)$  are the lower incomplete Gamma function and the Gamma function, respectively.

Note that (15) represents the exact expression for  $P_{FA}$ . The achievable values of this analytic expression are shown in Fig. 1 for the aforementioned three cases  $\Sigma_S = \Sigma_{S,1}$ ,  $\Sigma_S = \Sigma_{S,2}$  and  $\Sigma_S = \Sigma_{S,3}$  and different values of the observation length, namely  $K = 50$  and  $K = 100$ . As can be concluded

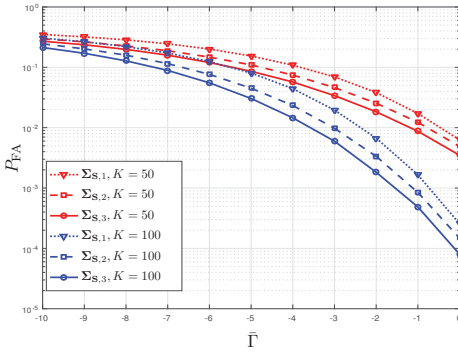


Fig. 1:  $P_{FA}$  for  $K = 50$  and  $K = 100$ .

from Fig. 1, the values of  $P_{FA}$  decrease for increasing correlation of transmitted signal samples. Furthermore, for a given value  $\bar{\Gamma}$  of the average SNR per sample, an increasing  $K$  clearly decreases  $P_{FA}$ .

### B. Probability of miss detection $P_{MD}$

As shown in the Appendix,  $\Sigma_{\mathbf{Y}|H_1}$  and  $\mathbf{A}$  are simultaneously diagonalizable. As a consequence,  $\tilde{\mathbf{Y}} = \mathbf{V}^T \mathbf{Y}$  modeled as a Gaussian vector has independent components. Therefore, for sufficiently large values of  $K$ , we can invoke the CLT again and considering  $L(\mathbf{Y})$  a Gaussian random variable leads to  $P_{MD} = \int_{-\infty}^{\tau_{opt}} p_{L(\mathbf{Y})|H_1}(l) dl$ . It turns out, however, that this approximation is sufficiently accurate only for large values of  $K$ . We therefore consider a numerical evaluation based on simulations of  $P_{MD}$  as a function of  $\bar{\Gamma}$ . The result is illustrated in Fig. 2 for  $K = 50$  and  $K = 100$  for the three different instances of  $\Sigma_S$ .

Again, the sensing reliability in the presence of a PU signal increases for increasing  $K$  due to the higher number of

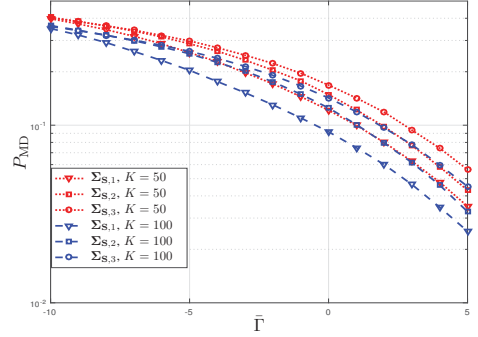


Fig. 2:  $P_{MD}$  from simulation results for  $K = 50$  and  $K = 100$ .

observations leading to a better separation of the distributions of  $L(\mathbf{Y})|H_0$  and  $L(\mathbf{Y})|H_1$  on both sides of the threshold  $\tau_{opt}$ . On the contrary, increasing the correlation in the transmitted signal degrades the performance in terms of  $P_{MD}$ . A comparison of  $P_{FA}$  in Fig. 1 and  $P_{MD}$  in Fig. 2 reveals that for a given value of  $\bar{\Gamma}$ ,  $P_{MD}$  is considerably larger than  $P_{FA}$  so that the overall probability of error  $P_E$  will be dominated by  $P_{MD}$ .

### C. Probability of error $P_E$ for Bayesian and energy detectors

In order to characterize the performance of the proposed Bayesian detector, the values of the resulting  $P_E$  is compared to the performance of the standard sensing algorithm, namely the ED, where the threshold in the latter is chosen numerically to provide the minimum value of  $P_E$ .

As can be seen from Fig. 3, in case of  $\Sigma_S = \Sigma_{S,1}$ , i.e. a white transmit signal  $S_k$ , both schemes show identical performances. This is a direct consequence of (12). Since for a white signal,

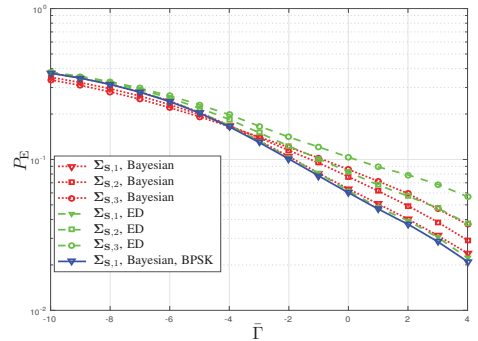


Fig. 3:  $P_E$  of the Bayesian detector in comparison to the ED for  $K = 50$  and Gaussian and non-Gaussian signals.

we have identical singular values  $\lambda_1 = \lambda_2 = \dots = \lambda_K$  which can be absorbed into the threshold so that the Bayesian test indeed corresponds to the ED.

In the case of correlated PU signals, i.e.  $\Sigma_S = \Sigma_{S,2}$  or  $\Sigma_S = \Sigma_{S,3}$ , the performance of the Bayesian detector outperforms the ED, since the former exploits the information about the transmit covariance matrix and the PDP in contrast to the ED. Finally, Fig. 3 shows the performance of a PU signal not being Gaussian distributed. Here, as an example, we consider a binary phase-shift keying (BPSK) signaling to evidence the ability of the proposed algorithm to detect non-Gaussian PU signals reliably.

#### IV. CONCLUSIONS

A novel Bayesian hypothesis test for sensing a primary user signal in a single antenna system and an uncorrelated scattering Rayleigh fading environment with an exponential power delay profile is derived. The central limit theorem is used to derive the optimum threshold for the test. The Bayesian detector can be applied in cases of both Gaussian and non-Gaussian primary user signals. The probability of false alarm of the proposed algorithm has been derived and the probability of miss detection and the probability of error are studied based on simulations. The probability of error achievable by the Bayesian detector is considerably lower as compared to the energy detector which ignores the correlation properties of the primary user signal.

#### APPENDIX

To show that  $\Sigma_{Y|H_1}$  in (5) and  $\mathbf{A}$  in (10) are simultaneously diagonalizable, we need to prove that they commute, which means  $\mathbf{A}\Sigma_{Y|H_1} = \Sigma_{Y|H_1}\mathbf{A}$ . First, we have

$$\begin{aligned}\Sigma_{Y|H_1}\mathbf{A} &= \sigma_W^2(\mathbf{B} + \mathbf{I}_K)\mathbf{A} = \sigma_W^2\mathbf{B}\mathbf{A} + \sigma_W^2\mathbf{A} \\ \mathbf{A}\Sigma_{Y|H_1} &= \mathbf{A}\sigma_W^2(\mathbf{B} + \mathbf{I}_K) = \sigma_W^2\mathbf{A}\mathbf{B} + \sigma_W^2\mathbf{A}.\end{aligned}$$

That implies that we need to check whether  $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ . These both expressions are given by

$$\begin{aligned}\mathbf{A}\mathbf{B} &= -\frac{1}{\sigma_W^2}(\mathbf{B} + \mathbf{I}_K)^{-1} - \mathbf{I}_K)\mathbf{B} \\ \mathbf{B}\mathbf{A} &= -\frac{1}{\sigma_W^2}\mathbf{B}((\mathbf{B} + \mathbf{I}_K)^{-1} - \mathbf{I}_K).\end{aligned}$$

Thus, we need to prove that  $\mathbf{B}(\mathbf{B} + \mathbf{I}_K)^{-1} = (\mathbf{B} + \mathbf{I}_K)^{-1}\mathbf{B}$ . The both expressions are given by

$$\begin{aligned}\mathbf{B}(\mathbf{B} + \mathbf{I}_K)^{-1} &= ((\mathbf{B} + \mathbf{I}_K)\mathbf{B}^{-1})^{-1} = (\mathbf{B}^{-1} + \mathbf{I}_K)^{-1} \\ (\mathbf{B} + \mathbf{I}_K)^{-1}\mathbf{B} &= (\mathbf{B}^{-1}(\mathbf{B} + \mathbf{I}_K))^{-1} = (\mathbf{B}^{-1} + \mathbf{I}_K)^{-1},\end{aligned}$$

which means that  $\Sigma_{Y|H_1}$  and  $\mathbf{A}$  indeed commute and are thus simultaneously diagonalizable. In the same way  $\Sigma_{Y|H_0}$  and  $\mathbf{A}$  are simultaneously diagonalizable, since they obviously commute according to

$$\Sigma_{Y|H_0}\mathbf{A} = \sigma_W^2\mathbf{I}_K\mathbf{A} = \sigma_W^2\mathbf{A} = \mathbf{A}\Sigma_{Y|H_0}$$

- [1] Carter, Kenneth R. and Lahjouji, Ahmed, and McNeil, Neal, and United States., *Unlicensed and unshackled: a joint OSP-OET white paper on unlicensed devices and their regulatory issues / Kenneth R. Carter, Ahmed Lahjouji, Neal McNeil*. Federal Communications Commission, Office of Strategic Planning and Policy Analysis Washington, DC, 2003.
- [2] McHenry, M.A., "NSF spectrum occupancy measurements project summary," *Shared Spectrum Company*, 2005.
- [3] Ian F. Akyildiz and Won-Yeol Lee and Mehmet C. Vuran and Shantidev Mohanty, "Next generation/dynamic spectrum access/cognitive Radio Wireless Networks: A Survey," *Computer Networks Journal (ELSEVIER)*, vol. 50, pp. 2127–2159, 2006.
- [4] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug 1999.
- [5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb 2005.
- [6] A. Goldsmith and S. A. Jafar and I. Maric and S. Srinivasa, "Breaking Spectrum Gridlock With Cognitive Radios: An Information Theoretic Perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.
- [7] D. Cabric and S. M. Mishra and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, vol. 1, Nov 2004, pp. 772–776 Vol.1.
- [8] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, First 2009.
- [9] E. Axell and G. Leus and E. G. Larsson and H. V. Poor, "Spectrum Sensing for Cognitive Radio : State-of-the-Art and Recent Advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, May 2012.
- [10] Zeng, Yonghong and Liang, Ying-Chang and Hoang, Anh Tuan and Zhang, Rui, "A Review on Spectrum Sensing for Cognitive Radio: Challenges and Solutions," *EURASIP J. Adv. Signal Process*, vol. 2010, Jan. 2010.
- [11] Y. C. Liang and Y. Zeng and E. C. Y. Peh and A. T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, April 2008.
- [12] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.
- [13] Haiyun Tang, "Some physical layer issues of wide-band cognitive radio systems," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005, pp. 151–159.
- [14] M. Oner and F. Jondral, "Cyclostationarity based air interface recognition for software radio systems," in *Radio and Wireless Conference, 2004 IEEE*, Sept 2004, pp. 263–266.
- [15] D. Cabric and A. Tkachenko and R. W. Brodersen, "Spectrum Sensing Measurements of Pilot, Energy, and Collaborative Detection," in *MIL-COM 2006 - 2006 IEEE Military Communications conference*, Oct 2006, pp. 1–7.
- [16] Yonghong Zeng and Senior Member and Ying-chang Liang and Senior Member, "Eigenvalue based spectrum sensing algorithms for cognitive radio," *IEEE Trans. on Communications*, pp. 1784–1793, 2009.
- [17] U. Y. Mohamad and D. Dahlhaus, "Cognitive RAdio sensing based on joint distribution of pseudo WISHart matrix Eigenvalues," in *2014 Wireless Telecommunications Symposium*, April 2014, pp. 1–6.
- [18] Z. Tian and G. B. Giannakis, "Compressed Sensing for Wideband Cognitive Radios," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 4, April 2007, pp. IV-1357–IV-1360.
- [19] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005, pp. 131–136.
- [20] G. Ganesan and Ye Li, "Agility improvement through cooperative diversity in cognitive radio," in *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, vol. 5, Dec 2005, pp. 5 pp.–2509.
- [21] Parsons, J.D., *The mobile radio propagation channel*. Halsted Press, 1992.



- [22] T. Cui and C. Tellambura, "Power delay profile and noise variance estimation for OFDM," *IEEE Communications Letters*, vol. 10, no. 1, pp. 25–27, Jan 2006.
- [23] Y. J. Kim and G. H. Im, "Pilot-Symbol Assisted Power Delay Profile Estimation for MIMO-OFDM Systems," *IEEE Communications Letters*, vol. 16, no. 1, pp. 68–71, January 2012.
- [24] H. V. Poor, *An Introduction to Signal Detection and Estimation*, ser. Springer texts in electrical engineering. Springer-Verlag, 1988.
- [25] Horn, R.A. and Johnson, C.R., *Matrix Analysis*. Cambridge University Press, 1990.
- [26] Yanai, Haruo and Takeuchi, Kei and Takane, Yoshio, *Singular Value Decomposition (SVD)*. New York, NY: Springer New York, 2011, pp. 125–149.
- [27] Keith Conrad, "The Minimal Polynomial And Some Applications." [Online]. Available: [www.math.uconn.edu/~kconrad/blurbs/minpolyandappns.pdf](http://www.math.uconn.edu/~kconrad/blurbs/minpolyandappns.pdf)
- [28] Walck, Christian, *Hand-book on Statistical Distributions for experimentalists*. Privately Published, dec 1996.

# Multi-current doubling used in multi-band wireless energy harvesting

Chen Xinying, Wu Muqing, Zhao Min

Beijing Key Laboratory of Network System Architecture  
and Convergence Beijing University of Post and Telecommunications  
Beijing 100876, P.R.China

Email: {chenxinying, wumuqing, zhaomin}@bupt.edu.cn

**Abstract**—Ambient RF energy harvesting has gain sufficient attention, on account of energy constrained in battery-less and wireless status. The attraction of ambient RF energy harvesting has generated massive research on it. Nevertheless the previous RF energy harvesters have concentrated on narrow-band or tunability, which has the restrain of universal implement as well as the inefficiency. In spite of the possibility of attaining higher voltage, the harvested current is conventionally inadequate. This paper present an approach to wide-band harvesting with multi-current enhancement. In this scenario, a wide-band antenna followed by several parallel wide-band voltage doubling modules is designed. The mentioned parallel modules can acquire sufficient current. In order to accommodate the wide-band harvesting the voltage doubling modules are comprised of several narrow-band rectifier chains. The scheme mentioned above has been verified valid in energy efficiency.

## I. INTRODUCTION

Electronic products are ubiquitous in our daily life, and have been extremely related to our life [1][2]. Nevertheless the routine products have developed rapidly, they still are constrained by the lifetime of battery. Hence the kernel is to figure out the solution of fast charging [3] or inventing a battery which has the ability of larger capacity. Otherwise we can find out a way to realize wireless charging. All of the ways we mentioned above will have tremendous contribution. However, the fast consumption of battery as well as the inefficient of current wireless charging [4] throws crippling roadblocks to scientific development.

These years, numerous work relevant to wireless energy has been developed [5]. These proposal scenarios give light to the solution of energy supply. There are unlimited potential of wireless energy harvesting [6][7] in our lives. It can not only realize the rational use of energy, but also escape the inconvenience of the wires of the devices as well as conquer the limitation of the products' life.

Most of the existing job is about narrow-band RF harvesting. That kind of job can extract power from ambient environment, which only covers very narrow spectral band. Even though that solution can solve the power supplying problem, it has its limitation. The narrow band harvesting can only fit in the particular situation. Once the using scene has been changed there are a great probabilities that the harvester can no longer get enough power from the ambient space, or even shut down. The failure is caused by the impedance

mismatching, such as change the city it has been implemented or plans to be implemented.

Some of the research is about wide band power harvesting [8]. This kind of scheme includes wide-band filter, multi-band harvester and tunable harvester.

The wide band harvester can scan wide band RF and harvest power from ambient. This method can cover wide band energy. Yet despite all that, there are disadvantages. For example, there is tradeoff between band width and circuit gain. In this proposal, with the band width increasing the gain of the circuit must reduce, which will give rise to the reduction of the power harvest, and also result in the low efficiency.

Most of the multi-band harvesters [8][9] consist of array antenna. Each antenna tunes to a specific band which it is interested in. The antenna provides power to an independent rectifier, through a tuned matching network. Ultimately the output DC voltages are cascaded together. Some other kinds of multi-band harvesters use a wide band antenna to feed a trunk node of parallel narrow band harvesters which are tuned in orthogonal frequency. The multi-band harvester still has its bootstraps, such as most of the voltage combination is serial which will cause the shortage of current.

The tunable harvester [10] can tune the circuit to match the frequency automatically which allows the system to adjust to any situation dynamically. Nevertheless the tunability settles the adjustment problem, there is still a fatal shortage which is called the cold-start problem. In other words, it still faces the efficiency problem which is the fact that it is still a narrow band system.

Our work aims to produce a much more efficient multi-band harvester, which only needs an antenna and can get more voltage gain as well as current gain. The state-of-the-art researches always aim at increasing the voltage gain such as [11]. We use a wide band antenna, followed by a truck node and M parallel tuned band-pass filter, to collect wide band signal from ambient space. And each branch is then followed by an impedance matched rectifier and then followed by N times doubling circuits. Finally they parallelly connect together. This structure can produce N times voltage doubling and M times current doubling.

In the remaining paper, we will present the single antenna harvester with N-times voltage and M-times current doubling. Then a strategy will be discussed and a solution will be given.

Simulation for the efficiency of two matching network and the influence of series of voltage multiplier as well as the relationship between impedance and series will be performed.

II. M-TIMES CURRENT HARVESTING

Considering that the amount of harvested energy is extremely constrained, we use a super-capacitor to store the energy. The capacitor is used as a charge pump. However only the charge pump and the basic voltage doubling circuit are far from our target, in addition we adopt M times current doubling circuit to gain higher current.

Our prototype consists of K multiple-voltage rectifier, which can cover K frequency band that is adjacent to each other. Each band consists of a filter follows a N times voltage doubler. Take Band1 as an example, Band1 focus on specific frequency  $f_1$ , and the output voltage is three time as big as the input. Because there are K bands, if the frequency satisfies all the K bands, the output current will be K times greater than each band's output. The above-mentioned structure is illustrated as Figure 1. The scenario we mentioned above enables us to make the voltage  $N \times K$  times as big as the origin AC we received, without thinking of the cost of the component.

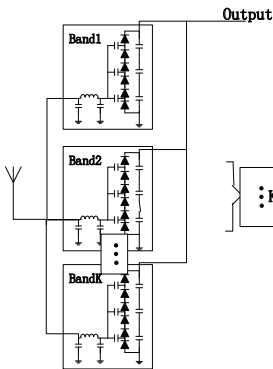


Fig. 1: rectifier that can cover K adjacent frequency band, each band can double the voltage by 3 times, further more the rectifier we proposed can double the final output current by K times

The scenario of  $N \times K$  voltage doubling is used in previous research, even though we can make the voltage much higher than the input of the rectifier. Before the use of the circuit as Figure 2 shows, most of wireless harvesters use the prototype of Figure 3, which is usually gain higher voltage by using more series voltage doubling circuits. Nevertheless increasing the series of the rectifier can increase the value of the output, which is nonlinear. At first the output increase fast. With the series increasing, the output voltage increasement slows down at last or even worse it declines. The simulation will operate in the simulation part.

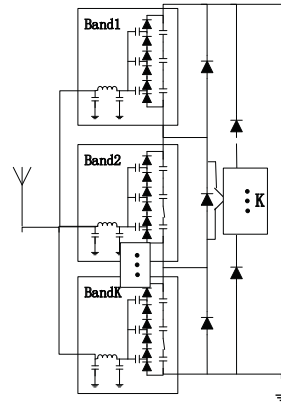


Fig. 2: rectifier that can cover K adjacent frequency band, each band can double the voltage by 3 times, the output voltage can be doubled by  $3 \times K$  times

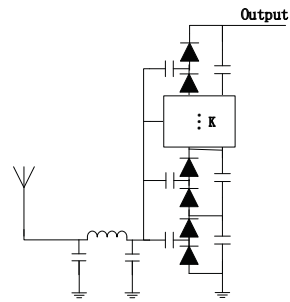


Fig. 3: gain higher voltage by increase the value of K

The reason that the multiple-voltage circuit is nonlinear is that the energy is consumed by the capacitor and the diode we used in rectifier. The diode we used in the rectifier is HSMS-2852, whose equal module is shown in Figure 4. The  $R_j$  is junction resistance, the  $C_s$  is junction capacitor, the  $R_s$  is parasitic resistance, and the  $L_p$  is parasitic inductance, the  $C_p$  is parasitic capacitor. The impedance of HSMS-2852 will cause power consumption, that's why we don't merely use the multiple-voltage circuit to gain higher voltage, in turn we use K different frequency N-time doubling rectifiers to get power from different spectral of RF. The series output of the rectifiers, is shown in Figure 3.

By applying the structure of parallel multi-band harvester our work outperforms the existing works which consist of parallel narrow band rectifier with the series combination on the voltage output. This structure enables us to gain more energy in the same time, which will cover the extra cost

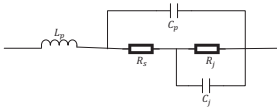


Fig. 4: equal module of HSMS-2852

of the component, such as the cost of leakage current that cannot be avoided. Whereas the published work has realized the opportunity of harvesting the multi-band RF energy, most of them use the multiple antenna. Authors in[11] rose the opinion of using wide-band antenna followed with the parallel narrow band rectifier which covers orthogonal narrow band frequency. Compared with the existing job, this paper has the advantage of energy and charging time efficiency. As a result this solution can not only covers a wide band frequency, acts as a super energy collector, but also gets more current (which can make up the problem in the wireless harvesting, that is the energy is very less as the current is low too), by using only one antenna.

The reason that we cannot keep increasing the series is that the diode will consume extra energy. Without the particular choosing series we cannot get the most efficient system. In the multi-band combination, considering that each narrow band harvester may not in the active mode, as a consequence cascade the voltage directly will reach the roadblock. Once there is one block inactive the whole system will powered off. As a result we use shortcut diode to conquer the band unexcited problem, which makes the system to choose an optimal scenario to overcome the the issue that the harvester bands are not all excited. Also considering that the energy we collect is limited, we choose the low threshold and low leakage diodes. Working in DC, the attributes of diode are easy to achieve.

### III. DESIGN STRATEGY OF THIS HARVESTER

This harvester consists of a single wide-band antenna which is used to collect energy from ambient space without considering the frequency. After the antenna follows M multiple-voltage rectifiers and each multiple-voltage rectifiers consist of K parallel N voltage doubling, which leads us to consider the matching network, as well as the optimal value of N (we have mentioned that the output voltage won't keep increasing as the number of series increase). Since the final output of the energy will supply a computing system, we need to locate an efficient voltage regulator chip to control the harvested power.

#### A. Matching network

The impedance matching network is responsible for the impedance transformation of antenna's output and the rectifiers' input. The perfect matching impedance of the network can reduce the power loss, and increase the voltage gain. On the other hand, the Impedance Matching Network has another effect of amplifying the voltage. The magnification of the matching network is determined by equation (1)

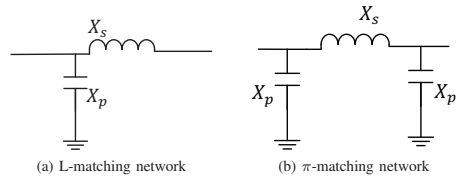


Fig. 5: Energy distributions of different algorithms

$$\frac{V_{out}}{V_{in}} = \frac{1}{2} \sqrt{1 + Q^2} \quad (1)$$

Where the  $V_{out}$  is the output voltage of the matching network,  $V_{in}$  is the input voltage of the matching network, and  $Q$  is the quality factor of the matching network. From the equation (1) we learn that we can obtain higher voltage gain by enlarging the value of  $Q$ . However it needs to be noted that even if we can gain higher voltage by increasing the  $Q$ , it still result in the reduction of the active band width. The relation between the band width and the value of  $Q$  is:

$$Q = \frac{f_c}{\Delta f} \quad (2)$$

Where  $f_c$  is the center frequency,  $\Delta f$  is the band width of the circuit.

By measuring the impedance of the rectifiers, we know that the impedance is complex impedance, and the impedance of the antenna is  $50\Omega$ . Thus we need the impedance matching circuit to realize the matching opportunity.

There are two kinds of matching networks, one of which is  $L$ -matching network as shown in Figure 5(a), and the other is  $\pi$ -matching network as shown in Figure 5(b). The difference between two networks is that the value of  $Q$  in  $L$ -matching network is fixed, while is variable in  $\pi$ -matching network. Because of that the  $Q$  value can be changed, we can optimize the output of voltage in  $\pi$ -matching network to gain higher output voltage, while is impossible in  $L$ -matching network.

#### B. The value N of series circuit

The conventional voltage doubling circuit is shown in Figure 6(a). As is shown, when the input RF signal in the negative half cycle, the circuit runs like Figure 6(b). If the input amplitude of signal  $V_{pk}$  is higher than the turn-on voltage  $V_{on}$  of  $D_1$ , the current will flows through  $D_1$  to charge  $C_1$ , meanwhile the right hand of  $C_1$  is the positive potential. As a consequence the voltage of  $C_1$  is  $V_{pk} - V_{on}$ . When the input RF signal in the active half cycle, the circuit runs like Figure 6(c), and the voltage  $V_{out}$  of  $C_2$  is  $2V_{pk} - 2V_{on}$ . Assuming  $V_{pk}$  is constant,  $V_{out}$  can be increased by reducing  $V_{on}$ , that's why we use schottky diode in rectifiers.

Hence in Figure 6(a) the output voltage is  $2V_{pk} - 2V_{on} + V_{ref}$ , where  $V_{ref}$  represents the earth potential. The multiple voltage rectifier takes the approach of changing the earth

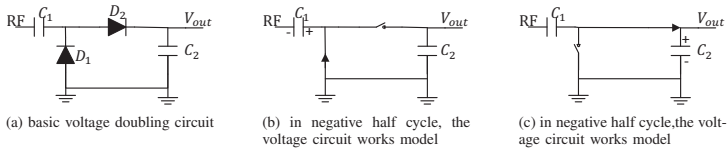


Fig. 6: Energy distributions of different algorithms

potential, and each output DC voltage is taken as the offset of next stage. Consequently the output voltage of the circuit is

$$V_{out} = 2N \times (V_{pk} - V_{on}) \quad (3)$$

From function (3) we can draw a conclusion that the output voltage can be enlarged by increasing the value of N. However, with the increase of the circuit's stage, more and more energy is consumed by diodes, PCB wires and capacitors. Consequently the value of N is important. In our test we set the stage as 3.

C. The regulator part

Considering that the energy we receive is constrained, we need a super capacitor to store the energy we harvested. When it is enough to power the computing system, the computing system will be started. As a consequence, an "on-off" control module is introduced into the system.

We use the BQ22504, which is an Ultra-Low Power Boost Converter with Battery Management for Energy Harvester Applications, designed by Texas Instruments. Once the  $V_{in}$  is higher than 80 mV, BQ22504 can continuously harvest energy. BQ22504 has ultra-low quiescent current  $I_Q$  less than 330 nA and low cold-start voltage ( $V_{in} \geq 330$  mV). We use BQ22504 to control the charge pump to supply or disconnect the computing system, which is more efficient than the exist research.

IV. SIMULATION

The result needs to be tested before implementation, as a result we construct a simulation to test the feasibility of this scenario. Our experiment tests many quota of RF harvesting, which include efficient of two matching network and the influence of series of voltage multiplier as well as relation between impedance and series.

Firstly, there are two kind of matching networks. The simulation is carried on the same circuit with different matching network structures, one of which is L-matching network and the other is  $\pi$ -matching network. Both network structures are shown in figure 6(a) and figure 6(b). Giving the same input power of -27.7dBm,  $V_{out}$  represents the output voltage of the voltage doubler and V-match represents the output voltage of the matching network. As we can see from result contrast shown in Figure 7, indeed the output voltage of  $\pi$ -matching network is higher. The reason is that the Q of  $\pi$ -matching network can be change to a proper value, to make the output

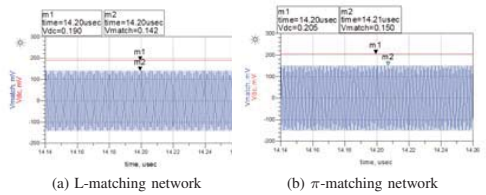


Fig. 7: input power is -27.7dBm, except for the matching network, others are the same.  $V_{dc}$  represent the output voltage of rectifiers, the result shows the-match network's output voltage is higher

voltage higher. As a consequence we use  $\pi$ -matching network in this work.

Secondly, the number of N means the times the voltage will be doubled. Nevertheless, it is not the greater the better. Thus we simulate the impedance of different value of N, the size of doubling times. The data is shown in TABLE I, which presents the impedance of different N. And TABLE II shows the output voltage of different N with input power of -20dBm and -10dBm.

From the TABLE I we can see that with the increase of N, the impedance decreases, resulting in the difficulty of impedance matching increasing. As a result, the series of the voltage multiplier should be reduced. Also from TABLE II we can see that with the same input power, as the series of the voltage multiplier increase, the output voltage goes higher too. Consequently, we can make a conclusion that there is a trade-off between the simple degree of matching impedance and the series of voltage multiplier.

In consideration of this situation, we set N as 3 and M as 2, which means voltage is triple doubled, and current is twice doubled.

V. EXPERIMENT RESULT

The simulation shows that the scenario is feasible. However the real situation can be more changeable. The leakage current and parasitic capacitance as well as the parasitic inductance can be more unpredictable in that frequency. Especially the leakage current will result in the failure of this work. To test the theory we make a PCB model to collect the energy. Considering that the main theory in this paper is the increase of the output current, we only test the narrow band energy

TABLE I: The impedance simulation of voltage multiplier with N series

series N	impedance $\Omega$
1	30.627-j356.514
2	15.332-j181.486
3	10.217-j121.716
4	7.656-j91.534
5	6.128-j73.364
6	5.103-j61.211
7	4.374-j52.51
8	3.827-j45.977
9	3.403-j40.891
10	3.062-j36.816

TABLE II: The simulation output of voltage multiplier with N series

Series N	Input Power(-20dBm)	Input Power(-10dBm)
1	631.6mV	1.415V
2	937mV	2.345V
3	943mV	2.926V
4	1.049V	3.216V
5	1.072V	3.519V
10	1.012V	3.928V

harvesting. The structure of experiment model is shown in Figure 8, and the circuit is shown in Figure 10. In this model we use triple voltage multiplier and twice current multiplier, followed by a big capacitor which can store enough energy to lighten the LED. In the vital control part, we use chip BQ25504 to control the voltage supply. This chip has ultra low current-leakage as well as ultrashort cold start time .



Fig. 8: The real prototype of triple double the voltage and twice double the current

The structure of control group is shown in Figure 9. Under the same condition, no matter how many series the voltage multiplier goes it still can't lighten the LED. According to the test, without the current doubling even though the voltage is enough to drive the voltage regulator chip, the current is too

less to drive the system.

The result is obvious, the control group cannot drive the computing system, while the experiment group can lighten the LED in every 10 minutes.

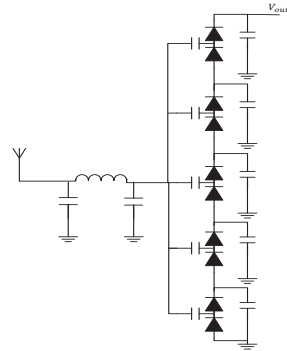


Fig. 9: The real prototype of triple double the voltage and twice double the current

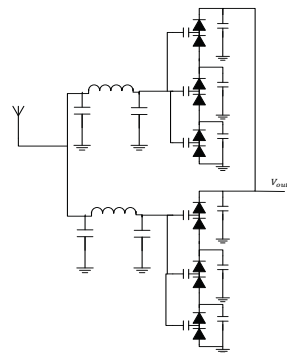


Fig. 10: The real prototype of triple double the voltage and twice double the current

VI. CONCLUSION

The voltage and current multiplier presented this paper gave us the benefit of power the system with lower energy. As the system can be driven by lower power, we can use the energy more efficiently without inefficiency due to leakage current.

The future work will focus on the efficiency of the energy harvesting. The main point is to shorten the leakage current consumption.

ACKNOWLEDGMENT

This work is supported by Beijing Laboratory of Advanced Information Networks.

## REFERENCES

- [1] A. Parks, A. Sample, Y. Zhao, and J. Smith. *A wireless sensing platform utilizing ambient rf energy*. in Radio and Wireless Symposium (RWS), 2013 IEEE, pp. 331-333, 2013.
- [2] A. P. Sample, A. N. Parks, S. Southwood, and J. R. Smith. *Wireless ambient radio power*. in Wirelessly Powered Sensor Networks and Computational RFID (J. R. Smith, ed.), pp. 223-234, Springer New York, 2013.
- [3] Liu, Zhao Verhallen, Tomas W. Singh, Deepak P. *Relating the 3D electrode morphology to Li-ion battery performance; a case for LiFePO4*. JOURNAL OF POWER SOURCES pp. 358-367, 2016.
- [4] Cove, Samuel R. Ordenez, Martin, Shafiei, Navid. *Improving Wireless Power Transfer Efficiency Using Hollow Windings With Track-Width-Ratio*. IEEE TRANSACTIONS ON POWER ELECTRONICS pp. 6524-6533, 2016.
- [5] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith. *Ambient backscatter: wireless communication out of thin air*. in Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM, pp. 39-50, ACM, 2013.
- [6] J. Koomey, S. Berard, M. Sanchez, and H. Wong. *Implications of historical trends in the electrical efficiency of computing*. Annals of the History of Computing, IEEE, vol. 33, no. 3, pp. 46-54, 2011.
- [7] J. R. Smith. *Range scaling of wirelessly powered sensor systems*. in Wirelessly Powered Sensor Networks and Computational RFID (J. R. Smith, ed.), pp. 3-12, Springer New York, 2013.
- [8] K. Niotaki, S. Kim, S. Jeong, A. Collado, A. Georgiadis, and M. Tentzeris. *A compact dual-band rectenna using slot-loaded dual band folded dipole antenna*. in Antennas and Wireless Propagation Letters, IEEE, vol. 12, pp. 1634-1637, 2013.
- [9] S. Keyrouz, H. Visser, and A. Tjhuis. *Multi-band simultaneous radio frequency energy harvesting*. in Antennas and Propagation (EuCAP), 2013 7th European Conference on, pp. 3058-3061, 2013.
- [10] G. Seigneuret, E. Bergeret, and P. Pannier. *Auto-tuning in passive uhf rfid tags*. in NEWCAS Conference (NEWCAS), 2010 8th IEEE International, pp. 181-184, 2010.
- [11] Parks, A. N. and J. R. Smith. *Sifting through the airwaves: Efficient and scalable multiband RF harvesting*. in IEEE International Conference on RFID (IEEE RFID), 2014 IEEE.

# Cyber Security Algorithms For Distributed Denial Of Service Attack In Botnet

Karuna S. Bhosale

Faculty of Telecommunication  
Technical University of Sofia  
Bulgaria

e-mail: bhosale.karuna@gmail.com

Maria Nenova, Georgi Iliev

Faculty of Telecommunication  
Technical University of Sofia  
Bulgaria

e-mail: mvn@tu-sofia.bg, gli@tu-sofia.bg

**Abstract:** - The Botnet has evolved to be most deadly threat to internet, so also the various attacks caused by it are dreadful hence detection and prevention had motivated many researchers to come up with many new and advance technologies. An overview of healing techniques for Distributed Denial of Service attack is proposed in the paper. The aim is gathering of information related to the malicious computer's user, after attack of black holing. Distributed denial of service attack is also termed as Black hole attack. To achieve early detection and stopping of distributed flooding attacks in network the paper analyses a framework based on CUSUM algorithm, Machine Learning Techniques, MULTOPS, which provides detection and more accuracy, simulation with high rate and with minimum disturbance.

**Keywords-** Botnets, DDoS, CUSUM algorithm, Machine learning, MULTOPS.

## I. INTRODUCTION

As Botnet has evolved to be most deadly threat to internet, the various attacks caused by it are dreadful, hence detection and prevention had motivated many researchers to come up with many new and advance technologies. In this paper different healing techniques for Distributed Denial of Service attack are overviewed. Here information related to the malicious computer's user, after attack of black holing is collected. Distributed denial of service attack is also termed as Black hole attack. This information may be Telephone Number, Address, Name, or other financial sensitive information. To achieve early detection and stopping of distributed flooding attacks in network the paper analyses a framework based on CUSUM algorithm which provides detection and more accuracy, simulation with high rate and with minimum disturbance. Some of the listed solutions here are efficient and could help to recover computers back to the system.

Nowadays Internet is being used on daily basis by millions of people all over the world. As internet is providing numerous amounts of potential opportunities, it has been attracting more and more people towards it. But as any new invention has its pros and cons, Internet

also has harmful and harmless users. They operate in Internet with various malicious attacks. Among them Botnet is one of the most dangerous attack responsible for huge volume of malware activity such as Distributed Denial of Service (DDoS), Spamming, Phishing, Spoofing, Scareware etc.

Botnet is the network of numerous computers called 'Bot', controlled by Botmaster. Botmaster has its command on Botnet through IRC, HTTP, Command and Control Channel. As attack of Botnet had proved to be hazardous giving rise to detection and healing techniques for Botnet. The detection techniques can be classified based on type of attacks from Botnet. They are namely Honeypot, Passive anomaly analysis, Signature based methods, DNS based and Mining based. Some of the detection techniques are been most successful in the attempt while other techniques had some flaws.

With the advent of computers and computer networks, the exploitation of information at unprecedented speed is available at all possible levels such as individuals, organizations and also governmental levels. Originally, the meaning of word 'hacker' was quite assertive in terms; hacker was a person who combined hardware and software to maximize the limits of technology. But gradually the criminalization of hacking had led to the decay of word's original meaning. As the design of internet was considered explicitly without considering its security it has become frequent medium for cyber criminals.

A Botnet comprises of a Bot master, Zombie pc's and Command and Control (C & C) which relays commands to the bots and report to the botmaster. Since July 2009, the various Botnets emerged are Forbot, PBot, ToxBot, MACHbot, PHP loot, among these Egg Drop is the first Botnet. The major activity of these Botnet is to send spam e-mail, conduct DDoS (Distributed Denial of Service), and steal sensitive information such as, personal information, credit card numbers and so on. The motive behind these activities is the financial benefit.



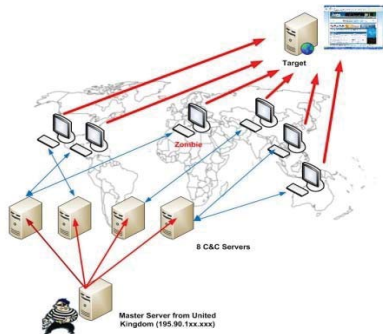


Fig. 1 Model of Botnet attack

The above figure depicts the typical Botnet attack scenario. Here, the attacker often termed as ‘Botmaster’, makes use of various hidden channel including protocols like IRC, HTTP and also popular sites such as Face book, Twitter, etc. With the help of these medium, Botmaster infects the nodes which are then called as ‘Zombies’ Several in number also termed as ‘Zombie army’ Zombies in turn infects the targeted machines through Trojan Horse or Viruses. Generally these targeted machines are not well equipped by strong firewalls. Hence they fall prey to such Botnet attack.

## II. METHODOLOGY

The paper is concentrated on healing the Distributed Denial of Service (DDoS) attack on application layer. With the help of combination of CUSUM Algorithm, Machine Learning Techniques such as decision tree or Naïve Bayesian classifier, Heuristic and Data mining mechanism such as MULTOPS, the solution for above stated attack is proposed. Here a framework, based on history and statistical analysis, for early detection and stopping of distributed flooding attacks in network is proposed which provides faster detection and more accuracy, simulation with high rate and with minimum disturbance.

The analyzed detection framework is based on basic attributes of most netflow formats include flow start and end time, source and destination IP addresses, source and destination ports, number of bytes and packets in the flow, as well as protocols used. Based on netflow attributes, following heuristics are defined:

- 1) Unidirectional flow of data.
- 2) Diversity in distinct source and destination ports by protocol and destination host.
- 3) Diversity in bytes per packet and packets per flow.

Three different feature selection algorithms are discussed:

- 1) Correlation based feature selection.
- 2) Consistency based Subset Evaluation.
- 3) Principal Component Analysis.

There are three different algorithms on three different machines learning techniques, namely Decision trees, Naive Bayes Classifier, and Bayesian network Classifier. The three key findings of these algorithms are:

- 1) The build time of the model with full feature set differs when compared to the build time of the models.
- 2) By reducing the no of feature required for classification and feature selection.
- 3) Using the full feature does give a higher accuracy. Correlation based feature selection (CFS) is a simple filter method. A good feature subset contains features highly correlated with (predictive of) the classification, yet uncorrelated with (not predictive of) each other. The subset giving the highest merit is of the CFS.

Consistency based Subset Evaluation (CSE) search the feature subsets & find an optimal subset of relevant features that are consistent to each other. The overall consistency of a subset is calculated using Inconsistency ratio (IR). Principal Component analysis (PCA) is not a feature selection but a feature reduction technique .It aims at reducing the dimensionality of the dataset. To measure the performance of a classification algorithm, different metrics can be used. The metrics employed here are as follows:

Accuracy of the classification is the ratio of sum of true positive (TP) and True negatives (TN) the sum of TP, FP, and TN and FN.

- 1) Detection Rate: It is ratio of True positive to Sum of TP and FN.
- 2) Normalized Build Time: The build time for the classification algorithm normalized over the range 0 to 1.
- 3) Normalized classification speed: Classification speed normalized over the range 0 to 1.

## III. COMPARATIVE ANALYSIS

One way to defend against botnets is to detect the presence of botnets Command and Control communication. After detection of C and C traffic, it must be prevented from reaching the victims, tract the C and C servers and shut down the systems. Systems like BotSniffer and BotMiner, detect synchronized activity coming from multiple, infected hosts in a network

Another approach is Signature Based detection of an infected host even a single connection matching one of their signature. Thus all the extracted strings are given rank (score), so the string receives high score when it likely capture C and C traffic, to a low score otherwise. This methodology, provide a novel approach for automatically extracting strings

signatures that captures Botnet C and C message. Evaluation of the ranking system is done by manually looking into the ranked list of candidate signatures. It focused on 100 entries from the top, the middle, and the bottom respectively. To understand whether the strings are meaningful or not, manually the payloads were matched by each sting. With the help of expert knowledge the judgment of whether the string is malicious, benign, or suspicious is made [1]. A Novel detection approach based on Control Flow Stability is used. On comparing stability of Storm bots to that of normal bots peer to peer clients, a stability detection algorithm is developed. The new emerging option uses P2P as C and C. Here no control server is present. The Botnets only joins the network as another peer and send the command to other peers. Several P2P based Botnets such as Sinit, Nugache; Storm Worm may choose different protocols. This paper proposed the C and C principle of most widespread storm worm Botnet and provides conception of Control flow stability. The analysis there is based on flow records. A flow is sequence of packets and flow records consists of addressing size of packets and bytes and information of flow. With the application of stability detecting method, the discrimination of Botnet flows from other P2P flows the accuracy and the completeness are consistent with the false negatives and false positives respectively. The P2P network, Botnet are more concealable and robust than traditional centralized Botnet [2].

A detection framework prototype is proposed and tested using real traffic. Two main approach for anomaly based detection, one involving previous network traffic, capturing what is considered as normal behavior and other involving the formal definition of normal or abnormal behavior. A relevant behavior in a bot is observed to be of scanning. the three important bot and Botnet behaviors likely to produce enough networks are C and C information exchanging, upgrading, updating and also attacking [3]. By applying CFS, CSE, PCA the reduced feature sets is obtained by each algorithm. Classification models for smaller subsets of feature give almost equivalent, accurate and detection rate as compared to full feature set. Benefits in terms of computational performance are significant. As, future scope, use of an ensemble of classifiers for robust detection of P2P Botnet traffic, also use of advanced space efficient data structures like Bloom Filters to efficiently extract statistical information from network [4].

It is advantageous that graph based representation of the infected computers allow to use graph partitioning algorithm to separate out the different Botnets when a network is infected with multiple Botnets at the same time. In this reference, the results indicate that framework can isolate Botnets in a network under varying condition with high degree of accuracy [5]. As the importance of controlling Botnets

is increased than ever before, as the malicious traffic from Botnets now threaten the network infrastructure of Internet service providers. Here the authors proposed, an advanced DDoS defense technique for use of ISPs with a real case study is illustrated and it has proven effective for controlling Botnets [6].

Distributed Denial of Service flooding attacks are nowadays major area of concern. DDoS typically exploit user's access to service. It is discussed, an attempt to prevent and where and when to prevent, detect and respond to the DDoS attack has been presented. In this paper authors specified various characteristics, models of DDoS to provide timeline mechanism. Also a mapreducing programming model is used to increase the efficiency [7]. Remotely placed computers are attacked by Botnet, such as DDoS and information Phishing. The authors show different Botnet attacks and defenses. Fast Fluxing and Domain Fluxing techniques are introduced. A comparison and evaluation of these fluxing detection methods are mentioned [8].

Compromised computer with malicious code are remotely controlled by Botnet infected by DDoS, Spamming and Spoofing. Here, a successful strategy for defense against Botnet is presented. The authors propose a scalable technique having self equipped healing system, collaboratively connecting to multiple peers, effectively combats the Botnets [9]. DDoS is the deadliest attack in history of network security. Among different proposed solution, Remote triggered black hole filtering for stopping DDoS is presented. In this reference, an improved framework for BlackHole filtering on the edge of ISPs is mentioned [10]. To avoids most of the computer related attack intrusion detection is versatile security paradigms which can be effectively cure the attack. In this reference proposed framework which is an amalgamation of intrusion detection and prevention technologies is presented [11]. In ubiquitous computing, significant problem have risen due to denial of service for both users and service providers. Also this problem gets more harmful as they behave in distributed denial of service manner. Prevention, detection, and response are involved as basic procedure for defending this attack. For a system to be effective, it should inculcate less detection time, low false positive rate, and low computational overhead. This paper presents TCP congestion window using Cumulative sum (CUSUM) and also uses network simulator platform [12].

To combat the serious attack of DDoS, DDoS test and evaluation is becoming more important. Here authors describe the design and implementation of DDoS attack defense testbed. The Testbed is co-simulation using OPNET and VMware workstation. Finally the experiments results in proving the testbeds to be more effective and convenient evaluation [13]. As the field of internet is widely occupied by wireless

Sensor Network, they became more prone to BlackHole attack. Various nodes are captured to prevent the packets reaching the base station resulting minimum output. In the authors proposed a novel method for measurement of Blackhole on network parameters.[14] The mitigation of DDoS had become a very serious issue. This reference proposed the architecture and algorithm of Firecol. Firecol is Intrusion Prevention Systems at Internet service providers. Firecol is based on simulation and real dataset proving effective and supportive for real networks [15].

It is discussed a promising solution for DDoS to be identifier /locator separation. Here the mapping approach is discussed which the Botnets found to be

uncontrollable. This method is also effective on attacks and zombies [16]. Effect of DDoS result in tremendous damages to users resources. This reference presented a statistical technique to detect and filter DDoS. This model is based on small storage capacity through proves to faster and time consumption were taken into account. This reference had resulted with 97% of accuracy and lesser faults [17]. Apart from Wireless Sensor network, software defined networks also have emerged as wider area for computing. Though Software defined Networks assures of more flexible network, it is surrounded by large no attacks. An Open Flow Protocol for intrusion of DDoS and suggests machine learning based techniques for such attacks are investigated in [18].

TABLE I: Evaluation of the Main Techniques

METHODOLOGY	PERFORMANCE	ANALYSIS	FUTURE SCOPE
TCP Congestion window using the Cumulative Sum(CUSUM). [12]	The CUSUM, C is calculated as $C_n = \max(0, C_{n-1} + y_n - \mu_0 - v)$ where $\mu_1 = \mu_0 + v$	Two metrics to evaluate CWD, false positive rate & detection time.	To evaluate the detection time to improve performance
Co-simulation using OPNET & VMware workstation. [13]	DDoS Testbed and system structure, Model designing & implementation.	1.Network topology module 2.Attack generation 3.Background flow 4.Data acquisition 5. Control & configuration module.	A detailed evaluation index is required, the configuration needed to be more automated.
Framework for black hole filtering on the edge of Internet Service Provider. [14]	RTBH Filtering techniques. Drop packet silently, sender is not informed about packet that did not reach destination.	RTBH forward traffic Null 0.Null 0 is pseudo interface. Detecting mechanism (self-trigger) and interface should exist.	Structure of self-triggered for processing & detecting malicious packet. Recording malicious information for Filter black holing.
Detection & Prevention algorithm is presented. [15]	Detection of BlackHole attacker node. Prevent attack from occurrence.	Build clusters from sensor node. Cluster coordinate is elected. Assign IDs. Router does not reply with response Packet & data Packet to Co, then that router is failed.	Analysis can be done on more topologies other than tree.
Theoretical Foundation, architecture and algorithm of Firecol. [16]	Firecol architecture-Ring based Overlay protection, subscription Protocol, Multiple customers.	A)Firecol metrics:- i)Frequency, $f_i = F_i / \sum F_j$ ii)Entropy = $-\sum f_i \log_2(f_i)$ iii) Relative entropy, $\Psi_i = \log f_i / f_i'$	Extend Firecol to support different IPS rule structures.
Identifier-to-locator mapping approach. [17]	1)Two types of attacks DDoS and DRDoS; 2)LISP (Locator /ID Separation Protocol)separates customer network from Provider network.	Mapping system: Network contains N Mapping system has one or more mapping servers. Identifiers are associated with servers assigning Key.	To evaluate scalability of mapping function for secure distribution
Statistical technique to detect & Filter DDoS attack. [18]	Highly reduces false alarm rates, TTL value extracted computes no. of hops, IP to HOP count is compared with TTL, packet is discarded. Deviation & anomaly detector.	To Compute divergence. $D_i = \sum p(i) \log p(i)/Q_i$ $D_i = 1/2 D (P_i    M_i) + 1/2 D(Q_i    M_i)$ $M = 1/2 (P+Q)$ where $P =$ true distribution of data $Q =$ theory _model or approximation	The employment of cross layer traffic analysis and defense mechanism
Software Defined networking accompanied with Openflow Protocol. [19]	Intrusion detection technology a)Signature detection tech, b)Anomaly detection tech c)Statistical analysis d)Machine Learning	1)Neural network-training & Testing 2)Support Vector Machine 3)Decision Tree 4)Genetic algorithm 5) Fuzzy logic 6)Bayesian network	Network virtualization and middle boxes based on cloud computing

IV. CONCLUSION

In the present paper an overview analysis of useful techniques for avoiding the effects from the attack of DDoS are discussed.

Firstly is investigated the methodology based on TCP Congestion Window which is analyzed using CUSUM algorithm on Network simulator platform. This method considers the parameters such as short detection time, low false position rate, and low computational overhead.

Secondly is analyzed a method based on Machine learning techniques. In this technique Identifier/Locator approach is examined for detection of DDoS attack. This approach makes attackers difficult to hold control on Botnets. This method is also effective on attacks by zombies.

Next Openflow Platform based on Heuristic Mechanism is investigated from the perspective of intrusion and DDoS attack and also to mitigate such attacks. Using defined heuristics, communication fingerprints are created to cluster similar network traffic. To determine the similarity of the detection framework, True positive and true negatives rates as well as false positive and false negatives rates were determined using 24 hours time window dataset. Such a system can be implemented in a real time if framework is able to process all the network hour traffic in less than hour.

By applying CFS, CSE, PCA the reduced feature sets is obtained by each algorithm. Classification models for smaller subsets of feature give almost equivalent, accurate and detection rate as compared to full feature set. Benefits in terms of computational performance are significant.

There are several aspects for next step of investigations, which remain to be done. The packet structure e.g. flow of id, different flags, time to live, delay parameters, comparison between traffic, bandwidth to be tested on Network Simulator Platform (NS2) which can be considered to reduce the effect of DDOS attack in future.

REFERENCES

[1] A. Zand, G. Vigna, X. Yan, Ch. Kruegel, "Extracting probale command and control signatures for detecting botnets", SAC' 14th march 24-28,2014.  
 [2] B. Wang, Z. Li,Hao Tu, Jie Ma, "Measuring Peer-to-Peer Botnets Using Control Flow Stability", International

Conference on Availability, Reliability and Security, 978-0-7695-3564-7/2009, IEEE, 2009.  
 [3] L. Mendonca, H. Santos "Botnets: A Heuristic-based detection Framework", SIN '12 October 23-27, Jaipur, India, 2012.  
 [4] P. Ratik Narang, Jagan mahan Reddy, Chittaranjan Hota " Feature Selection for Detection of Peer -To-Peer Botnet Traffic" COMPUTE'13,AUG 22-24, Tamil Nadu, India, 2013.  
 [5] P. Jaikumar and A. C. Kak,"A graph -theoretic framework for isolating botnets in a network" SECURITY AND COMMUNICATION NETWORKS Security Comm.Networks, 2012.  
 [6] Y.H Moon1,S B Choi1,H k Kim2and C Yoo3, " A Hybrid Defense Technique for ISP Against the Distributed Denial of Service Attacks" Appiel Mathematics & Information sciences An International Journal, 2014.  
 [7] S.Tripathi1, B.Gupta, A.Almoomani, A.Mishra, S.Veluru, "Hadoop Based Defense Solution to Handle Distributed Denial of Service(DDoS) Attacks" Journal of Information Security, 2013.  
 [8] L Zhang,S Yu,Di Wu paul Watters, "A survey on Latest Botnet Attack and Defense" 2011 International Joint conference of IEEE, 2011.  
 [9] Jitin, Mridul, Bhatnagar, S Tiwari, K Kumar, "AN AUTOMATED APPROACH TOWARDS COMBATING BOTNETS", Proceedings of SARC-ITR International Conference, Chennai, May 2014.  
 [10] A Sadeghian,M Zamani, "Detecting and Preventing DDoS Attacks in Botnets by the help of Self Triggered Black Holes" APCASE , Bali, Indonesia. 2014.  
 [11] F.drees, M. Rajaranjan, A.Y.Menon, "Framework for Distributed and Self -Healing Hybrid Intrusion detection and Prevention System",IEEE, 2013.  
 [12] M Alenezi and M J .Reed, "Denial of Service detection Through TCP Congestion Window Analysis" School of Computer Science and Electronics Engineering, UK, IEEE 2013.  
 [13] S.Nthg1,Q Hani, "Design and Implementation of DDoS Attack and Defense Testbed" Zhengzhou Information Science and Technology Institute, China, IEEE, 2012.  
 [14] M Wazid,A,Roshan S Sachan,R H Goudar and D P Singh, " Detection and Prevention Mechanism for Blackhole Attack in Wireless Sensor Network" International Conference on Communication and Signal Processing, India, IEEE April 2013.  
 [15] J. Francois, I Aib and R. Boutaba, "Firecol: A Collaborative protection Network for the Detection of Flooding DdoS attacks" ACM Transactions on Networking, IEEE, Dec 2012.  
 [16] H.Luo, Y. Lin and H. Zhang M. Zukeraman, Beijing Jiaotong University, "Preventing DDoS Attacks by Identifier/Locator separation", city University of Hong Kong, IEEE, 2013.  
 [17] P. Shamsolmoali, M. Zareapoor, "Statistical-based filtering system against DDoS attacks in cloud Computing" Jamia Hamdard University, IEEE India, 2014.  
 [18] J. Ashraf, S. Latif, " Handling Intrusion and DDoS attacks in Software Defined Networks Using Machine Learning Techniques" CSE Dept,MCS, Islamabad, Pakistan, 2014.

# Comparative Analysis of Current Security Techniques Employed in Various Attacks of Bluetooth Communication

Khatod Varsha Ritesh

Department of Electronics and telecommunication engineering  
ViMEET, India  
Varshabht.2@gmail.com

Agata Manolova

Associate professor  
Technical University of Sofia  
Bulgaria  
amanolova@tu-sofia.bg

**Abstract**— Bluetooth technology releases the user from any type of wires and cables. A user can have a laptop, PDA, Tablet PC, or a smart phone and connect to any device that supports the Bluetooth technology in the near vicinity. The convenience that Bluetooth offers has transformed the market from wired to wireless in less than a generation, and has helped user in the era of wearable technology. Yet, there are numerous security risks associated with using Bluetooth—ranging from the ability to eavesdrop on some devices, to the ability to crash devices and drain the batteries on others. In fact, the core specification for Bluetooth Low Energy (LE) states explicitly: "The overall goal of keeping the cost of the Controller and the complexity of a slave device to a minimum was used in making compromises on security capabilities in LE." Various tradeoffs serve as potential roadblocks to making Bluetooth-enabled devices fully secure: vendors do not wish to make their products more difficult to use, shorten their battery life, increase the cost of their manufacture, or make their devices incompatible with other products. Therefore, this paper is motivated to examine the most popular current security vulnerabilities and threats imposed by the inherent open nature of Bluetooth communications and review and compare several latest security techniques, including encryption, public and private keys, authentication protocols etc.

**Keywords**—Bluetooth, Authentication, confidentiality, pairing, encryption

## I. INTRODUCTION

In the year 1994, a team of researchers at Ericsson Mobile Communications started a study of universal short-range and low-power wireless connectivity primarily for exchanging data between mobile phones, headsets, vehicular devices, and computers which ended in a group known as Bluetooth Special Interest Group (SIG). This technology was officially approved in the year 1999 [1]. Since then it has widely been used in various electronic devices as IEEE 802.15.1. The Bluetooth technology drives over the 2.4 GHz Industrial-Scientific-Medical (ISM) unlicensed frequency band which is predominantly for low power transference. In the reach of 10 to 100 meters, Bluetooth can provide 700Kbps, 2.1 Mbps, or up to 24 Mbps data rates depending on its version. Bluetooth security and users ease of operation involves some serious tradeoffs due to which the attacks are more frequent in Bluetooth communication. The most popular attack is blue

jacking, which lets the attacker send anonymous and unsolicited messages to certain Bluetooth phones. Then the blue snarfing, which makes it possible to get access to a phone wirelessly without the owner's knowledge and download the stored phonebook and calendar and many more. More recently, reports have described blue bugging, in which someone can theoretically take complete wireless control of Bluetooth phone and use it for all kinds of illicit purposes.

## II. BLUETOOTH COMMUNICATION

The different key steps involved in Bluetooth communication are authentication, authorization (pairing) and encryption. The communication procedure may be understood with the help of Fig.1.

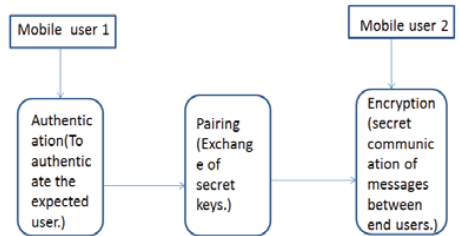


Fig.1 Bluetooth communication process

## III. ATTACKS IN BLUETOOTH COMMUNICATION

There are various types of vulnerabilities present in Bluetooth system. The bouts can be executed at any stage in the communication process. But most popular and hazardous bouts occur during the process of authentication and pairing. The various attacks are enlisted in Fig.2. In Blue jacking the unauthorized user can send unsolicited messages or business cards to Bluetooth-enabled devices like email spam and phishing. In blue bug an attacker can download contacts, call log, even send or receive message and the most hazardous thing is it allows the hacker to almost complete takeover of a

phone. The most hazardous type of active attack is man in the middle [11]. In this type a third person (unintended recipient) will try to watch the communication or modify the messages or pretend to be the intended receiver. The attack can be intruded either during authentication (pretending as intended receiver) or during pairing (modification of messages) [12], [23]. The attacks are classified as per key steps for Bluetooth communication as shown in Fig.2.

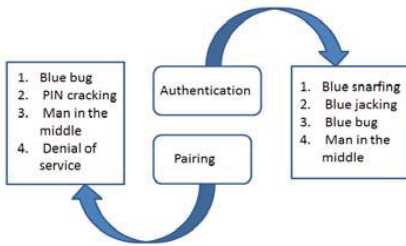


Fig.2 Classification of Bluetooth attacks

IV. COMPARISON OF CURRENT SECURITY TECHNIQUES

It is needed that the protection level of knowledge transmission for Bluetooth communication ought to be inflated. There are several analysis papers addressing this issue through the years. are aiming to concentrate our efforts to match solely papers printed once 2011 and publicly out there.

Rajput et al. [2] present a security mechanism that produces use of Data Encryption Standard (DES) to cypher the information and Rivest Shamir Adleman (RSA) to cypher the Key. RSA formula uses difficult factorization theme hence it is appropriate for encrypting a tiny low quantity of information, distribution and updating of key is relatively simple. Whereas International Data Encryption Algorithm (IDEA) algorithm will encrypt an outsized quantity of information at comparatively a lot of operational speed. In order to use advantages of the two algorithms, the authors have projected a brand new encoding algorithm, based on IDEA, RSA and Message Digest Algorithm (MD5) to unravel the present authentication connected risks of Bluetooth technology.

In [3], the authors have proposed Identity-Based Cryptography (IBC) jointly of the ways that to shield communications between sensible wireless devices within the home. It combines the employment of Identity-Based Cryptography for Wi-Fi and Bluetooth communication. Identity-Based Cryptography may be a form of public-key cryptography where a public piece of information connected to a node is employed as its public key. IBC as a way for implementing secure, useful and efficient schemes in sensible mobile devices.

According to [4] the blue bug is one in every of the damaging kind of Bluetooth attack because Radio Frequency Communication (RFCOMM) is a direct connection between

two devices and Bluetooth RFCOMM has no authentication hence assailant will simply enter the new device and prove direct connection without authentication. The authors have thought of a changed authentication approach for defense of blue bug attack wherever RFCOMM ought to use database for Bluetooth device address or Mac address. After scanning, once match is found then only RFCOMM should give permission to communicate the actual device, otherwise it should not get permission.

In [5] users have proposed a general purpose verification tool named as SPIN to to influence the problem of authentication as a result of contemporary security verification tools are compatible for analysing protocols in isolation however it is not specified a way to use them for protocols intended to be run in more “dynamic” settings. The rules requiring sharing of parameters are used as building blocks in more advanced and compound ways. SPIN finds all the rumored attacks on authentication in terribly short time.

The authors have planned an algorithmic program in [6] for the problem of authentication and pairing. The method makes use of address of each unit concerned in pairing and RSA algorithm is used for the aim of secret writing. Since each unit involve their Bluetooth device authentication (BTDA) in pairing thus third unit cannot interfere throughout pairing and cannot get any data of the act entities.

As per B. Manthan et al. [7] the foremost wide used Secure and Fast Encryption Routine (SAFER+) is prone to some vulnerabilities. The existing SAFER+ algorithm is modified to supply higher data output and frequency with the assistance of some modifications like introduction of Rotation block between each round to modify the problem of authentication, pairing and encryption. A number of the advantages of SAFER+ algorithm are its speed, simplicity, transparency, flexibility of use etc.

As indicated by [9] the E0 calculation has downsides like constrained assets limit of linear feedback shift registers, low believability of PIN and address ridiculing. On the off chance that RSA calculation for key encryption and the DES key for information transmission are utilized, there is no compelling reason to exchange DES key covertly before correspondence. The creators have built up a plan of Bluetooth security by extending DES and RSA crossover encryption calculation and demonstrated that it is moderately more secure, accordingly guaranteeing information transmission between the Bluetooth gadget safety and real-time.

In [10] the authors have thought about a half breed encryption system utilizing triple DES for cipherment of the key and Tiger algorithm. Keeping in mind the end goal to outline vitality proficient cell phones, the battery utilization of the cell phones ought to be low. Rivest Cipher 4 (RC4) cipherment technique has better encryption speed contrasted with other encryption strategies, it applies a versatile length of key from 1 to 256 bytes to set up a 256-byte state table.

In [11] the author has focused on a technique to boost the client privacy and operational potency. For first time communication a random number is initialized and is incremented once after every successful communication and it

is used like a link key then encrypted using Blowfish algorithm.

In [12] the author has mentioned concerning Man-In-The-Middle (MITM) attack collectively of the wide used techniques for stealing the sensitive information of an user and to deal with it the writer has employed the secure the exchange of public keys in SSP. For the secure exchange the public keys are enciphered employing an acknowledged cryptographic function and reverse method is performed at the receiver to urge the original data.

In [13] the writers have given a theme named Enhanced secure simple pairing (ESSP) for the authentication of public keys of the two consumers concerned within the communication. A main device is selected to store the complete information of records of all devices by entering a password for public key to mitigate the attack.

In [14] Diffie-Hellman Key Exchange protocol is employed to securely compute the shared secret key, hash function MD5 is used to resolve the attacks associated with PIN, and Hummingbird-2 encryption formula is used to secure the pairing process with efficient use of processing power and the storage requirements.

A verification convention known as Mobile-ID is displayed to beat the worry of MITM which is a kind of dynamic assault exhibited in [15]. Digital signature is one of the best answer for the validation assaults. A similar idea is utilized here yet now a protected component like SIM card is utilized for confirmation of credible client by signing process however without association of a client.

The author in [16] has adjusted the confirmation procedure amid SSP on the grounds that to start the way toward pairing, authentication of both the clients is a vital undertaking to evade the MITM assault. A symmetric encryption process is utilized to scramble the dedication esteem (key). At the point when the sender starts the correspondence the receiver will compare the decrypted commitment value (key) to authenticate the user.

Table 1 gives a brief overview of the distinctive security glitches, approaches used to conquer those issues, a brief investigation of these strategies and as per us their confinements which can be further reached out for future work.

TABLE I. ANALYSIS OF TECHNIQUES.

Security glitches	Methodology	Performance	Proposed solution	Limitations
Authentication	An Integrated Encryption Scheme algorithm based on IDEA, RSA and MD5. [2]	Cipher text of Key (CK) is generated by calculating hash value (MD5), which is encrypted by RSA.	Receiver will calculate and compare hash value (MD5) of both plain text and CK for integrity.	Low energy efficiency, operational speed can be improved, heavy processing power.
Authentication	Identity-Based Cryptography (IBC) to protect communication. [3]	It is an evolution of Public-Key Cryptography with shorter encryption keys and more efficient algorithms.	Public piece of information is used as public key to encrypt the data.	Security of key is compromised.
Authentication	A modified approach of RFCOMM implementation. [4]	A database of device address, location, company name, details, date of manufacturing etc. all necessary information is maintained by both the users.	To commence the communication scanning is performed and if the match is found then only link will be established.	Security for first communication, memory requirement is more.
Authentication	A verification tool (SPIN) is used. [5]	It is finite state model checker which analyzes the extent to which a model can be weakened still it remains efficient.	A restricted class of security protocols with precise formation and operation semantics is used.	Only key distribution protocols are considered.
Authentication, pairing	RSA algorithm for secure pairing. [6]	Bluetooth addresses of both units which are involved in communication are used for the pairing process.	For calculation of cipher text using RSA, hexadecimal prime values of Bluetooth devices address are used as input to RSA.	More time for connection establishment, high complexity, and low operational speed.

Authentication, Pairing, Encryption	SAFER+ algorithm.[7]	SAFER+ makes use of 8 identical rounds where every round calculates a 128 bit word using two sub keys and a 128 bit input from previous round.	A rotation block is used after every round and round 1 input and the round 2 output are Xor/Add Modulo 16 byte-by-byte. This result is used as the input of round 3.	Dissimilar coding and decoding process, no proof of overall security.
Encryption (Confidentiality)	RC4 algorithm for data encryption. [8]	A state table is used to generate pseudorandom stream which in turn used to produce the cipher text.	Fast encryption speed reduces battery consumption for low energy devices like mobile phones.	As data size increases the data transfer rate is reduced, Key collision.
Encryption (Confidentiality)	A hybrid encryption algorithm based on DES and RSA is used.[9]	A combination of RSA key encryption and the DES for plain text encryption is used.	TDES encrypts data as much as three times than DES and uses a different key for any of the three passes.	More computational complexity, less operational speed.
Encryption (Confidentiality)	A hybrid encryption technique using triple DES and Tiger algorithm. [10]	The message to be transmitted is generated by the combination of the session key from Tiger algorithm and the cipher text from triple DES encryption algorithm.	A Triple Data Encryption Algorithm (TDEA) is a compound operation of DES encryption and decryption operations.	Less energy efficiency, higher complexity, limited operational speed.
Pairing, Man in the middle attack	Method of secure simple pairing. [11]	Authentication is confirmed by visual number and encryption using Blowfish algorithm.	For first pairing a random number is initialized and is incremented for every successful transaction.	Blowfish algorithm is just used for transfer of public key.
Pairing, Man in the middle attack	Method of secure simple pairing.[12]	Encryption of public key with a cryptographic function.	Public key is stored in every device database makes SSP efficient.	Reduced computational speed, higher memory requirement.
Pairing, Man in the middle attack	Method of enhanced secure simple pairing. [13]	A unique password is assigned by exchanging the public keys and device addresses for all the devices.	A main device stores the records of all devices and all the devices copy record from main device then communication begins.	Extra memory requirement.
Pairing, Man in the middle attack	A hybrid pairing protocol based on Diffie-Hellman Key Exchange protocol, MD5 and Hummingbird-2 is proposed. [14]	Public keys are shared to calculate private keys and a hash function is used for secure key generation.	A lightweight algorithm is used to speed up the cost effective pairing process efficiently.	Secure key distribution.
Authentication, Man in the middle attack	A mobile authentication protocol named Mobile-ID is used. [15]	Messages are signed by genuine user in order to identify the attacker.	Users register with their preferred server and then use the same for authenticating themselves to any web site which accepts the Mobile-ID protocol.	Confidentiality aspects are not studied, applicable only for mobiles.
Authentication, Pairing, Man in the middle attack	Modified secure simple pairing. [16]	Encryption of the key computed by slave by using Elliptic-curve Diffie-Hellman Cryptography.	By using symmetric key encryption procedure user will perform the verification.	Reduced computational speed.



DISCUSSION ON GAP ANALYSIS

All encryption schemes are compared based on the parameters like complexity, Speed, security, memory requirement and power consumption in the proposed research. It is observed that the algorithms AES, DES, 3DES, RC4 are having the benefit of faster speed and less complexity but on the cost of compromised security for certain attacks like Man in the middle attack. Security is the key requirement of encryption. From security point of view the algorithms like RSA, DH are better however on account of more complexity and slower speed.

The proposed method combines Galois field and encryption algorithm. The method of Galois field(Elliptic curve cryptography) is having smaller key size, faster speed , High security, less memory storage requirement and low cost as major strengths . The proposed system deals with lightweight encryption system for near field communication devices hence to deal with various attacks related to near field devices (attacks related to authentication, pairing, MITM etc.) ECC offers a best suitable solution.

CONCLUSIONS

The Bluetooth Special Interest Group has anticipated that by 2018, more than 90% of Bluetooth-empowered cell phones are relied upon to bolster Bluetooth LE, while the number of Bluetooth-enabled cars is expected to prime 50 million. This paper presents diverse sorts of assaults that can be presented by assailant during the communication between two Bluetooth empowered gadgets.

The encryption algorithms like IDEA, RSA, MD5, and DES have an extent of change in speed of encryption, processing power. RC4 coding technique requires additionally handling time as information size continues expanding. The encryption functions like Triple DES and Tiger needs an improvement in energy effectiveness and computational speed. The distinctive pairing schemes require more memory for processing. The proposed arrangement combines Galois field having the potential benefits like higher computational speed, low cost with cryptographic algorithms can be turned out to be an all the more capable and secure approach to set up an association between near field communication devices. One of the practical application of this proposed security framework may be in recent ATM frauds.

ACKNOWLEDGMENT

This work was supported by contract DFNI I02/1 for research project:”Intelligent man-machine interface for assistive medical systems in improving the independent living of motor disabled users” of the Bulgarian Research Fund of the Ministry of Education and Science.

REFERENCES

[1] N. Minar1 and M. Tarique, “Bluetooth security threats and solutions: a survey”, International Journal of Distributed and Parallel Systems (IJDP) Vol.3, No.1, January 2012.  
 [2] B. Singh Rajput, P. Gupta, S. Yadav, “An Integrated Encryption Scheme Used in Bluetooth Communication Mechanism”, International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1 , Issue 2.

[3] V. Mora-Afonso, P. Caballero-Gil, J. Molina-Gil, “Strong authentication on smart wireless devices”,2013 IEEE.  
 [4] H. Hossain, U. Kabir and S. Rahman, “Modified Approach of RFCOMM Implementation to Protect Bluetooth Technology from Bluebug Attack”,2011 IJCIIT.  
 [5] N. B. Henda, “Generic and Efficient Attacker Models in SPIN”, SPIN '14, July 21-23,2014, San Jose, CA, USA.  
 [6] P. Vishwakarma, Professor B. Patel, “Bluetooth Security–Secure Data Transfer over Bluetooth” , IJCSMS International Journal of Computer Science and Management Studies, Vol. 11, Issue 02, Aug 2011.  
 [7] B. Manthan A, Dr. A. S. Shingh, “Multilevel Security Algorithm for Bluetooth Technology”, International Journal for Research in Technological Studies ISSN: Vol-1, Issue - 1, Dec 2013.  
 [8] S. Latchmanan, Dr.S. Parveen, “Applicability of RC4 algorithm in bluetooth data encryption method for achieving better energy efficiency of mobile devices”.  
 [9] K. Raganath Reddy, G.Srinivas Raju, “A New Design of Algorithm for Enhancing Security in Bluetooth Communication with Triple DES”, International Journal of Science and Research (IJSR), India Volume 2 Issue 2, February 2013.  
 [10] P S Patheja, A. A. Waoo, S. Nagwanshi, “ A Hybrid Encryption Technique to Secure Bluetooth Communication”, International Conference on Computer Communication and Networks CSI-COMNET-2011.  
 [11] N.Dave, K. I. Lakhtaria, “A New Approach to Handel MITM Attack on Secure Simple Pairing”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 7, July 2015.  
 [12] P. K. Mishra, “Analysis Of MITM Attack In Secure Simple Pairing”, Journal of Global Research in Computer Science, Volume 4, No. 2, February 2013.  
 [13] I. ALMomani, Md. Al-saruri and M. AL-Akhras, “Secure public key exchange against man-in-the-middle attacks During Secure Simple Pairing (SSP) in Bluetooth”, World applied science journal, 2011.  
 [14] J. T. Lalisl, B. D. Gerardo2 and Y. Byun, “ Securing Bluetooth Communication with Hybrid Pairing Protocol”, International Journal of Security and Its Applications Vol.8, No.4 (2014).  
 [15] K. Bicakci, D. Unal, N. Ascioğlu, O. Adalier, “Mobile Authentication Secure Against Man-In-The-Middle Attacks”, The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014), 2014.  
 [16] Md. A. Alam and Md. I. Khan, “Security Enhancement of Pairing and Authentication Process of Bluetooth”, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.  
 [17] T. Sharma , Attacks on bluetooth devices, Wegilant Net Solutions Pvt. Ltd.  
 [18] K. Haataja et al., Bluetooth Security Attacks, Springer Briefs in Computer Science,2013.  
 [19] J. Li, Hua Li, Wei Dong “A Flexible Architecture for the Multiplier in Hybrid Cryptosystems”, IEEE conference,2011.  
 [20] A. M. Sagheer,S. S. Al-Rawi,O. A. Dawood, “Proposing Developed Advanced encryption standard”, IEEE conference,2011.  
 [21] M. Shrivastava, S. Murarka, S. Narware,“Performance Analysis of Quadratic Prime Code for Modified Bluetooth FH Communication System”, International Journal of Emerging Technology and Advanced Engineering, January 2012.  
 [22] M. J. Ying, J. P. Wang, C. B Sheng, “PASA: Passive Broadcast for Smartphone Ad-hoc Networks”, IEEE conference,2014.  
 [23] Yashpal , Kr.Y. Sinha , “Bluetooth Secure Simple Pairing Personal Area Network Countermeasure DOS Middle Attack” , International Journal of Engineering Trends and Technology (IJETT), Volume 12 Number 5 - Jun 2014.  
 [24] M. A. Chenaghlu , S. Jamali, N. N. Khasmaki, “ A novel keyed parallel hashing scheme based on a new chaotic system”,Elsevier,2016.  
 [25] P. Jindal, B. Singh, “RC4 Encryption-A Literature Survey ”, International Conference on Information and Communication Technologies, 2015.  
 [26] N.Arora, “ Light Weight Cryptographic Algorithms:A Survey”, IJRDTM – Kailash | ISBN No. 978-1-63041-994-3| Vol.20 | Issue 08,2013.

# State of the Art Protection Techniques against Denial of Service Attacks In E-commerce Network

Javed R. Shaikh  
SKNSITS, Lonavala, India  
jrs.sknsits@sinhgad.edu

Vandana M. Rohokale  
SKNSITS, Lonavala, India  
vrohokale.sknsits@sinhgad.edu

Georgi Iliev  
Technical University of Sofia  
gli@tu-sofia.bg

**Abstract**— In today's era of Digital world, the use of Internet is very common so is the use of it for online shopping (E-shopping). Nowadays, maximum Generally, are using online shopping due to its many advantages over traditional physical shopping such as low-cost, real-time, interactive, personalized, cross domain, etc. Due to the use of online shopping, human life becomes easier and they can get the real experience in the virtual world. The percentage of customers using the E-shopping is increasing rapidly because it is fast, efficient and which is of economic advantage too. The consumer's interest for adaptation of E-commerce affected by three factors which are consumer attitude towards online transaction systems, security and the last is trust in the reliability of online product suppliers. Personal information security and Debit or Credit card fraud are major concerns for customers and merchants in E-commerce. As the popularity of e-commerce increases so the threats grow. The DoS attack has a significant effect on e-commerce. In this paper, different security preventions are considered to protect e-commerce assets. This paper also highlights the available security Measures against DoS/DDoS attacks.

**Keywords**—E-commerce ; DoS; DDoS; Security of E-commerce network;

## I. INTRODUCTION

Nowadays Electronic payments are becoming an important part of our everyday life. For most people, it is difficult to imagine a single day where we do not make a use of credit/debit cards in a physical store or to perform some mode of online payments or money transfer over the internet. We almost trust on e-commerce system without consideration for how they work. E-commerce security is generally a part of an information security and it is applied to the factors of security such as Data security, Computer Security, server security and other factors of information security. Service providers as well as merchants who process credit card and debit card became the easy targets for computer hackers to steal information of cards and commit frauds. E-commerce security is the protection of E-commerce assets from unauthorized access, its alteration, or destruction. Different dimensions of E-commerce are security, Authenticity Integrity, on-repudiation, Confidentiality, Privacy, and Availability. A denial of service (DoS) attack is a malicious attempt to make a server or a network resource unavailable to users, usually by temporarily interrupting or suspending the services of a host connected to the Internet. A Denial-of-service attack is characterized by an explicit attempt by attackers to prevent authorized users of a

service from using that service. Generally, two types of DoS attacks are there; one in which it crash services and second that flood services. Viruses are a nuisance threat in the E-commerce world. They only disrupt E-commerce operations and should be classified as a Denial of Service (DoS) tool. The Trojan horse remote control programs and their commercial equivalents are the most serious threat to e-commerce.

DoS attacks are low-cost and difficult to counter without the right tools. Due to these reasons Dos attacks are highly popular even for people with technical knowledge. In fact, DoS services are offered on some websites and have grown more and more sophisticated. These services can effectively exploit application vulnerabilities and evade detection by firewalls.

## II. DENIAL OF SERVICE(DoS) AND DISTRIBUTED DENIAL OF SERVICES(DDoS) ATTACKS

### A. Denial of Services(DoS)

The most serious attacks are distributed type generally which involves forging of IP sender addresses (IP address spoofing) so that the location of the attacking machines cannot easily be identified, and filtering based on the source address is also difficult in this case.

Another approach of DoS attack is where an attacker can use spam email messages (also known as spamming) to launch an attack on your email account. Either you have employees email account or one available through a free service, these are available with a specific quota, which limits the amount of data you can have in your account at any given time. By sending many email messages to the account continuously, an attacker can consume your quota, preventing you from receiving regular messages.

### B. Distributed Denial of Services(DDoS)

A distributed denial-of-service (DDoS) is where incoming traffic comes from thousands - of unique IP's, either from botnets or via various types of reflection attack. DDoS attacks target a wide variety of important resources, from banks to news websites, and present a major challenge to making sure people can publish and access important information.

By taking advantage of security weaknesses or

vulnerabilities, an attacker could take control of user computer. An attacker could then force user computer to send huge amounts of data to a website or send spam to particular email addresses. The attack is "distributed" because the attacker is using multiple computers, including users, to launch the denial-of-service attack. that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

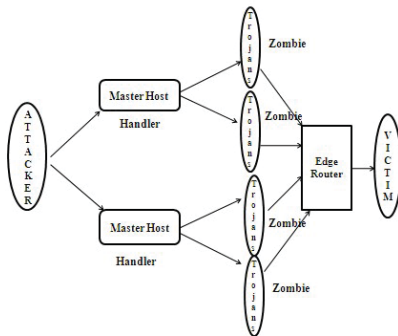


Fig.1 Architecture of DDoS attack

III. SYMPTOMS OF DOS ATTACK

DoS attack has different significance when it is applied to the networks. There are different indications from the network when it is affected by the DoS attack.

1. Unusually slow network performance (opening files or accessing web sites).
2. Unavailability of a particular website.
3. Inability to access any website.
4. Dramatic increase in the number of spam emails received—(this type of DoS attack is considered an e-mail bomb).
5. Disconnection of a wireless or wired internet connection.
6. Long term denial of access to the web or any internet services.

IV. TYPES DOS ATTACK

A. Volume Based Attack

Volume based attacks include UDP (User Datagram Protocol) floods, ICMP (ping) floods, and other spoofed packet floods. The main target of attack is to saturate the bandwidth of the attacked site. Its magnitude is measured in bits per second (Bps).

B. Protocol Attack

Protocol attack includes fragmented packet attacks, SYN floods, Ping of Death, Smurf DDoS and more attacks. This type of attack consumes server resources, or some intermediate communication equipment, such as firewalls and load balancers. Its Magnitude is measured in Packets per second.

C. Application Layer Attack

Application layer attacks include Slowloris attack, Zero-day DDoS attacks and DDoS attacks that target Apache, Windows or Open BSD vulnerabilities and more. These group of attacks comprised of seemingly legitimate and innocent requests and the goal of these group is to crash the web server. Its magnitude is measured in Requests per second.

V. SPECIFIC DOS ATTACK TYPE

1. UDP(User Datagram Protocol): Using UDP flood attack the attacker attacks random ports on the targeted host with IP packets containing UDP datagrams. This type of attack floods random ports on a remote host with a large number of UDP packets, causing the host to repeatedly check for the application listening on that port, and reply with an ICMP Destination Unreachable packet. Sometimes this DoS attack is in the form of DNS amplification attack, also called as alphabet soup attack.
2. ICMP (Ping) Flood: It devastates the target resource with ICMP Echo Request packets (ping packets), generally sending packets as fast as possible without waiting for replies. As victim server will busy in responding ICMP echo packets which result in system slowdown. This type of attack can consume both outgoing and incoming bandwidth.
3. SYN Flood: This attack consumes resources on the targeted server and makes it unresponsive. SYN flood DDoS attack exploits a known weakness in the TCP connection sequence, the host must respond with SYN-ACK to SYN request sent by the requester to initiate a TCP connection with a host and then confirmed by an ACK response from the requester. In an SYN flood, the requester sends multiple SYN requests, but it does not respond to the host's SYN-ACK response and it does not send the SYN requests from a spoofed IP address. During this process the host system continuously waiting for an acknowledgment for each of the requests which ultimately results in denial of service.
4. Ping of Death: This is a type of DoS attack where attacker sending multiple malicious pings to crash, destabilizes, or freeze the targeted computer or service. The maximum length of an IP packet is 65,535 bytes. However, the Data Link Layer usually has limits to the maximum frame size. In this case, a large IP packet is split across multiple IP

packets (fragments) and the recipient host considers IP fragments as a complete packet. This malicious manipulation of fragment content, the recipient ends up with an IP packet which is larger than 65,535 bytes when reassembled. This can overflow memory buffers allocated for the packet, causing a denial of service for legitimate packets.

5. **Slowloris:** Slowloris is a highly-targeted attack. It takes place generally by enabling one web server to take down another server, without affecting other services on the target network. Slowloris works by opening multiple connections to the targeted web server by continuously sending partial HTTP requests, none of which are ever completed and keeps web server open as long as possible. The targeted server keeps each of these false connections open, which yields overflows the maximum concurrent connection pool, which yields a denial of connection from clients.
6. **Amplification:** In this attack, the perpetrator exploits publically-accessible Network Time Protocol (NTP) servers to overwhelm the targeted server with User Datagram Protocol (UDP) traffic. NTP amplification is essentially a type of reflection attack which involve eliciting a response from a server to a spoofed IP address. In this attack, attacker sends a packet with a victim's IP address and the server replies to this address. Any attacker that obtains a list of open NTP servers can easily generate a devastating DDoS attack.
7. **HTTP Flood:** HTTP flood DDoS attack involves exploits of HTTP GET or POST requests to attack a web server or application. HTTP flood attacks are volumetric attacks in nature. It often uses a botnet "zombie army" which is a group of Internet-connected computers, each of which has been maliciously taken over, with the assistance of malware like Trojan Horses. The attack is most effective when it forces the server or application to allocate the maximum resources possible in response to each single request. HTTP floods do not use malformed packets, spoofing or reflection techniques. Generally, this attack requires less bandwidth than other attacks to bring down the targeted server.

VI. AVAILABLE PROTECTION TECHNIQUE FOR DDoS ATTACK

There are different protection tools are available to protect the servers or systems from DDoS attack. The selection is dependent on the type of DDoS attack. By implementing one of the below protection tool we can minimize the effect of DDoS attack and we can prevent server of the system from DDoS.

Table 1- Protection tools against DoS/DDoS

Protection tool	Description
<i>Firewalls</i>	In the case of a simple attack a firewall is used. Firewall has simple rule added to deny all incoming traffic from the attackers, based on protocols, ports or the originating IP addresses. However more complex attacks will be hard to block with simple rules.
<i>Switches</i>	Most of the switches have some rate-limiting and ACL capability. Some switches provide automatic rate limiting and/or system-wide limiting traffic, delayed binding (TCP splicing), deep packet inspection and bogus IP filtering (Bogon filtering) to detect and remediate denial-of-service attacks. These schemes will work as long as the DoS attacks can be prevented by using them.
<i>Routers</i>	Routers have some rate-limiting and ACL capability. They too are manually set. Most routers can be easily weighed down under a DoS attack. Cisco IOS has optional features which can reduce the impact of flooding.
<i>Application front end hardware</i>	Application front end hardware is intelligent hardware placed on the network before traffic reaches the servers. It can be used on networks in conjunction with routers and switches. This hardware analyzes data packets as they enter the system, and then identifies them as priority, regular, or dangerous. There are more than 25 bandwidth management vendors available.
<i>Application Level Key Completion Indicators</i>	In case of application level DDoS attacks against Cloud based applications, approaches may be based on an application layer analysis, to indicate whether an incoming traffic bulk is genuine or not and thus enable the triggering of elasticity decisions without the economic implications of a DDoS attack.
<i>IPS based prevention</i>	Intrusion-prevention systems (IPS) are effective only if the attacks have signatures associated with them. However, the trend among the attacks is to have rightful content but bad intent. IPS which works on content recognition cannot block behavior-based DoS attacks.
<i>DDS based defense</i>	DoS Defense System (DDS) can block connection-based DoS attacks and those with genuine content but bad intent. A DDS can also address both protocol attacks such as Teardrop and Ping of death and rate-based attacks such as ICMP floods and SYN floods.
<i>Blackholing and sinkholing</i>	With blackholing and sinkholing technique, all the traffic to the attacked DNS or IP address is sent to a "black hole" which is a non-existent server. To be more efficient and avoid affecting network connectivity, it can be managed by the ISP.
<i>Clean pipes</i>	By using clean pipes all traffic is passed through a cleaning center via various methods such as proxies, tunnels or even direct circuits, which blocks bad traffic (DDoS) and also other common internet attacks) and only sends good traffic beyond to the server.

Above table 1 describes different protections tool available in the market against DoS/DDoS attacks. While dealing with E-commerce we need to consider security at different levels, such as front end security (Authorization) Transaction security and Server side security. To protect the e-commerce server from DoS attack we need to use any one protection techniques mentioned in above table. The work is

going on to design the best prevention-detection mechanism against DoS attack in E-commerce.

#### VII. CONCLUSION

E-commerce security is the protection of E-commerce assets from unauthorized access, use, alteration, or destruction. Dimensions of E-commerce security are Integrity, No repudiation, Authenticity, Confidentiality, and Availability. This paper highlights the existing DoS/DDoS attacks in E-commerce security and related techniques applied in cyber security field along with the major challenges. Also in the paper different Protection Techniques available against DoS attacks are considered for Securing E-commerce Network. Though these techniques being applied in the field of cyber security seem to be capable of enhancing cyber security measures, but still require a lot of research to deal with the influx of new threats in large computer networks.

#### References

- [1] McDowell, Mindi, "Cyber Security Tip ST04-015 - Understanding Denial-of-Service Attacks". Archived from the original on 2013-11-04. Retrieved December 11, 2013.
- [2] Taghavi Zargar, Saman, "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks" IEEE COMMUNICATIONS SURVEYS & TUTORIALS. pp. 2046–2069. Retrieved 2014-03-07.
- [3] Ahmad Mukaram, "Cyber Threat Landscape: Basic Overview and Attack Methods" Cyber Threat Intelligence, June 3 2014
- [4] Park, H., Jung, O., Lee, H., In, H., "Cyber weather forecasting: forecasting unknown internet worms using randomness analysis", in Gritzalis, D., Furnell, S., Theoharidou, M. (eds.) Information Security and Privacy Research, AICT, vol. 376, pp. 376–387. Springer, Heidelberg, 2012.
- [5] Yang Liuqing and HU Yu,Huang Weimin "Study on Power e-commerce Network Security Technique and Design Based on Firewall", The 6th International Conference on Computer Science and Education (ICCSE 2011) August 3-5, 2011. SuperStar Virgo, Singapore.
- [6] Reyhaneh Tamimi, Prof. Dr. Mohammad Ebrahim Mohammad pourzarandir, "The Application of Web Usage Mining In E-commerce Security", The 7th International Conference on e-commerce in Developing countries with focus on E-security 2013
- [7] Behzad Yahid, Dr. Assadollah Shabbahrami, Dr. Mohammad Bagher Nobakht "Providing Security for E-wallet using E-cheque ", The 7th International Conference on e-commerce in Developing countries with focus on E-security 2013

# *User Privacy in Big Data Analytics for eHealth*

## *Data Privacy Model*

*Nidhi*  
 Vishwaniketan iMEET  
 Navi Mumbai, India  
[nidhi@es.aau.dk](mailto:nidhi@es.aau.dk)

*Albena Mihovska*  
 Department of Electronic Systems  
 Aalborg University, Denmark  
[albena@es.aau.dk](mailto:albena@es.aau.dk)

*Ramjee Prasad*  
 Department of Business and  
 Engineering,  
 Aarhus University, Denmark.  
[prasad@es.aau.dk](mailto:prasad@es.aau.dk)

**Abstract**—Big data analytics can benefit the healthcare sector by incorporating improved situational analysis, database management, real-time decision making and new ways of diagnosis and treatment. However, its use opens critical security and privacy concerns. This paper surveys the open challenges of collecting and accessing health data, and the different types of possible breaches of privacy and security that are the key to the successful deployment of eHealth systems. To mitigate the privacy hindrance issue with the medical data, we propose an eHealth data privacy model, which will add transparency in to the personal data collection, aggregation, handling and storage. Transparency in healthcare sector have different interpretation at different level. We'll look at some different segments in the healthcare industry working to adapt to the call for transparency. The model builds upon the Information Accountability protocol for the transparency. The user will be the player and take decision on their data, how is to be used and shared.

**Keywords**—Big Data; e-Health; e-Health Data; Data Security and Privacy

### I. INTRODUCTION

A typical eHealth system should be highly secured, responsive, and controlled and one, where the users' privacy and the protection of their personal data, remains intact. eHealth systems demand integrity, accessibility and availability along with interoperability, which is even more important with the colossal pool of data defining the infrastructure of today. Many everyday applications and services rely on the collection, storage, processing and analysis of data, often user-related, and often made available to different sectors irrespective of boundaries ranging from machine learning and engineering, to economics and medicine [1].

The amount of data generated in healthcare sector is increasing and will continue to increase with the technological enactments, creating room for new data handling and analysing techniques. Big data analytics provide tools to benefit healthcare for example; it provides customized medications, anticipated analytics, risk- intervention etc. [2]. It has marked a presence in handling and analysing data generated via the social media but it offers promising solutions for handling efficiently eHealth data (also, commonly referred to as "health big data"). Big data analytics includes data aggregation, processing, storage of eHealth data to make decisions and evolve new ways of treatment, keep the population healthy etc. [1].

The health-related data are usually stored and processed at distributed repositories at different levels [3]. There are numerous security and privacy concerns in moving the health data under the big data approach. The privacy of the patients and the safeguarding of their personal data are major issues in applying big data analytics to eHealth.

A recent survey published in [4], suggested that the lack of adequate security measures had resulted in numerous data breaches in the healthcare sector, exposing certain patients to economic threats, mental stress, and even social embarrassments. Sharing the patients' personal information without the user's official consent is one critical privacy breach for the healthcare sector. The authors in [5] have summarized the issue of connected healthcare and requirement of appropriate protections to safeguard the privacy of the patients and for minimizing the medical error. Therefore, an appropriate equity is needed to maintain privacy and security of data and the patient's personal space in healthcare.

In this paper, we survey and analyse the privacy and security issues in healthcare when using big data analytics. Based on the investigation, we propose an eHealth Data Privacy Model for enabling transparency in the flow of data over the network and that only the data relevant to a particular health service provider would be delivered. The model is based on the concept of Information Accountability [6] which enable patients to decide the usage of their data on a shared platform. It advocates transparency in the data usage and enable one with the ability to track the appropriate use of data under the predefined rules.

The paper is organized as follows. Section II describes the current state of the art in the area. We elaborate the concept of eHealth data, and survey the associated threats and vulnerabilities, the potential attack zones and how data are transmitted and received in the network. Section III analyses the key factors and issues related to the flow of information in an eHealth scenario. We explore the different aspects of information security related to healthcare and the user. We highlight the issues in eHealth for data privacy and formulate the need for a data privacy model. In Section IV, we propose the health data privacy model and the related functionalities, modules, protocols and required networking. Section V highlights the issues and challenges involved in implementing the proposed model. Section VI concludes the paper.

II. STATE OF THE ART

The healthcare sector spans over a vast landscape demanding cooperation and the active participation of public and private bodies, the individual user along with innovations and initiatives from various fields including marketing, finance, education and many more. The eHealth's objective is to avail medical services and amenities accessible and available at a reasonable cost and available resources while maintaining the quality of care and productivity. In an eHealth scenario, both, the patient and the medical service provider will be connected for the health monitoring, routine check-ups and even emergency services, facilitated by real-time secure data exchange. The healthcare industry dominates the data volume per person per day ratio generated. To handle data, Big Data can make significant revolution without resulting into additional infrastructure. It is an emerging technology for the future generations, which can analyse wide variety of voluminous data. It enables the processing of high-volume, high-velocity, and/or high variety of data aiding optimized results, better and efficient analysis, improved decision-making etc.

A. eHealth Data and Data Flow Structure

In an eHealth scenario, human and associated data are the most valuable and vulnerable assets. The medical reports generated electronically, called Electronic Health Records (EHR) are documents containing the patient's personal details (i.e., that have been used for registration over the network, the personal social security numbers used for the medical insurance, the medical reports, the diagnosis reports, the discharge summaries etc.) The medical data represents the patient-doctor relationship (e.g., the information including the patient-identification, the medical history, the digital renderings of the medical images, the treatment received, dietary habits, sexual preference, genetic information, psychological profiles, employment history, income, and physicians' subjective assessments of personality etc. [3], [7].)

Figure 1 shows the flow of medical data within the healthcare system. The patient shares their medical history, symptoms and personal identification details to the primary health services' unit. The primary health service unit then registers the patient using a unique identifier and creates the patients' file, which is shared with restrictions with the billing unit (relevant treatment charges and genuine user detail) and organization's IT unit (billing details and treatments' summary). The primary health services are responsible for the various test reports, clinical data, laboratory activities etc. and accordingly communicate with the secondary health services, pharmacy and regional health centres using patients' identifier and hiding other background details. The primary health services also share the details with its employees, which are strictly service-based i.e. ground staff are only exposed to the details like medication timings, test routines while doctors get detailed medical history irrespective of personal details. Third party IT services are used to store the health record files. They have an access to patients' personal data (identifiers) and medical data (contributed by health organizations) and make the same available on demand.

The information transfer and storage in organization's and third party's IT servers are critical [4] as even a single careless activity can expose a patient's details. The data of the patients can be used to improve the efficiency within the healthcare system, to drive the public policy development and the administration at a state and federal level, and in the conduct of the research to advance the medical science apart from the personal care [8], [9].

B. Data Breaches

A medical data breach can lead to everything from an identity theft to billing fraud to blackmail, some breaches ultimately have little consequence on the patients affected. Whenever a medical data breach occurs, it signifies that there is a lack in security, while handling the information of the patient. In recent years, the survey [4], [10] recorded that 43 per cent of the total data breaches involved healthcare data. The healthcare breaches

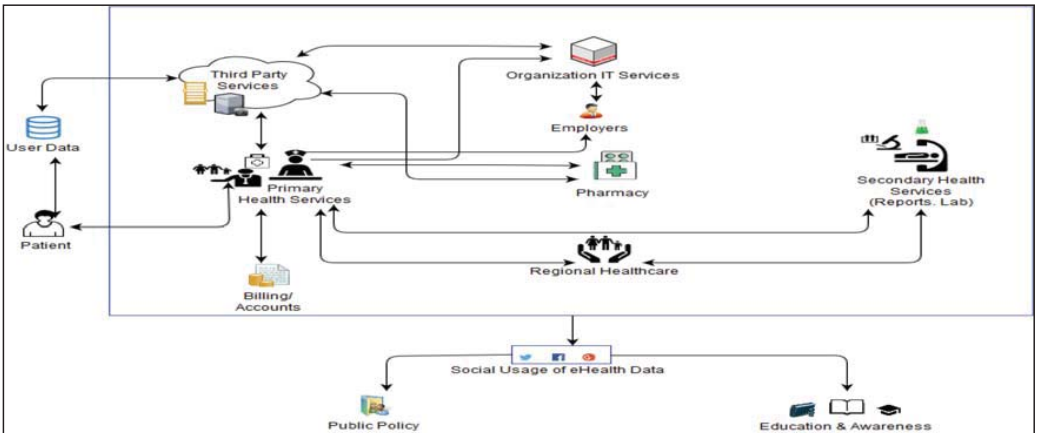


Figure 1 Data Flow in eHealth Scenario

originated mostly from the service provider organization and/or a third party associate. It has been reported in [10], that about 90 percent of healthcare organizations had suffered data breaches, such as cyber attacks, employee mistakes, theft etc. Most common breaches are data and identity theft, unauthorized access, hacking the transmission, the loss of data in transmission, improper disposal, denial of service etc.

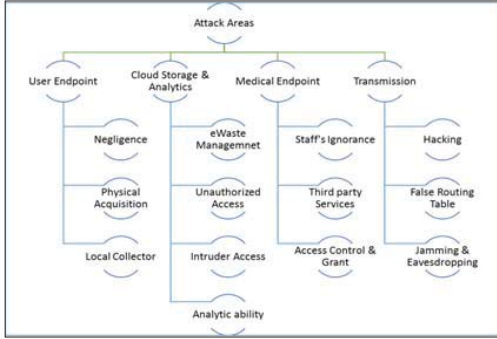


Figure 2 Attack Areas in Healthcare system  
 Figure 2 summarizes the vulnerable areas in each segment of the eHealth scenario cycle. The healthcare sector can be defined broadly into four sub-sectors, namely, the user-endpoint, the cloud storage and analytics, the medical endpoint and the transmission. At the user end, the breaches mainly would be the result of ignorance or access to wearables, documents etc., unethically. At the storage and the analytics end, the unauthorized access, a compromised node/ employee, the lack of security measures and the improper disposal of data may cause medical breaches. The staff's negligence, unauthorised access, third-party service dependency are some of the root causes to breaches. Hacking or jamming the transmission of information also contribute to compromised and lost data.

C. Threats to Patients' Privacy and Data Security

A threat scenario is defined by the motives, resources, accessibility and technical capability. The threats to the privacy of the patient are a major outcome of the illegitimate usage of data either by internal/ external agents or by a third party agent in the data flow chain. The authors in [7], [8] summarized various threats to the user integrity and life. The threats impose different level of risk depending on the motive, the attack zone, the sensitivity and the mitigation and prevention strategies. The threats were categorized as the threats arising from accessing the patients' data inappropriately either by internal or external sources, and as the threats arising from data exposure over the network.

III. INFORMATION IN HEALTHCARE

The increased use of Web-based services has significantly raised the bars for the privacy concerns of the users. Published research has summarized the user content among a wide range of users, which include students, employed, senior citizens etc. The disclosure of data is user-dependent, i.e. a user agreement

is needed in eHealth to verify the consolidation of the user to disclose personal data for research and development and/or other healthcare related needs. The current security and privacy in health data was summarized into the following sub-categories [11-17].

A. Data Access and Security in eHealth:

The healthcare institutions appeal to have security measures to govern the data access. Some of the common steps taken in that direction include access control systems, intrusion detection systems, policies etc. In [11], the authors have used a game theoretic approach to model the optimal levels of access. Remodelling the existing access control policies in the healthcare scenario is challenging apart from being highly expensive. In healthcare, we have individuals having different sets of roles, dependent data streams, independent data systems, dynamic configurations etc.

Data security requires concrete frameworks and defined protocols to identify, solve and mitigate the related security issues as stressed in [3], [5]. The emergence of ubiquitous access to patient data via mobile devices has exposed the vulnerabilities of the patients even further.

B. Authorized Data Disclosure and Integrity:

Attributing public health, the privacy policies should be made strongest when it comes to individual and communal interests. For each solution proposed, it should be carefully outweighed how much data gets disclosed and at what span [12]. Health services should be available on demand, which requires a full-time data protection.

Healthcare systems are getting more and more vulnerable to cyber security incidents nowadays. Factors like voluminous data generation; extensive usage of IT technologies to connect patients and healthcare utilities; data exposure over the network; diversified nature of healthcare systems; outdated applications and systems; poor security algorithms; expansion in devices with enhanced capabilities and many more have contributed to the exponential increase in the number of incidents in the healthcare [16], [17].

C. eHealth as a Critical Data Platform

The above mentioned factors along with the data breaches makes the healthcare sector critical. A healthcare platform deals with asset classification and requirement to form a base layer of the healthcare system and the components may include index services of the user and/or service provider, registration proofs, identifiers etc. The data privacy, security and integrity involves network elements and storage (internal/external clouds). Access is determined using identifiers. Availability is the crucial among all as it can cost even life of the patient in case of emergencies.



**D. Challenges in the Healthcare Data**

The authors of [14], [16] and [19] suggested that to maintain the users’ privacy and in order to establish a balance in the economic constraints, quality of service and care and availability are the main challenges in healthcare. They advocated on the efficient and effective solutions for privacy maintenance at affordable and operative costs.

**E. Transparency in Healthcare**

In healthcare sector, transparency is needed at every sector and individual end-point. Transparency has its individual definition from patient, doctor, healthcare organization, payers and providers [20]. From patients’ perspective, transparency is needed in data acquisition and its usage and the cost for quality and services. The data collected from patients are in silos hence it becomes more important to check how data moves into the network both online and offline. With each bit of data comes an individual role and responsibility of data-managers in healthcare sector. Data are subjected to limited access grants to protect patients’ confidentiality.

**F. Data Analytics**

“Big Data” in health is defined as a voluminous complex and distributed data set, which imposes difficulty for conventional technologies in analysing and maintaining the information [19].

In order to safeguard the user privacy and tackle issues with interoperability and data repositories, we need a data model, which will be transparent, secure and able to analyse to produce the desired results. The data model should be able to manage widely distributed and scattered data. The data scheme should address user privacy and data sharing within agreement.

**IV. PROPOSED MODEL FOR DATA TRANSPARENCY**

Health data are sensitive and demand appropriate and authenticated usage. While implementing digital data records, the required security measures for data storage, access and monitoring should also be put in place [14], [15]. Our proposed data privacy scheme allows for transparency in data-handling. It reinforces mechanisms to mitigate the illegitimate (unauthorised access, modification, disclosure to unintended user etc.) use of data [13], [21] and to best exploit the benefits gained from sharing the health data.

In our proposed data privacy model, the patients set the usage authority for their information and decide the extent, to which data can be aggregated with others. The data is shared under well formulated set of rules, guidelines and policies.

**A. Components and Participants**

The proposed model is shown in Figure 3. The following agents are the entities in the eHealth scenario;

**Patient:** Patients are the data owners. They can view the records of accesses made at any time and can submit their queries on unauthorized accesses.

**Healthcare Utility:** These are the healthcare authorities, responsible to aggregate the patients’ personal data and medical history, test reports, medication briefs etc. These are often

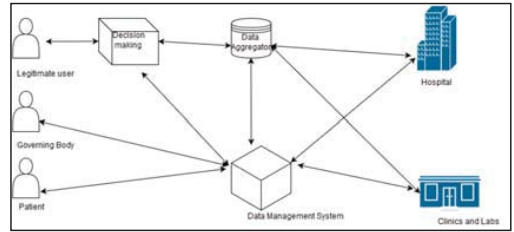


Figure 3 Proposed Data Privacy Model

referred to as Data Providers. They are responsible to provide data to the data aggregator, which maintains, stores and analyzes using big data analytics and policies imposed by a system manager. They also maintain data logs to manage risks.

**Governing Body:** these are the system managers who make and set policies to regulate the shared eHealth system and avoid misuse. These are responsible for the data integrity and log authenticity.

**Data Users:** Data users would make use of the aggregated data. Healthcare professionals, approved researchers, and government studies are examples of data users. Data users will be able to access specific log entries regarding their own access to patient information. They will be able to review the entries when they receive an inquiry requesting that they justify why they needed to access the relevant information in the given situation.

**Data Aggregator:** It collects information from the data providers. It works based on a policy set that considers the data provider and owner and allows aggregation of unobjectional data.

**Data Management System:** It acts as the main governing body that maintains interoperability among various participants as well as the services in the eHealth scenario.

The Data Management System regulates and governs each of the related entities, to ensure that there is no hindrance to the patient’s privacy. It sets usage policies that allow for in-flow and out-flow of data for research-based applications, for maintaining logs, and for making the logs available for the patients to review anytime.

The data management system is responsible for establishing default policies and maintaining logs and other information for the governing body. The data aggregator and data management system coordinates while maintaining the retrieval policies and recording the logs events. The data management system also sets policies for the smooth functioning among the clinics, hospitals, laboratories and other medical end utilities. The data aggregator sends and receives data requests and responses from the laboratories and other users responsible for generating viable health information. The legitimate user sends requests to the decision making body to execute the request. The decision making body on receiving the request from a legitimate user, generates a query data and sends it to the data aggregator for approval.

*B. Description and Working Principle*

The proposed data protection model, shown in Figure 3, allows for transparency on how the patient’s data are used within the healthcare system. The model guarantees no data usage without the users’ consent or agreement. The model uses the secure key management scheme as proposed by the project MAGNET and MAGNET Beyond [23]. It explains a new key agreement protocol based on elliptical curve cryptography and personal public key infrastructure. The patients have control over the data access by intended or third-party users. The Governing Body, responsible for regulatory policies have power to grant/deny any request to access the user-data without hampering their privacy.

The model creates log for all the successful and unsuccessful accesses, which can serve as a database to validate future requests. The patients can refer to the log to check their access details periodically. The data management system monitors and controls all accesses together with the system manager, a patient and data aggregator. In the case of an unauthorized request, patient can go for inquiry and ask justification. The data management system is responsible to answer the user’s inquiry.

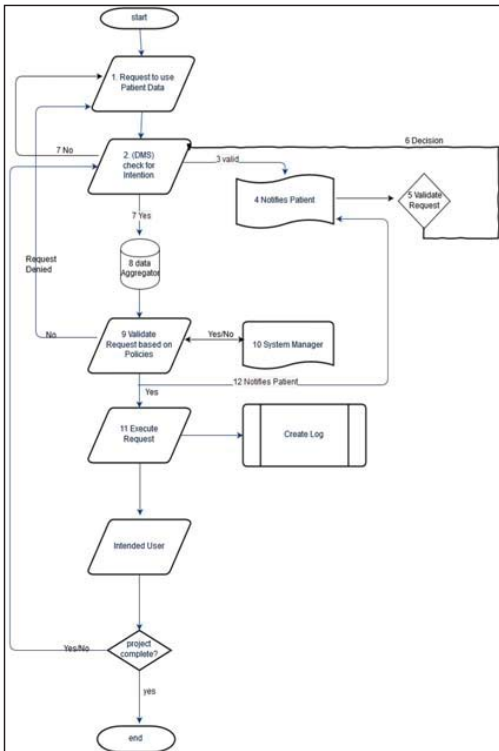


Figure 4 Working Principle of Proposed Data Privacy Model

The basic working principle is explained through the flowchart in Figure 4. The model works basically based on the policies that take into consideration the interests and concerns for the patient as well as the healthcare utilities. The following section describes briefly the Model’s working;

*Governing Policies*

**Healthcare Utility:** Data providers share their data under predefined set of usage and collection policies. These policies are user-friendly and do not interfere in the services provided. Policies would only govern the data usage and access for the healthcare development and facilitate research.

**Patient:** The user controls the personal data shared on the network for maintaining health and disease control by setting up policies, which let them decide when, how and by whom the data is to be accessed. The policies also govern the amount of visibility depending upon the usage, purpose and motive. The users can invoke filters at their choice.

**Governing Body:** the system managers are the government bodies, which set up certain rules and regulations to maintain a social balance and to restrict privacy hindrance to the individual.

*Data Collection and Aggregation*

In the model, a data aggregator collects information from the data providers and the data owners. Only the permissible data would be stored and analysed further for other purposes intending to develop an overall eHealth scenario.

*Request to Access Data*

When a data user executes a query in the system, the query service retrieves a policy for the data user. This can include rules regarding, which data they can access, how they can use data, and required de-identification of the results. If they are permitted to perform the query, the retrieved rules are then applied to filter the result set, removing restricted information. The information access request is logged, and the policy versions used to determine the access request is stored with the context-aware log entry.

*Maintaining Access Records*

The logs produced in an accountable system can contain sensitive information themselves and must be appropriately protected, including restricting who can view these logs and for what purpose.

V. IMPLEMENTATION

The implementation of the proposed model involves the formulation of rules and policies, data collection and aggregation, request to access data and creation and maintenance of access record-logs.

The governing policies are framed at three levels or working groups, comprising the healthcare utility, the patients and at the governing body as explained in the previous section. Each of the entities have their own requirements and accordingly set policies. For example, the doctors may share patients’ symptoms and behaviour for expert comments but the system would hide the personal data and identification; similarly, the patients would participate in guiding the usage of their data. The policies are generally depicted in the Open Digital Rights Language (ODRL) [22], which encourages the adoption of open

international specifications for expressing policies in language. For data collection, aggregation and access, we will implant filters that will prevent restricted data access and maintain context-aware log entry. The model also incorporates a context aware security management scheme, which allows to have virtual identities and include various agents to ensure trust, privacy, security and disclosed information. It authenticates uncompromised nodes and make decision on what to be shared keeping the patient aware of it.

*Challenges:*

The proposed eHealth data privacy model would have to comply with the required scalability, heterogeneity and performance metrics assuring the data storage in the knowledge of the user and the concerned authorities. The successful implementation of the proposed data privacy model in eHealth Big Data Analytics use cases, requires to be overcome the challenges imposed by big data alone and also the adverse effects in healthcare sector.

Analysing random as well as discrete data in eHealth will be complex and it will be difficult to maintain the exponential growth in the operation and computation time. The log maintenance will be difficult. Access grant against the query while maintaining its privacy will be a challenge. How an information is accessed, queried or stored including their log records will be a challenge. The major challenge for the implementation of the model is to accumulate and correlate the information coming from heterogeneous data sets.

VI. CONCLUSION

The paper proposed a data privacy model in healthcare using big data analytics, added transparency in the data handling and accessing. The proposed model triggers the privacy issues in data aggregation and allowed access by maintaining logs and seeking due consents from the users who are the data owners. It also promotes the data sharing and risk mitigation in healthcare.

Future work will incorporate solutions to the imposed challenges in system scalability, interoperability, heterogeneity of data sets and implementation challenges.

ACKNOWLEDGMENT

The research was supervised under Centre for Teleinfrastruktur (CTIF), Aalborg University, Denmark. The author would like to show their gratitude towards Ambuj Kumar and other colleagues for providing the timely efforts and cooperation.

REFERENCES

[1] Wang, W. and E. Krishnan, Big data and clinicians: a review on the state of the science. *Journal Medical Informatics*, 2014, 2(1): p. e1.  
 [2] Huser, Vojtech, and James J. Cimino. "Impending challenges for the use of Big Data." *International Journal of Radiation Oncology• Biology• Physics* (2015).  
 [3] Ball, Marion J., and Jennifer Lillis. "E-health: transforming the physician/patient relationship." *International journal of medical informatics* 61.1 (2001): 1-10  
 [4] "Data Breaches In Healthcare Totaled Over 112 Million Records In 2015." Ed. Dan Munro. N.p., n.d. Web. 28 Sept. 2016.

[5] Terry, Nicolas. "An eHealth diptych: the impact of privacy regulation on medical error and malpractice litigation." *American journal of law & medicine* 27 (2001).  
 [6] Weitzner, Daniel J., et al. "Information accountability." *Communications of the ACM* 51.6 (2008): 82-87.  
 [7] Hodge, James G. "Health information privacy and public health." *The Journal of Law, Medicine & Ethics* 31.4 (2003): 663-671.  
 [8] Campos, Maria João Magalhães Pereira. "Identity in eHealth-from the reality of physical identification to digital identification." (2012).  
 [9] J. J. Rodrigues, I. de la Torre, G. Fernández, and M. López-Coronado, "Analysis of the security and privacy requirements of cloud-based electronic health records systems," *Journal of medical Internet research*, vol. 15, no. 8, 2013.  
 [10] By Greg Slabodkin. "Survey: No Cure In Sight for Healthcare Data Breaches." *Information Management RSS*. N.p., 2016. Web. 14 Sept. 2016.  
 [11] Zhao, X., and Johnson, M.E. (2008) —Information Governance: Flexibility and Control through Escalation and Incentives, *Workshop on the Economics of Information Security*, Hanover, NH  
 [12] Wilkowska, Wiktoria, and Martina Ziefle. "Privacy and data security in E-health: Requirements from the user's perspective." *Health informatics journal* 18.3 (2012): 191-201.  
 [13] Dimitra Liveri, Anna Sarri, Christina Skouloudi and ENISA. "Security and Resilience in EHealth: Security Challenges and Risks." *Security and Resilience in EHealth: Security Challenges and Risks*. European Union Agency for Network and Information Security (ENISA), 2015. Web. 13 Sept. 2016.  
 [14] N. H. Shah and J. D. Tenenbaum, "The coming age of data-driven medicine: translational bioinformatics' next frontier." *Journal of the American Medical Informatics Association*, vol. 19, no. e1, pp. e2-e4, 2012.  
 [15] L. P. Garrison Jr, "Universal health coverage—big thinking versus big data." *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 16, no. 1 Suppl, p. S1, 2013.  
 [16] Appari, Ajit, and M. Eric Johnson. "Information security and privacy in healthcare: current state of research." *International journal of Internet and enterprise management* 6.4 (2010): 279-314.  
 [17] NRC National Research Council (1997) —For the Record: Protecting Electronic Health Information  
 [18] J. Feigenbaum, A. D. Jaggard, and R. N. Wright, "Towards a formal model of accountability," in Proceedings of the 2011 workshop on New security paradigms workshop. ACM, 2011, pp. 45-56.  
 [19] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability," *Communications of the ACM*, vol. 51, no. 6, pp. 82-87, 2008.  
 [20] healthcatalyst. "3 Best Practices for Payer-Provider Collaboration to Improve Patient Care." *Health Catalyst 3 Best Practices for Payer/Provider Collaboration to Improve Patient Care Comments*. N.p., 2016. Web. 10 Oct. 2016.  
 [21] R. H. Sloan and R. Warner, "Developing foundations for accountability systems: Informational norms and context-sensitive judgments," in Proceedings of the 2010 Workshop on Governance of Technology, Information and Policies, ser. GTIP '10, ACM, New York, NY, USA: ACM, 2010, pp. 21-26.  
 [22] "Community & Business Groups." *ODRL Community Group*. N.p., n.d. Web. 27 Sept. 2016.  
 [23] Prasad, Ramjee, ed. *My personal adaptive global NET (MAGNET)*. Berlin: Springer, 2010.

# An Energy-Efficient Link with Adaptive Transmit Power Control for Long Range Networks

P. Lynggaard

CMI, Aalborg University Copenhagen  
Copenhagen, Denmark  
perlyn@es.aau.dk

T. Błaszczuk

CWSA, DTU Diplom  
Technical University of Denmark  
Ballerup, Denmark  
tomb@dtu.dk

**Abstract** — A considerable amount of research is carried out to develop a reliable smart sensor system with high energy efficiency for battery operated wireless IoT devices in the agriculture sector. However, only a limited amount of research has covered automatic transmission power adjustment schemes and algorithms which are essential for deployment of wireless IoT nodes. This paper presents an adaptive link algorithm for farm applications with emphasis on power adjustment for long range communication networks.

**Keywords**—IoT; WSN; energy saving; transmission power adjustment; adaptive wireless links; solar power in WSN

## I. INTRODUCTION

The demand for smart sensor system based IoT devices in the agriculture sector is huge. Especially, IoT devices for plant and crop protection which monitors diseases, insect attacks, and extreme weather anomalies are in focus. To deal with these challenges the IoT devices have autonomous computational capability and are able to communicate with a cloud system for reporting and offloading data in an energy efficient manner. Thus, saving energy is one of the key challenges in these systems [1]. In this paper we focus on energy management in a distributed wireless sensor system for agriculture usage. Energy management is needed to deal with the challenges of limited power (battery rechargeable units) and long communication distances between the nodes, i.e. it has impact on the application usability, reliability, and the life time of overall network. We present an algorithm and simulate its behaviour in the light of energy savings, message propagation time, message size, and dynamic behaviour.

As an alternative to using maximum available transmission power for each sensor node it should adjust its transmission power to the needed level for achieving a reliable communication channel regardless of weather conditions, obstacles, dynamic distance variation, manmade noise and other disturbances. The presented energy efficient adaptive link model provides an adaptive transmission power algorithm which uses interaction between the sensor nodes in order to determine the transmission power. The algorithm can be deployed in static networks as well as in dynamic ad-hoc networks.

## II. RELATED WORKS

In our recent work [1] we have presented an adaptive link model based on message propagation between the nodes. It presents an algorithm that is able to adapt to weather conditions, available power, and time of day/night for a sustainable and energy limited system based on solar cells. This work uses the CC1120 hardware RF modules and it assumes that the sensor nodes do not have dynamic transmit output power adjustment. However, for fast deployment and limited configuration and maintenance efforts it is strongly desirable to rely on “automatic” transmission power adjustment to set the proper link margin and run the nodes at their optimal transmission power. Depending on the topology and the communication distance an automatic transmission power adjustment approach can save a lot of energy in battery operated devices.

A large proportion of the IoT devices on the market do not have automatic transmission power adjustment. However, new devices such as the SIGFOX GPS network tracker [2] offers a possibility to implement such algorithms, but its defaults the transmit power to a constant value (maximum power) [3]. A lot of research is ongoing to find efficient algorithms for adaptive links which regulate the transmit power in terms of energy savings. One example is the work performed by Chen et al. [4] which proposes method to minimize energy in communication network by reducing overall sensor transmission power. Their algorithm uses an approach based on constructing a Relative Neighbourhood Graph and optimizes the performance in relation to this. Nonetheless, their algorithm focuses on cluster of nodes whereas this work presents an algorithm which is capable of handling node to node communication as well.

## III. THE CONCEPT AND THE CHALLENGES

The context for this work is limited to devices distributed in farm fields and forests. Due to large distance and lack of power resources in this context the IoT nodes must operate on ambient energy sources such as photovoltaic cells, mechanical harvesting, etc).

There is plenty of use-case scenarios in the farm-domains where IoT sensors solve problems in cost effective manner,

e.g. scenarios which cover disease control for plant and crop protection as presented in paper [1]. One of these scenarios is presented in the next section.

A. A case study of bee family condition in modern beehives

The map presented on Fig. 1 shows beehives placed in 2 spots in distance of 600m and 1.1 km away from farmhouse shown in left bottom corner. Additionally close to one bee spot there is weather station.



Fig. 1. . A farm equipped with IoT devices for tracking bee family condition

For a proper bee family growth and development after winter period, it is important to provide the right air conditions inside the beehive. Modern beehives are constructed with light styrofoam material which has many beneficial properties such as low thermal conductivity, low weight, and low size. Nevertheless, this type of construction requires frequent inspections to ensure that the air exchange between the interior and the external environment is sufficient to control the moisture level inside the beehive. One method that deals with this challenge regulates the ventilation valve and at the same time monitoring the temperature loss [5]. Improper control of temperature and humidity in the context of changing weather condition might lead to large food consumption and it can have fatal consequences for a bee family.

Currently bee-families are monitored by manual inspection which is costly and cumbersome. Such systems can easily be automated and implemented by using smart sensor devices to monitor humidity, temperature.

B. Radio and sensor hardware choices

In order to relate the simulation model used in this paper to the real world settings we have used the parameters from selected HW modules (as presented on Fig. 2) as input for the simulations. The RF modules used in this work provide a transmit power range of: -11,-6,-3,0,1,2,3,4,5,6,7,8,9,10,11,12 ,13,14,15,27 dBm. These output power settings as a function of input power are presented in Fig 2.

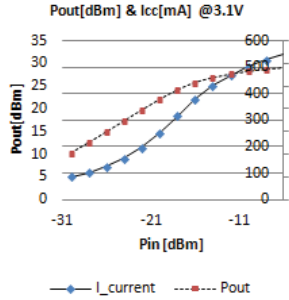


Fig. 2. Current consumption @3.1V for RF169MHz module with PA activated (leftmost) Hardware RF modules based on the wireless CC1125 device (rightmost).

C. An adaptive power control algorithm

The presented algorithm has been designed and tested in a farm application where a weather station sends data over a RF link to a gateway (every 10 minutes) which collects the data and uploading these to a Cloud, Fig. 4.

The basic algorithm flowchart is presented in Fig. 3. This algorithm will be invoked every time a package is sent.

```

This algorithm can run on RF1 or RF2 (*RF2).
(*RF2) - for algorithm running on RF2.

1. Power up IoT node RF1 (*RF2)
2. Initializing values:
L=0
R=19 // sizeof(T)-1
T[] = {-11,-6,-3,0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,27} // array of dBm values

3. If L<(R-1) is true then:
middle=round(L+R-1)/2
tx_power= T[middle]

4. Packet will be send with output power tx_power = T[middle] from RF1(*RF2).

4.1 If RF1(*RF2) received confirmation packet from RF2(*RF1) so:
if (send == confirmation) is true then:
last_good_power=tx_power
R=middle

4.2 Otherwise, when RF1(*RF2) did not received confirmation packet from RF2(*RF1) so:
if (send == confirmation) is false then:
L=middle

4.2.1 When RF1(*RF2) did not received confirmation packet from RF2(*RF1) and if (R=L && R<sizeof(T)-1) is true then:
R=R+1;
middle=R;
L=L;
txPower=T[middle]

5. Transmit output power is set again -> jump to point 3.

Before next packet is sent, again transmitted output power is calculated (from point 3).
If condition if(L<(R-1)) will be false and condition if(send==confirmation) will be true so if RF1(*RF2) received confirmation packet from RF2(*RF1), then minimum output power will be set, last_good_power will be lowest transmit power step where RF2(*RF1) is sending confirmation.
    
```

Fig. 3. Algorithm steps for RF Firmware modules (RF1 and RF2)



Fig. 4. Automatic Weather Station with RF1 module inside

This algorithm is implemented in simulation model for energy efficient adaptive power link analysis in next sections.

IV. AN ENERGY EFFICIENT ADAPTIVE WIRELESS LINK

The key factors for overcoming the challenges of creating an energy efficient link between the receiver and a transmitter discussed previously means dealing with the path loss and the noise sources in the link. To control these is tricky and complex because they vary as a function of time. Examples are: antenna position and polarization are changed by the wind; atmospheric losses changes with rain etc.; adjacent and co-channel interferences increase the receiver noise floor; and new object can arrive in the first Fresnel zone. However, some of these challenges have more impact on the link losses that other, thus the ones with the highest impact are explored and elaborated in the following sections.

A. Path loss

The free space loss is one of the most significant contributions to a wireless link loss. It accounts for the spreading of the transmitted energy as a function of distance. A commonly known formula which covers this is the Friis transmission formula (1).

$$\frac{P_r}{P_t} = G_r G_t \left( \frac{\lambda}{4\pi R} \right)^2 \tag{1}$$

Where: Pr and Pt are the received and the transmitted power respectively. Gr and Gt are the antenna gain for the receiver and the transmitter. λ is the wavelength and R is the distance between the receiver and the transmitter.

As discussed the path loss is a key contributor to the link loss. Thus, the path loss can be as large as 100 dB if the distance between the receiver and the transmitter is 10 km. An additional loss is the loss provided by objects that blocks the line-of-sight (LOS) between the transmitter and the receiver. Especially when an object is partly blocking the first Fresnel-zone has a considerably impact on the path loss [6].

Another key loss factor is the white Gaussian noise sources, which is random processes that occurs in the nature such as thermal vibration from atoms in conductors, black body radiation, etc. In addition to these are the manmade noise sources such as noise from electrical equipments and spill-

over from other radio sources placed in the same frequency range. These noise sources can be modelled by using an Additive White Gaussian Channel (AWGN channel) model which adds Gaussian noise power with a constant spectral density [6].

B. A simulation model

To proof link stability in different environments scenarios for the suggested adaptive power algorithm behaviour has been derived and simulated. This simulation model relates to real world settings by in-cooperating parameters from the previously discussed radio-hardware. This concept allows for experimenting and fine-tuning the model parameters to explore and evaluate the pros and cons of the proposed adaptive power algorithm.

The model consists of two parts, where the first part (model-1) is a model that is based on the equations discussed earlier in this paper. It calculates the signal to noise ratio (SNR) at the receiver end based on the parameters in table I and table II. The second model (model-2) simulates the bit-errors in a sequence of transmitted frames, which are transferred through an AWGN channel that incorporates the SNR from the model-1.

TABLE I.

Parameter	Value
Distance from the field to the farmer house	10 km
Height (over ground) of farm house antenna	3 m
Height of the antenna - the field IoT device	1 m
Attenuation of object in first Fresnel zone	20 dB
Antenna gain (Rx+Tx)	0 dB

TABLE II.

Parameter	Value
Operating frequency	169 MHz
Sensitivity	-120dBm @ 1200 bps, BER=10E-2
Modulation	2-FSK
Radio consumed power RX	69 mW
Radio consumed power TX	150 mW
Microcontroller consumed power	3 mW
PA efficiency	50 %

In general, the simulations assume that the ZigBee wireless transmission protocol is used and that some power saving can be achieved by the radio-interface [7]. The ZigBee protocol uses frames with an overhead of 31 bytes and a maximum payload of 127 bytes. Regarding the radio-interface power savings it is assumed that the microcontroller provides a power saving mode (sleep mode) when no traffic takes place and it is assumed that the sensors have a very low power-consumption (it can be ignored).

The second model (model-2) simulates bit-errors in a sequence of transmitted ZigBee frames. Basically, 2-FSK modulates a short ZigBee frame which only contains one acknowledge token. This frame is piped through an AWGN channel which has a SNR in the range of 0 to 20 dB. Finally, the frame is demodulated and compared to the modulated frame to detect bit errors. If one or more bit-errors are present in the frame it is assumed that these cannot be corrected, i.e. the frame is considered faulty and not received.

By deploying the adaptive transmit power algorithm to this model it can be explored and elaborated. Thus, this algorithm has been integrated into the model together with logging-information which at a frame to frame basis shows: the chosen transmit power level, frames that are in error, and the algorithm state information.

C. Elaborated simulation results

By using model-1 and the previously discussed simulation settings it is possible to estimate the total received power which is illustrated in (Fig. 5).

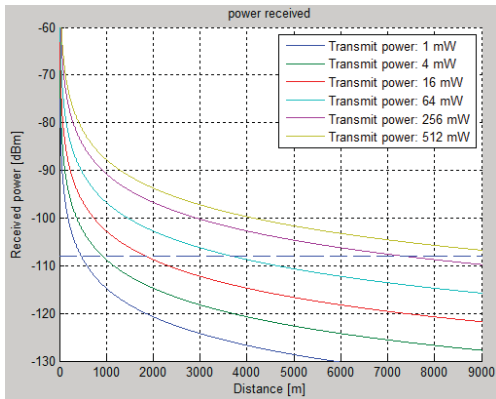


Fig. 5. The received power as a function of transmitted power and distance. The dashed line indicates the receiver sensitivity level.

As illustrated in Fig. 5 the minimum power settings (1 mW) provides a distance of approximately 0.5 km meters and the highest power settings (512 mW) a distance beyond 9 km. In this context it is noted that the datasheet for CC1125 transceiver uses a very low BER limit (10E-2) [8]. Most real world systems require a BER in the range of 10E-4 to 10E-5 [6] why the “real world” power level margin must be approximately 4 dB higher. However, as elaborated previously fluctuations in the received SNR must be expected. This means that it is not possible to select the transmit power level in a stationary manner, why it is common to add extra link-margin at the cost of extra battery power.

An alternative to a concept which uses fixed link power is an adaptive concept where the transmit power adapts to the instantaneous link and noise conditions. This has been

implemented in model-2 in form of an adaptive transmit power algorithm as previously discussed.

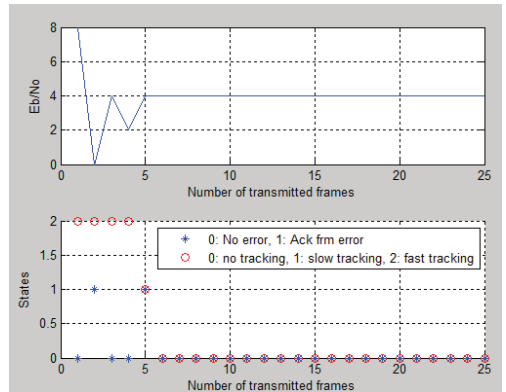


Fig. 6. Trace of the tracking algorithm. The upper figure shows SNR (Eb/N0) as a function of transmitted frame number. The lower figure shows the state-information of the tracking algorithm.

The behaviour of the adaptive transmit power algorithm is shown in Fig. 6 where the upper figure shows the used SNR (Eb/N0) as a function of the received frames. Similarly, the lower figure shows the state of the received frames (i.e. the crosses). If the state is 0 the frame is error-free, but if it is 1 the frame contains one or more errors. The circles indicate when: the algorithm does not change the power level (state 0), perform slow tracking (state 1), and perform fast tracking (state 2).

It is noted that Fig. 6 is one example run of the algorithm; nevertheless, it highlights and illustrates the important main points. The algorithm starts by setting the transmit power level to its middle value as seen in Fig. 6. Thus, the transmit power level (TPL) start with a SNR value of 10 dB. When the first error-free frame arrives the TPL adapt to this and changes its level to 0 dB. But, this level is too low why the next received frame is in error. Hence, the algorithm regulates the TPL to 3 dB where it settles, i.e. fast tracking is stopped. However, the next received frame contains errors as shown this triggers the slow tracking mechanism (circle is in state 1) which increases the TPL to 4 dB. From this point the rest of the frames are received without errors why the algorithm is in state 0, i.e. no tracking. It is noted that if one or more error-frames were received the slow-tracking algorithm would have increased the TPL further.

The energy savings which can be achieved by regulating the transmit power adaptively is shown in Fig. 7. The x-axis is transmitted bytes in the ZigBee frame and the y-axis is energy savings in percent. The lines in the figure map the energy usage as a function of number of transmitted bytes and the used transmit power level. It is noted that the energy usages are normalized with the highest energy value, i.e. a 24 dBm TPL with 128 bytes of payload. Some examples illustrate the possible savings. For example, regulating the power level down from 24 dBm to 18 dBm with a frame length of 128

bytes saves 52 percent energy per frame. Similarly, regulating down to 10 dBm and shorten the frame length to 60 bytes saves 80 percent.

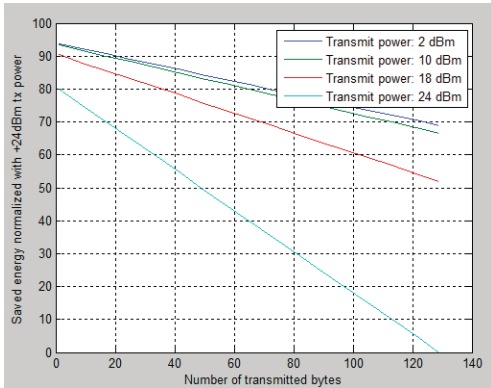


Fig. 7. Normalized saved energy as a function of transmit power and number of bytes transmitted. The highest setting (24 dBm and 128 bytes set to 100 percent and used for the normalization).

These energy savings can be visualized and elaborated in the light of using a solar power resource as discussed in [1]. The results of combining the settings from [1] with the adaptive transmit power algorithm is shown in Fig. 8.

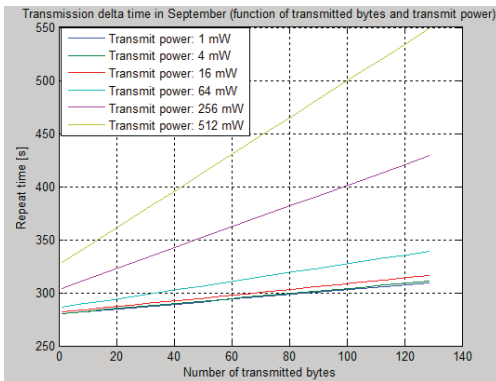


Fig. 8. The time between each transmission as a function of message length, transmit power and solar power available in september month (Denmark)

The “repeat time” shows how often a message can be send as a function of message length and the used transmit power. This provides some insight into the compromises that are needed with respect to the used average energy level provided by the solar cell. An example is using a TPL of 256 mW (24 dBm) which means that a message of length 128 bytes can be repeated every 430 second. However, lowering the TPL by

using the adaptive algorithm to e.g. 64 mW (18 dBm) with the same payload length the messages can be repeated every 340 seconds. It is noted that often it is possible to shorten the transmitted messages considerably as discussed previously. Using a shorter message provides additional savings, i.e. an adaptive link can be established which reduces the energy consumption and thereby sustain the usage of solar power as the main power source.

More generally, in addition to an adaptive transmission power algorithm similar IoT device behaviours can be programmed into its embedded microcontroller and thereby enable it to behave pseudo-intelligently. Such behaviour can be remotely programmed into it in form of manually setup the needed parameters. An example could be the time before harvest (September for many grain types in Denmark) where the farmer decides to monitor the humidity in the grain and therefore program the field IoT device to only transmit these data with a fixed time interval. Similarly, an pseudo-intelligence system in the IoT device could monitor the available battery-power and regulate the transmitter parameters further. So, if the field IoT devices has been running for some time without any harvested power the transmit parameters can be adjusted so the battery last longer and when the solar energy is back on the parameters can be adjusted so the battery charges between the transmissions.

V. CONCLUSION

An energy efficient adaptive wireless link for farms which uses an adaptive transmission power algorithm has been explored and elaborated by using two simulation models. These models simulate the vital parameters in a wireless link based on ZigBee technologies and real world hardware in form of the transceiver module CC1125 from Texas Instruments. It has been found that it is possible to place a wireless IoT device far from the farm-controller by using the elaborated adaptive technology. Hence, power savings in the range from 50 to 70 percents can be achieved by regulating the transmit power level instead of setting it to the much higher fixed level which includes a considerable safety margin. Similarly, considerable savings can be achieved by adjusting the amount of payload so it is sufficient to transmit vital data like those presented and discussed previously in the bee life monitoring scenario. Additionally, the presented simulation results support and enable a future design of adaptive algorithms for IoT nodes, i.e. algorithms could scale the IoT node power consumption according to weather conditions and thereby behave adaptive with respect to the sustainable power available.

ACKNOWLEDGMENT

This research was partly funded by the FP7 EU FRAMEWORK PROGRAMME under grant agreement No. 604659 (project CLAFIS).



REFERENCES

- [1] T Blaszczyk and P Lynggaard. (2015) "An Energy Efficient Adaptive Wireless link for Farms based on IoT technologies" [Online] <http://vbn.aau.dk>
- [2] SIGFOX. (2016) SIGFOX IoT device. [Online]. <https://www.hidnseek.fr/shop#st1a>
- [3] Blob. (2016) GitHub. [Online]. <https://github.com/hidnseek/hidnseek/blob/master/arduino/libraries/HidnSeek/HidnSeek.cpp>
- [4] Xiao Chen and Neil Rowe, "Saving Energy by Adjusting Transmission Power in Wireless Sensor Networks," in *AsiaSim 2012*.: Springer, 2011, pp. 37-44.
- [5] Slawonir Trzybinaski, "Gospodarka w ulach styropianowych," *Bee&Honey*, 2013.
- [6] B Sklar, *Digital communications, fundamentals and applications*, Second edition ed.: Pearson, 2002.
- [7] ZigBee Alliance. (2008) ZigBee specification. pdf.
- [8] Texas Instruments. (2015) CC1125 Ultra-High Performance RF Narrowband Transceiver. pdf.

# CCM-R: Secure Counter Synchronization for IoT Wireless Link

Upal Roy  
DTU Compute  
Technical University of Denmark  
Copenhagen, Denmark  
Email: s127819@student.dtu.dk

Jiachen Yin  
DTU Compute  
Technical University of Denmark  
Copenhagen, Denmark  
Email: s127628@student.dtu.dk

Birger Andersen  
DTU Diplom  
Technical University of Denmark  
Copenhagen, Denmark  
Email: birad@dtu.dk

**Abstract**—In this paper we propose and evaluate a new version of the CCM mode of operation, CCM-R, which is an extended and alternate version of the original CBC-MAC with Counter Mode (CCM) that was created to address the problem of counter synchronization. While CCM is considered secure when used/implemented correctly, it is still vulnerable to Denial of Service (DoS) attacks where packages in the stream are removed or modified with the purpose of destabilizing the synchronization of the counter states. Another possible way it can be destabilised is through random loss of packets due to noise or weak signals on wireless link. Therefore, in order to solve this problem, we have proposed a solution where we have embedded a secure counter synchronization approach into the original CCM mode of operation. The approach makes it immune to package loss and DoS attacks. The overhead is also kept at a minimum, making it suitable for low-power wireless IoT devices.

**Keywords**—CCM-R; IoT wireless link; CCM; counter synchronization; Network Security.

## I. INTRODUCTION

CCM (CBC-MAC with CTR mode) was created by Niels Ferguson, Doug Whiting and Russ Housely [3]. Concern was the main issues of wireless IoT devices, such as data link issues where messages are usually small (need low overhead) and energy efficiency.

Often the messages are not retransmitted in cases where messages are lost or received with errors in order to save energy and not to use more bandwidth. In our case (with our protocol stack developed with wireless weather stations in mind), the messages lost or received not in order will be retransmitted three more times and then withdrawn in cases messages are still not received correctly.

At this time, many points of contention came up in the sense that while implementing the RFC version of CCM mode would definitely work it would lead to problems in case the recipient loses track of the counter and would not be able to authenticate and decrypt the message. According to these issues, it is very urgent to have efficient counter synchronization and therefore we modified and extended CCM to fit in with our purpose and to actually customize the code so that it would function best with our situation. At the same time, we carefully considered the possible attacks

against counter synchronization that could lead to, 1) counter would become unknown, 2) synchronization would fail (DoS attack) and/or, 3) a battery powered device would use up the power during many synchronization attempts (power drain attack). The security concerns of our proposed solution will also be then discussed. Finally, we want to introduce as little overhead as possible by only adding few bits/bytes to messages, which can be important in a battery powered domain.

## II. ENVIRONMENT

We are here describing hardware and software used. The most important software that was used to build and run application with protocol stack is MPLABX [11], which was used to code and build the software directly to the hardware. The connection kit we used is the PICkit connector [12], which allowed us to build and execute the program by using MPLABX. We were using MPLABX 3.10 to code and program a 16-bit PIC micro-controller, which is made by microchip complied by using a XC16 compiler. For testing purpose we had circuit boards of a weather station and an IoT gateway both with wireless transceivers with PIC microcontrollers. More precisely, our hardware was a RF1 module, which consists of three individual printed circuit boards (PCBs), a 9V battery and a switch, as well as a RF2 module that consists of two 1.5V batteries, a PCB with serial connector and a switch. RF1 was for weather station while RF2 was for IoT gateway.

## III. ALGORITHM

The CCM-R algorithm consists of three essential parts: the authentication algorithm, the encryption algorithm, and the random number generator.

### A. Authentication

The authentication part is one of the most important parts because it ensures that the message has not been modified by any means and that the message is from expected source. The first step is to compute the authentication tag  $T$ . This acts as the MAC tag. This is done by using CBC-MAC. First a sequence of blocks has to be defined,  $B_0, B_1, \dots, B_n$  and then the CBC encryption operation is run on these. CBC-

MAC uses an initialization vector(IV) of 0. The first plain-text block is XOR'ed with this IV and this result is then encrypted with the key to form the cipher-text for the first block. This continues till the result of the second to last block i.e. the cipher text of it is XOR'ed with the last block plain-text and then this is encrypted with the Key to form the cipher-text of the last block. The most significant bytes or MSB of the cipher-text of the last block is taken. Assuming blocks  $B_0, B_1, \dots, B_n$  are defined, then the first ciphertext block,  $X_1$  can be represented as:

$$X_1 := E_{K, B_0} \oplus IV \tag{1}$$

Based on that the final cipher text block can be represented as:

$$X_n := E_{K, B_n} \oplus X_{n-1} \tag{2}$$

Thus the authentication field value or the MAC tag is defined as:

$$T := MSB(X_n) \tag{3}$$

**B. Encryption**

For encryption we are using counter mode. Counter mode uses a nonce and counter as input generally where there is a fixed nonce and the counter is updated using the nonce as a base. But in our case we use  $CTR_0 = \text{nonce}$  and as the counter updates by 1,  $CTR_1$  just becomes the "nonce +1" and  $CTR_2$  as "nonce + 2". The counter is generally a fixed value throughout an entire session. For our implementation we used a variable and secret 8-byte counter which changes for each packet being sent. The counter is generated using a true random number generator. This would increase the security since using a fixed nonce with counter generates sequences which with enough time and computation power can be broken. Another advantage of generating our own randomized counter is that since we are sending it along with the packet there will never be any counter synchronization problems. If packages are lost the counter would become unsynchronized and cause incorrect authentication and decryption of the packet. This is the major advantage of our implementation as CCM-R in the sense that it completely takes out the need for maintaining synchronization between the counters on the receiving and sending ends using additional synchronization messages. Due to the entire counter being sent the only thing that needs to be taken into account is that the counter generating sequence is the same, i.e.  $\text{counter}_0 + 1, \text{counter}_0 + 2$  and so on is the same on both sides which is programmatically very simple. The first CTR or  $CTR_0$  is generated using a true random number generator function. This counter is then encrypted with the secret key and the result is then XOR'ed with the plain-text first block. The next keystream is block is generated by encrypting successive values of the counter. So for the final block to be encrypted the final counter which is  $CTR_0 + \text{number of blocks encrypted}$  is encrypted with the key and the result is then XOR'ed with the plaintext.

For the first block being encrypted, it is represented by:

$$X_1 := E_{K, CTR_0} \oplus B_0 \tag{4}$$

And therefore the final block to be encrypted is:

$$X_i := E_{K, CTR_i} \oplus B_i \tag{5}$$

**C. Output**

The output which is to be sent is then the encrypted message from  $X_1$  to  $X_i$  in counter mode which we dub as "m", the MAC tag and the CTR0. While the cipher text is secure enough to be sent as is, the MAC tag and the counter is then encrypted with AES-128 [2] to provide added security since this is impossible to crack as of now. To note is that the MAC tag and the  $CTR_0$  are each 8 bytes in length to fit in with CCM mode restrictions.

**D. Decryption**

For decryption the exact opposite needs to be taken place. Using AES-128 the  $CTR_0$  and MAC tag are decrypted. The  $CTR_0$  along with the key is passed through the encryption cipher and the result is XOR'ed with the first cipher text block and the plaintext of the first block is the result. Same is repeated till the last block, where the final counter which is basically counter + number of blocks is passed through the encryption cipher and the result XOR'ed with the last cipher text block to produce the plaintext.

**E. Frame Format**

Although not being part of CCM-R, we have present its implementation at the link layer in details. Some parts of frame header was already defined by the protocol in previous version where AES-CBC was used but with only encrypted CRC fields for protecting integrity.

Figure 1 shows the wireless link frame format while packages are transmitting. It starts with one-byte value identifying the packet length, and 4-byte value for the address from the destination, a 2-byte CRC16 value, and then, the AES-CTR mode is used for encrypt the set of 4-byte source address, 1-byte command, status value along with the data that needs to be transferred. The initial counter value as well as the CBC-MAC Tag value are combined into one block of 16-byte encrypted by AES-ECB mode transferred along with the entire frame.

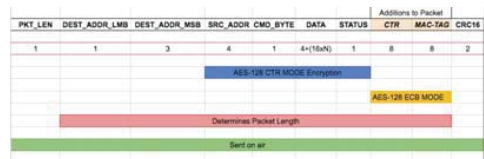


Fig. 1. Wireless Link Frame Format

IV. EVALUATION

Our implementation of CCM-R version deviates from the RFC version considerably. Also concerning of the issues of what a wireless link would occur as well as the importance to solve the counter synchronization problem to maintain the security of the system, a sequence of packets lost and receiver does not know about the sequence number of the counter and which counter will be the next one, and also an attack could remove or change a long sequence of packets, e.g., DoS attack. Both of the cases discussed above could have terminated the network communication and cause threats to the system for instance. In such a situation, one counter synchronization algorithm would not help a lot as for instance the attacker could just simply remove or change a long sequence of packets once again. Therefore, in order to maintain the security level of the RFC CCM version and also improve it by solving the counter synchronization problem, CCM-R comes to the place. Figure 2 shows the basic overview

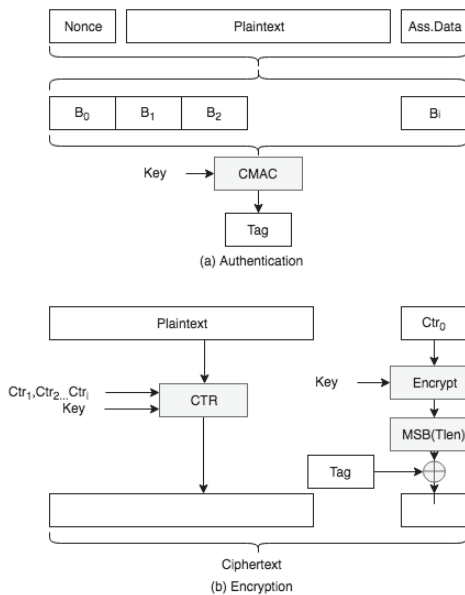


Fig. 2. CCM Structure and Encryption/Decryption

of the RFC CCM version. The diagram also describes the entire encryption and authentication happening and also the results, which are being sent. Figure 2 shows CCM-R, where the packet structure can be easily seen. We put the 8-byte initial counter, generated by using a true random number generator, for the AES-CTR mode and 8-byte CBC-MAC tag value as one block encrypted and decrypted by AES-ECB when transfer the packet. We are using the tag value to check if the packet is sent correctly and send the initial

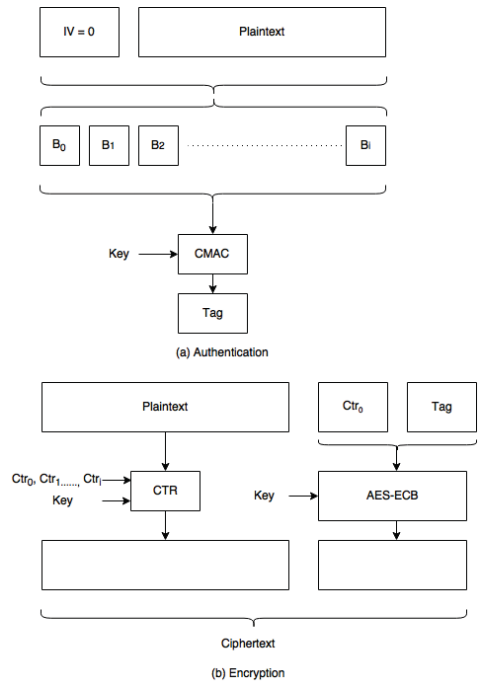


Fig. 3. CCM-R Structure and Encryption/Decryption

counter for decrypt the packet in AES-CTR mode to tell the packet sequence number. As we can see here, the difference we make here from the RFC CCM version is that we take out the initial counter and send it along with the CBC-MAC tag value as one 16 bytes block by using AES-ECB to encrypt, which can also help CCM-R to solve the counter synchronization problem that exists in the RFC CCM version, where if the packet is not completely transferred, the counter synchronization will be messed up. Since using this strategy we solved this problem by sending the encrypted initial counter. There are several test cases are made in order to make sure there are no such security holes in the system that can be exploited by attackers to potentially steal the data or cause irreparable changes to it, and also how is it better than the current version. The tests are all on CCM-R by comparing results on before encrypt and after decrypted, also changing some parts of the data to see how CCM-R deals with it and by testing for a weaker signal that leads to more packet loss.

Concerning the possible attacks against CCM-R, such as DoS and analyzing messages. A DoS attack, which is cyber-attack that the attackers is attempting to make the users not able to access the information or service is immune by CCM-

R since the message itself is self-contained. Also to be noted, the counter value is randomized with an 8-byte value, for instance the counter might be used for a second time or maybe even third time. CCM-R is also very secure even with this issue, because the counter is secret, which is encrypted sent along with the Tag by using AES-ECB. It is very rare to happen that both of the Tag and counter will be the same again, and we believe that CCM-R is enough secure at the current level of the network and at least at the same security level as the original CCM. Therefore, CCM-R should provide the following security services:

- Confidentiality, which ensures only authorized parties can access the information.
- Authentication, which provides proof of authority of the user.
- Integrity, which ensures changes can be detected.
- Access control in relations with layer management.

Theoretically, the meet-in-the-middle attacks can be used to limit the key size of  $2^{k/2}$  operations, where k is the size of the key in bits. Due to CCM-R still remains the same key size as CCM, it is secure enough against the attacks to  $2^{64}$  steps of operations.

#### V. EXPERIMENTS

As mentioned in the previous section, there are several tests being made. In order to prove that CCM-R works properly, we therefore executed the program with the CCM-R cipher on the software simulators along with two RF modules to compare the inputs and outputs, which lead to great success after several attempts. For the security manner of CCM-R, we decided to test a bit more on the algorithm and see how it reacted to protect the data packet. Therefore, we decided first to make the signal of the packet transfer between the gateway and the simulator for the packet sender weaker by using attenuator, with the purpose to take into account of reality that there may be some jamming of networks leading weaker signal or lead to more packet loss.

By testing this, the same equipment was then being used with an attenuator to change the strength of the signal. The CBC-MAC tags for authentication of the messages from both ends keep match each other until the weakest point of the attenuator can be changed to. When the signal is very weak, the packet cannot be received, and therefore we decided to drop the packet that cannot be received after three times attempts, and for testing purpose the interval of the attempts was set to 3 seconds, but it can be changed to have longer time interval, for instance 20 seconds for reality. There is another test for CCM-R, which is to deliberately change of the packet structure and the initial randomized counter, on the purpose of where the hackers tries to attack and modify the packet or the generated random counter. By doing the tests, we could make sure that by solving counter synchronization problem by CCM-R, it is also being more secure. As has

been explored, we believe that CCM-R is secure enough and also authenticated.

#### VI. CONCLUSION

In this paper we have presented a simpler solution to the counter synchronization issue faced with the use of CTR mode. We have discussed the implementation as well as the tests we have performed to ensure the working and security of the algorithm. CCM in the original RFC version matches well with the CIA (Confidentiality, Integrity and Availability) criteria and was a very useful algorithm to use. Regarding to the pivotal issue of synchronization of counters as well as the possible attacks against CCM-R such as DoS and analyzing messages, the security level of the CCM-R has been improved by solving the counter synchronization problem as well as the solutions against those possible attacks. In spite of this we believe due to the simplicity of solving the counter synchronization issue and also maintain security level it could be used in other real life scenarios practically to complement the original CCM.

#### VII. RELATED WORKS

CCM was originally designed by D. Whiting, R. Housely, and N. Ferguson [3]. CCM is considered secure when used correctly but does not provide a solution to the counter synchronization problem. In case an unknown number of packages have been lost due to link layer failures or DoS attack, recipient will fail to decrypt because the counter values are no more in sequence. Solutions where counters are displayed in clear-text like in WiFi are in general unsafe or at least risky.

One approach to counter synchronization is taken in the SPINS protocol [7]. Here counters are agreed upon and synchronized by means of plaintext messages, meaning that the counter value is not secret and therefore the counter (assumed never reused) only guarantees freshness and prevents against reply attacks. However, it was suggested to send the counter encrypted with each message in order to protect against counter synchronization DoS attacks, but this approach was not further developed.

Zigbee [8] has been based on the SPINS approach but does not include encrypted counter values. IP ESP [9] takes a different approach and uses a window where a limited number of succeeding counters are tried at recipient side until successful authentication (including integrity validation). Counters are never transmitted as plaintext. Start value is always zero. IP ESP will fail in case the resynchronization fails. Therefore, a DoS attack against a number of succeeding packages can efficiently interrupt an IP ESP session. Also power drain might be consequence.

MiniSec [10] uses the time as replacement for the counter and messages are accepted at recipient whenever their timestamps (included with the packages) are within the allowed window. This approach is exactly opposite to SPINS which

is not transmitting the counter value. MiniSec implements weak freshness as two messages with same timestamp will be accepted if they arrive within the window. A message with older timestamp than the previous will not be accepted. SPINS and IP ESP provides strong freshness (another message with same counter will never get accepted).

#### ACKNOWLEDGMENT

This research was partly funded by the FP7 EU FRAMEWORK PROGRAMME under grant agreement No. 604659(project CLAFIS) [1].

#### REFERENCES

- [1] CLAFIS-Project, *Crop, Livestock and Forests Integrated System for Intelligent Automation*, <http://www.clafis-project.eu/>, Website
- [2] Joan Daemen and Vincent Rijmen, *The Design of Rijndael: AES-The Advanced Encryption Standard*, Dated: 2002, Springer, Journal Submission
- [3] D. Whiting, R. Housley, N. Ferguson, *Counter with CBC-MAC*, RFC, <https://tools.ietf.org/html/rfc3610>, RFC/Website
- [4] Jakob Jonsson, *On the Security of CTR+CBC-MAC*, RSA Laboratory Stockholm.
- [5] Pierre-Alain Fouque, Gwenalle Martinet, Frederic Valette and Sebastien Zimmer, *On the Security of the CCM Encryption Mode and of a Slight Variant*, Ecole Normale Supérieure/DCSSI Crypto Lab/CELAR, <http://www.di.ens.fr/~fouque/pub/acns08.pdf>, Journal/Paper/Website
- [6] William Stallings, 'Network Security Essentials Applications and Standards, Fourth Edition', 2011, Textbook
- [7] Adrian Perrig, Robert Szewczyk, J.D. Tygar, Victor Wen and David E. Culler, *SPINS: Security Protocols for Sensor Networks*, Wireless Networks 8, 521?534, Kluwer Academic Publishers 2002.
- [8] Paolo Baronti, Prashant Pillai, Vince W.C. Chook, Stefano Chessa, Alberto Gotta, Y. Fun Hu, *Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards*, Computer Communications 30 (2007) 1655?1695, Elsevier 2007.
- [9] Stephen Kent, RFC 4303 IP Encapsulating Security Payload (ESP), IETF 2005.
- [10] Mark Luk, Ghita Mezzour, Adrian Perrig, Virgil Gligor, *MiniSec: A Secure Sensor Network Communication Architecture*, IPSN'07, April 25-27, 2007, Cambridge, Massachusetts, USA, ACM 2007.
- [11] MPLAB-X IDE, <http://www.microchip.com/mplab/mplab-x-ide>
- [12] PIC KIT, [www.microchip.com/pickit3](http://www.microchip.com/pickit3)

# Comparative study of Internet of Things Infrastructures & Security

Bhupjit Singh  
DTU Diplom  
Technical University of Denmark  
Ballerup, Denmark  
bhsi@dtu.dk

Bipjeet Kaur  
DTU Diplom  
Technical University of Denmark  
Ballerup, Denmark  
bika@dtu.dk

**Abstract**— With increasing use of IoTs in diverse fields has increased the demands of different parameters for high level of security, trust and applications. Several companies have invested millions of dollar to fulfill the needs of the market which has given rise variant infrastructures of IoTs. In this paper we have compared the different infrastructures and their parameters along with establishing the requirements of security in IoTs. The various vulnerabilities in the IoTs architecture and consideration for privacy control is also discussed. After identifying the security issues in IoTs , this paper suggests solutions from existing technologies as a starting point for establishing a standardized security paradigm in IoTs

**Keywords**—LoRaWAN; SigFox; Symphony; IoT architecture; requirements for security paradigm in IoTs; security issues in IoTs;IoT Security

## I. INTRODUCTION

The term Internet of Things (IoT) has gained enormous popularity with the explosion of wireless sensor networks, smart meters, home automation devices, and wearable electronics. The IoT spans long-range outdoor networks such as the smart grid and municipal lighting, as well as shorter-range indoor networks that enable the connected home and residential security systems. Each year will see exponential growth in devices connected to the Internet. Gartner predicts there will be 25 billion connected “things” by 2020. [1] It is not matter of things but actually connectivity and services.

Numerous companies have introduced innovative solutions for the IoT market that provide security, status, and other convenient services. A connected system architecture comprises a number of wireless nodes ranging from simple remote control devices to complex wireless networks featuring a gateway to connect to the Internet. One of the major issues for machine to machine, M2M communications used for applications like the Internet of Things, IoT is to enable communications over long ranges using very low power levels [2].

This paper is divided into six sections; Introduction, Infrastructure Of IoTs, Comparative Study Of IoT Infrastructure, General Architecture of IoTs Network, Security Issues in IoTs Networks, Conclusion and Future Work

## II. INFRASTRUCTURES OF IOT

### A. LoRaWAN

One scheme for addressing long range communication dedicated to Internet of things is known as LoRa. It gains its name from the fact that it is able to provide 'LongRange' communications using very low power levels. It uses low-power, long-range wireless connectivity in the widely used sub-GHz band.

It is a LPWAN (Low-Power Wide-Area network), currently deployed in Western Europe, San Francisco, and with ongoing tests in South America & Asia. These are best suited for connecting devices that need small amount data and long battery life.

Actually LoRaWAN is based on server-side implementation of a multiple access protocol. It is specially designed to minimize collisions with a large number of endpoints. It requires a server application to run the MAC functions over a network connection. Its architecture is typically laid out in a star-of-stars topology in which gateways are a transparent bridge relaying messages between end-devices and a central network server in the backend.

It is designed primarily for uplink-only applications with many endpoints, or applications where only a few downlink messages are required (limited either by application or by number of endpoints). In this type of architecture, the gateway within the same network require synchronization. Communication between end-devices and gateways is spread out on different frequency channels and data rates. The selection of the data rate is a trade-off between communication range and message duration. Different data rates do not interfere with each other instead create a set of “virtual” channels increasing the capacity of the gateway.

The LoRaWAN network server manages the data rate and RF output for each end-device individually by means of an adaptive data rate (ADR) scheme that is typically updated once every 24 hours. Multiple layers of security/encryption (EUI64 on network level and application level and EUI128

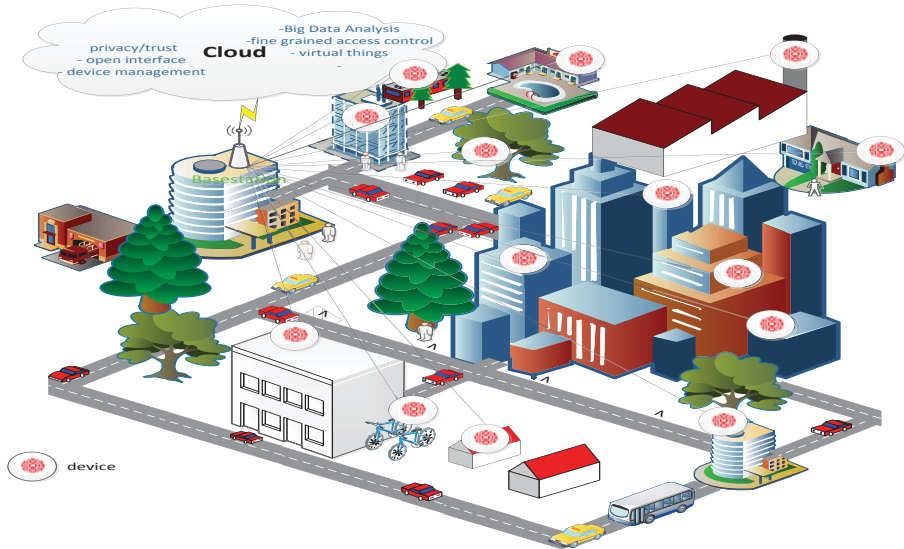


Figure:1 IoT Network in Smart City Application

device specific key). AES CCM (128-bit) for encryption and authentication is available in this infrastructure. It works within the confines of the ETSI 1% and 10% duty cycle on transmission time in the 868 bands. Draft revision of class B for downlink nodes that can poll for a beacon every 1s to 128s ( $2^n$ ) where n is 0 to 7. It has also got antenna diversity. It is due to that all gateways' listen to the same uplink channels. LoRaWAN has Adaptive Data Rate which is driven by the server, if a node's link suddenly fades, the server has no way of telling it to change spreading factors to compensate. ADR for LoRaWAN issued to optimize the capacity of the channel.

**B. SigFox**

SIGFOX is an operated telecommunication network, dedicated to the Internet of Things. It is an operated network, meaning you do not have to handle any installation or maintenance operations.

SIGFOX is seamless and out-of-the box, allowing you to forget about communication and keep focused on the core of your project. SIGFOX is operational in 24 countries covering about 1.4 million Km<sup>2</sup> and 358 million people. This is world largest IoT network.

SIGFOX allows a bidirectional communication, both from & to the device. The communication is always initiated by the device. The SIGFOX network is designed for small messages sent every now and then. It is not appropriate for high-bandwidth usages (multimedia, permanent broadcast). Its focus

on energy efficiency allows you to build connected devices able to last years on a standard battery.

SigFox sets up antennas on towers (like a cell phone company), and receives data transmissions from devices like parking sensors or water meters. These transmissions use frequencies that are unlicensed, which in the US is the 915 MHz ISM band; the same frequency a cordless phone uses. (Europe has a narrower band around 868 MHz, and most of the world has some version of this band either like the US or Europe, all with different rules that govern their use.) SigFox wireless systems send very small amounts of data which is 12 bytes and at very slow rate of 300 baud using standard radio transmission methods namely phase-shift keying – DBPSK – uplink and frequency-shift keying – GFSK –downlink. The long range is therefore accomplished by very long and very slows messages. This technology is a good fit for any application that needs to send small, not very frequent bursts of data. Examples such as basic alarm systems, location monitoring, and simple metering are one-way systems that might make use of such network infrastructure. In these networks, the signal is typically sent a few times to “ensure” the message goes through. While this works, there are some limitations, such as shorter battery life for battery-powered applications, and an inability to guarantee a message is actually received by the tower.

Another way to design a network is bi-directionally (like your cell phone). SigFox has not deployed any bi-directional networks, though they have said to be working on the



technology. If they are successful in deploying a two-way network, this will enable a wider variety of applications on their networks, though it will not have a symmetrical link because of the underlying technology they have chosen. SigFox has faced challenges in US due the law that limits the use of the unlicensed radio spectrum with the maximum transmission time on the air to be 0.4 seconds. But SigFox transmissions are 3 seconds or so, this makes too difficult to enter the market and now it requires a new architecture design. The frequency band in the US is also subject to much higher levels of interference than the band SigFox uses in Europe. This issue has brought problem in SigFox business for example: the pet tracking company, Whistle, announced a partnership to sell a solution on the SigFox wireless network in May 2014, but has been unable to ship product.

C. *Symphony*

It is developed by Link-labs to overcome the pitfalls of the LoPoWAN. It guarantees message receipt. There is very high loss of packets in case of SigFox and LoPoWAN. Symphony link acknowledges every message both in uplink and downlink. It uses light TCP like architecture. Symphony Link allows updates the host firmware on the device after it has been fielded. This is huge advantage early in the IoT evolution, as it allows customers to get to market more quickly, and with less risk. This lowers the network management struggles in case of networks with hundreds of device in large networks.

Symphony Link uses the Frequency Hopping Listen Before Talk plus adaptive frequency agility band, thus removes the duty cycle limit. In the 900 MHz Band, there is no duty cycle limit. The duty cycle of 1% prevents LoRaWAN from being used in systems that need the ability to send lots of data at a time. Since Symphony Link is a synchronous protocol, repeaters allow users to expand the range of the network dramatically without impacting latency. In Symphony Link, the host device configuration is the same for all devices of the same type, and key exchange is handled via PKI based Diffie Hellman AES architecture. Symphony Link infrastructure, before every transmission, an end device calculates the reverse link to the gateway, and adjusts its transmit power and spreading factor or modulation rate to match. This way node throughout the network has a balanced link budget. Close nodes are transmitting quietly and quickly, and far nodes are transmitting loudly and slowly. ADR in Symphony Link is about optimizing performance and reliability. Symphony Link uses a dynamic channel mask that is controlled by the gateway, it ensures as few collisions as possible. By using asynchronous features like slotting, and uplink/downlink coordination, a Symphony Link network has over 4 times the capacity of LoRaWAN. And when you couple that with quality of service, Symphony Link is a much more robust choice for users that need it.

III. COMARATIVE STUDY OF IOT INFRASTRUCTURES

No single architecture and service provider can cover all the application and services needed for the diversity applications in the world of internet of things. Each service provider aims at high availability and intensive level of services but security in Internet of things restricts its way. The Figure:1 shows the comparison between different IoT infrastructure. The internets of things' networks are highly vulnerable to security risks.

	Range	Data rates	Power consumption	Duty Cycle	Modulation
LoRaWAN	10km Bidirectional	5000 bit/s	5-10 years	1%	chirped spread spectrum (CSS)
SigFox	Bidirectional	12 bytes/s	Less than 5 years	1%	BPSK
Symphony	Greater than 10 km Bidirectional	Adaptive Data Rate	7-10 years	No limit	Frequency hopping plus adaptive frequency agility band

Table:1 Comparitive table at different IoT infrastructures

IV. GENERAL ARCHITECTURE OF IOT NETWORKS

Figure 1 shows the smart city infrastructure in which a single base station can cover several devices and the cloud solution will provide privacy, trust–open interface, device management, The IoT Architecture can be stacked into 4 layers in which the most basic level layer is the perceptual layer (also known as recognition layer), which collects all kinds of information through physical equipment and identifies the physical world, the information includes object properties, environmental condition etc; and physical equipment include RFID reader, all kinds of sensors, GPS and other equipment. The key component in this layer is sensors for capturing and representing the physical world in the digital world. The second level is network layer. Network layer is responsible for the reliable transmission of information from perceptual layer, initial processing of information, classification and polymerization. In this layer the information transmission is relied on several basic networks, which are the internet, mobile communication network, satellite nets, wireless network, network infrastructure and communication protocols are also essential to the information exchange between devices. The third level is Data Fusion layer. This layer will set up a reliable support platform for the application layer, on this support platform all kind of intelligent computing powers will be organized through network grid and cloud computing. It plays the role of combining application layer upward and network layer downward.

The application layer is the topmost and terminal level. Application layer provides the personalized services according to the needs of the users. Users can access to the internet of thing through the application layer interface using of television, personal computer or mobile equipment and so on. Network security and management play an important role in above each level. Then we will analysis the security features.

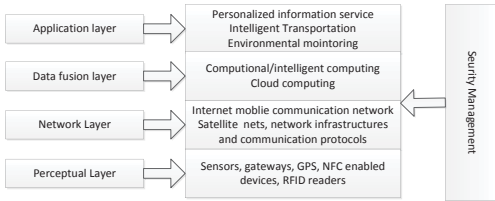


Figure 2: Internet of things layered architecture

V. SECURITY ISSUES IN IOTs

The issue of security is one of the most important issues considered as IoTs devices should be able to communicate in heterogeneous system to provide on the clock service in a long term deployment without having to perform regular checks. This setup leads to various intermittent and locale specific failures and could also lead to more permanent failures. Some of these failures in IoTs are covered by the obvious redundancy that is needed in such kind of deployments but due to the demand for IoTs to perform consistently and have the ability to recover from security attacks to normal operations. Thus, the security solution should cover the possibility of security updates, easy connection, attack detection capabilities with a standardization in IoT architecture among all layers to operate in different modes to provide networking with the bare minimum services in order to convey and recover from various security attacks. These modes would be able to provide attack detection, diagnose, apply repairs and countermeasures, there is also the need to keep in mind the computational limitations of IoTs while building such a security solution [9]. To recover and repair, it is sometimes needed to go offline, reprogram, go into recovery modes. Thus, in order to switch between recovery mode and other modes which might be without networking, ad hoc connections with easy connection, authentication and small overheads for configuration of network with light weight security techniques will be needed. The above mentioned factors are a good starting point for the security paradigm in IoTs.

The data protection and privacy is one of the key aspects with respect to IoT as access control of data should ultimately be the decision of the user.

Access control in IoTs has to consider aspects such that limiting or granting access is on the discretion of user. It should possible to give and remove access to various systems involved, on the fly with some kind of leasing. The type of access to data should also be deterrent to context of data usage, that is the data available from IoTs should be

categorized into various data sets with heuristic trust- so that data is only to be used in specific context. Certain assurances and certification could also be provided that data would not be used out of context of pre agreed terms (some legislations and other policies would be needed to be passed in order to fully realize this aspect [9])[10].

Influenced by existing solution, data distortion and data encryption is the driving force along with key management and authentication as IoTs integration is within heterogeneous, multi-layer networks. [10]

Identify security issues in the specific layers: [11,14]

1. Security challenges faced in perceptual layer are node authentication, confidentiality of information. Attacks like distributed denial of service and poor physical security with respect to installation of IoTs “things” cause separate set of problems.
2. In network layer security attacks like man in the middle attack and counterfeit attack are experienced along with data congestion and other problems relating to network layer are consistent in this layer. As perceptual layer and network layer are very closely related, problems like exploiting devices through unsecure network services are common in this layer.
3. In data fusion layer, malicious information, attacks from internet due to lack of transport encryption, insufficient authentication are common along with other insecure cloud interface.
4. In Application layer privacy protection due to data sharing plays an important role with respect to access control along with all the implications of data privacy. The data fusion layer works closely with application layer thus the issues from data fusion layer related to data integrity and corruptness creeps in this layer.

Proposed security solution

To deal with various security issues identified in the above section, a need for a framework specific for IoT security and privacy which has layer specific attack detection and repair capabilities (discussed in detail in next paragraph) along with privacy constraints. It should be able to determine and define the context of data in real time and dynamic propose privacy policies on the fly. It should also be able to facilitate secure inter domain data interaction and data fetching or querying in line with the various aspect of data access control discussed above.

In perceptual layer, equipment backup or limiting access to the site could be one immediate solution to counter poor physical installation of nodes. Illegal node access could be avoided by node authentication based on various attributes of the hardware by means of a digital certificate within the confines of the extranet established by VPNs to provide data integrity and confidentiality between nodes to gateway.

In network layer, there exist a set of communication security solution but it is difficult to be applied in IoT systems, as it should contain identity authentication along with

confidentiality and integrity mechanism, thus public-key cryptography to sign resources to guarantee origin authenticity and integrity of delivered information. IPsec in network layer can provide integrity, authenticity and confidentiality. But to counter DDoS attacks due to less processing power can be quite severe, limiting access to the nodes to only through VPN could provide some control but on the other hand gateways needs enhanced processing and inter domain communication capabilities.

In the data fusion layer, there are various kind of solutions like strong encryption algorithms, data validation algorithms along with other data verification logics can be applied with existing two factor, heavy weight encryption. The cloud based web interface should not be susceptible to XSS, SQLi, CSRF with bogus attempt detection capabilities.

In the application layer, key agreement across heterogeneous networks is a key aspect which can be covered by the existing solutions, but the privacy aspect of privacy protection is discussed in the conclusion. TLS/SSL protocols while transiting networks, with message payload encryption with key handshaking and data verification are needed in this layer.

## VI. CONCLUSION AND FUTURE WORK

To conclude the security issues in IoT, communication security and data privacy are two different aspects of security which should be addressed in their own right. The data protection should be such that, it should provide a balance between providing access to data. Its aim is to achieve full exploitation of data to reap benefits of IoTs without users worrying about repercussion and implications for providing access to their private data. This could be achieved via Anonymity and Pseudonymity[15,16] along with legislations and heuristic trust solutions which could provide access to the data with respect to the context of data usage.

The aspect of communication security has couple of key points to be considered, firstly, security solutions should be light weight due to the limitation of computational power of IoTs especially in perceptual and network layer. Secondly, it should have the ability to detect and repair the IoT nodes themselves. This can be achieved by establishing standardized secure communication framework which would include troubleshooting, recovery modes to perform attack detection and self-repair. This framework expands and adapts its services, functionality and security patching to cover layer specific failures and security risk but keeps the bare bone structure running in a virtual environment.

Looking into proposed solution we intend to work into extension of design and implementation of proposed framework in several areas of applications of IoT Networks

## ACKNOWLEDGMENT

This project is carried out in close collaboration with Prof. Dr. Joseph Kueng, Institute for Application Oriented Knowledge Processing, JKU, Linz. This research was funded by the FP7 - EU FRAMEWORK PROGRAMME under grant agreement No. 604659 (project CLAFIS).

## REFERENCES

- [1] <http://www.gartner.com/newsroom/id/2905717>
- [2] J. Allmendinger, G. Lombreglia, R.: Four Stages for the Age of the Smart Services (2012) <http://courses.ischool.berkeley.edu/i2901/f08/readings/StrategiesSmartServices.pdf>
- [3] IBM Market Insights, Cloud Computing Strategy Research, July 2009
- [4] Low Power Wide Area Network Technology : Symphony Link vs. LoRaWAN A Guide for Engineers and Decision Makers. White paper links lab 2016
- [5] A Comprehensive Look at Low Power, Wide Area Networks For 'Internet of Things' Engineers and Decision Makers
- [6] C. tian, X. Chen et.al : Analysis and Design of Security in Internet of Things, 8th International Conference on BioMedical Engineering and Informatics, 2015.
- [7] Advances in Cyber-Physical Systems Research :Jiafu Wan, Hehua Yan, Hui Suo and Fang Li
- [8] The research of several key question of Internet of Things: Zihua Hu
- [9] Internet of Things – New security and privacy challenges: Rolf H. Weber - University of Zurich, Zurich, Switzerland, and University of Hong Kong, Hong Kong
- [10] H Suo, J. Wan Security in the internet of things: Review, 2012 International Conference on computer Science and Electronics Engineering
- [11] John A. Stankovic, Life Fellow, IEEE "Research Directions for the Internet of Things"
- [12] A comprehensive look at Low Power, Wide Area Networks For 'Internet of Things' Engineers and Decision Makers
- [13] G.Gang, LU Zeyong: Internet of Things Security Analysis
- [14] [https://www.owasp.org/index.php/Top\\_IoT\\_Vulnerabilities](https://www.owasp.org/index.php/Top_IoT_Vulnerabilities)
- [15] Security and Privacy Grand Challenges for the Internet of Things : Glenn A Fink, Dimitri V. Zarzhitsky, Thomas E Carroll and Ethan D Farquhar
- [16] The Information Security for the Application of IoT Technology : Jia Jiang and Donghai Yang
- [17] LoRa vs LTE-M vs Sigfox - <http://www.nickhunn.com/lora-vs-lte-m-vs-sigfox/>

# Analysis of Deployment Options to Enhance Horizontal Information Sharing and Networking in Internet of Things

Andrej Mihailovic<sup>1,2</sup>, Marko Simeunovic<sup>2</sup>, Milica Pejanovic-Djurisic<sup>2</sup>

<sup>1</sup>King's College London, London, UK

<sup>2</sup>Research Centre of ICT, University of Montenegro, Podgorica, Montenegro  
andrej.mihailovic@kcl.ac.uk, marko.simeunovic@ac.me, milica@ac.me

**Abstract**—Proliferations and uses of Internet of Things (IoT) services are greatly underlined by purpose-based and autonomous compositions of IoT eco-systems. This trend is commonly encountered in large scale IoT environments such as Smart Cities. Consequently, expansion of IoT services affirms a specific property of their compositions, that is, verticality. This means shielded information flows for control, management and data and results in web-level visibility of IoT data. Following the recent initiatives towards horizontality-enabling solutions for IoT, our work revisits some basic assumptions encountered when an IoT system is built from scratch. These are relevant when considering various deployment possibilities, applications, use of equipment, data provisioning/sharing and facilitation of networking tools for specifics of IoT communications. We give an analysis of two possible deployment paths for enhancing the horizontal information sharing and networking. One is based on the default view of the integration of stand-alone devices and small scale IoT networks into the IP networking suite with scalable adaptations of discovery options such as DNS and APIs. The other one takes a radically different approach based on novel concepts of Information Centric Networking/Named-Data Networking and adapts them to IoT specifics. It applies it as a long term research direction for horizontal data distributions and IoT search-and-discovery models in future large scale deployment environments such as Smart Cities.

**Keywords**—*Internet of Things, Future Internet, Smart Cities, Information Centric Networking, Named-Data Networking*

## I. INTRODUCTION

Proliferation of Internet of Things (IoT) devices, services and technologies has undoubtedly been one of the most impacting novelty and expansion areas of telecommunications and accompanying technology. The term IoT has been applied as a universal meaning to a plethora of technicalities, devices and systems with wide ranging compositions and uses. In terms of the compositions, IoT refer both to stand-alone or small-scale installations of communicating device(s) as well as large scale installations of many system components in grand deployment environments such as Smart Cities. IoT eco-systems in Smart Cities effectively become operational as IoT services based on the operational integration of: devices, cloud-like databases, tools for data processing, control, management and final delivery at the application/web-level. Observing the properties and technicalities of the actual devices draws attention to their varying characteristics and uses, i.e. IoT devices refer to

simple low-processing sensor nodes or actuators (or clusters), to vehicles, then, unconstrained machines capable of smart operations and PC-like processing [1] etc. IoT deployment trends also show [2][3] great diversity of application areas: from small scale eHealth, smart homes, to automation of production processes, then, agriculture applications, smart cities, transport applications etc.

Many of the explanations of this paper are influenced by the diversity of issues related to IoT. Progress in the areas of IoT shows many, both novel and ongoing, initiatives for more organised and standardised approaches to many dimensions of IoT deployment: from PHY/MAC to application layer [3], integration into new standards for cellular technologies (i.e. 5G [4]), to achieving coherence of data representation tools between databases/clouds of different IoT eco-system providers [7][8]. It can be postulated that there is an ongoing global effort towards aligning and consolidating the understanding of diversity of IoT and their impact on the existing systems. In this paper, we focus on two angles: i) the particular technical space of networking and data provisioning tools; ii) large scale coexisting implementations of independent IoT devices or groups of IoT eco-systems. These are relevant to both the support for IoT in the Internet, eco-system architecture compositions and take their particular significance in dense large scale urban deployment areas called Smart Cities [9][10].

To accompany our study we observe a chronological and consequential process of forming many relevant IoT implementations. Many IoT-termed implementations of today have occurred as a transformation of Wireless Sensor (and actuator) Networks (WSNs) [1][2][5] forming IoT eco-systems typically bounded by functional and commercial structures (e.g. a Smart City company installations). A technical formulation can be stated from an engineering stance as a crosscheck and reflecting on the current IoT implementations: IoT was idealistically introduced to signify IP-reachability of various devices as “tiny” IP hosts being facilitated by large volume of IPv6 addresses. This approach has its discrepancy when reflected against the current IoT eco-systems structures where achieving independent IP network (and data) visibility for every IoT device is not the practical requirement. Rather, the IoT eco-systems provide visibility in an integrated and processed form at the application level.

II. FORMULATION OF IOT SPECIFICS AND CHALLENGES

A. Issues with Vertical Composition of IoT eco-systems

An IoT system is described as vertical to denote the property of system components organisation from devices/“things” on the ground to presentation of data at the service/application level [29]. Verticality is applied in a generic sense highlighting the flow directions of control/data and setup of architecture elements that support the deployment of IoT (e.g. for data collection purposes)<sup>1</sup>. Figure 1 sketches a simplified layout of IoT eco-system architecture components. The chosen IoT eco-system setup shows a large scale deployment of IoTs in Smart City environments, consisting of installations of “things” at the physical (ground) level, their communications via access infrastructure (e.g. cellular networks, WiFi hot-spots etc.), then, via Internet communications to the installations of a vertical IoT service provider with data processing facilities (e.g. a cloud infrastructure) and final delivery of digested data in the forms of services and applications towards the users. This setup assumes digestion of situations on the physical/ground levels, exposed via data collections, by processing data and presentation at the application level. This summarises the vertical composition of an IoT eco-system.

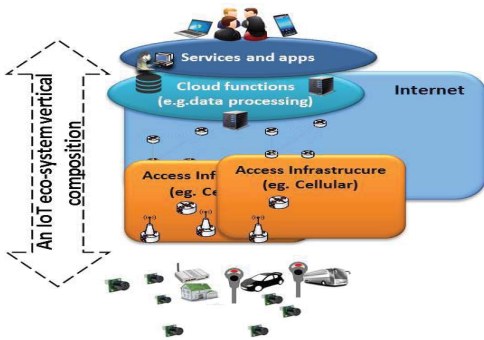


Figure 1. An example generic setup of a vertical IoT eco-system

Existence of many separate IoT eco-systems in Smart Cities of today induces a challenge of enabling their interoperability in exchanges and correlations of data at the data processing and/or cloud level (e.g. using standardized semantics or ontological data structures [7][8][11]). Aligning complex data structures between different eco-systems is a burdensome task at the system-level both in terms of the logic applied to link the data and lack of openness for complete interoperability. In addition, some properties of verticality would still apply (visibility of processed data at the application level) even if ideal aligning between the data models is achieved.

<sup>1</sup> European Commission project call ICT 30 in 2015 “The biggest challenge will be to overcome the fragmentation of vertically-oriented closed systems, architectures and application areas and move towards open systems and platforms that support multiple applications.”

B. Challenges for Horizontal Information Sharing and Networking

We apply the concept of IoT horizontality as existences of deployment situations and supporting functionality that allow processes of IoT discovery and service provisioning in manners not bound to vertical installations of IoT eco-systems. Hence, the IoT service is not to be discovered and utilised only using specific web-level presentation of IoT data but as inherent and integrated features of the future networks. We note that this might already be a work-in-progress whereas we discuss some open and needed changes in the Internet functionalities and use of novel networking paradigms. Some specific challenges and observations are listed below, highlighting the need for IoT horizontality (more is given in [29]):

- An underlying consensus of the global IoT initiatives and visions confirms IP reachability and connectivity as the defining property of IoT communications, i.e. facilitated by IPv6 addresses. It is projected that there will be billions of IoT devices by the end of this decade [6] by unleashing the IPv6’s addresses. Many of the deployed devices would be independent installations that are not bound to commercial and functional compositions of an IoT eco-system provider. As IETF provides the tools for hosting IP protocol on IoT “tiny” devices [12][13][14], the challenge is how to accommodate large population of IP hosts regarding the Internet functions such as DNS, APIs, topological configuration of IP address, Service Provider logic etc.
- Many of the challenges surface when IoT are considered from the commercial deployment perspective specifically in deployment areas such as Smart Cities<sup>2</sup>. Communication paradigm for IoT differs from the conventional Internet, which is built on “human-to-human” or “human-to-server” communication model [13][23]. IoT communications are combinations of “device-to-device/human/database/automation...” models and vice versa.
- IoT deployment is marked by the specific action of device configuration. This is a process of software and hardware setup that ensures operation and communication of IoT devices. This action requires dedicated engineer skills and has greatly been the reason for the nature and properties of current IoT systems (i.e. verticality and evolution of early sensor network installations). It is a significant constraint to global and liberalised deployment of IoT devices as many users often do not possess sufficient skills to configure and setup IoT devices. Hence, horizontal enablers are needed to ease the deployment process. Commercial forecasts accordingly highlight the device management as the risk associated with deployment of IoT devices at large scale<sup>3</sup>.
- Another facilitator of horizontality can be looked for in the vision of Named-Data Networking (NDN) [15][16][17] for IoT that stems from the novel ideas and proposals of

<sup>2</sup> <http://www.iottechnews.com/news/2016/apr/11/why-network-cornerstone-smart-cities/>

<sup>3</sup> <http://www.iottechnews.com/news/2016/apr/07/why-device-management-presents-risk-iot-adoption-scale/>

Information Centric Networking [18][19]. The vision is founded on the principle that IoT communications are typically concerned with data provisioning of small chunks of data (several bytes) that are named as the types of sensory or actuation purposes of IoT devices (e.g. temperature, pollution, meter readings, then, light/equipment on/off statuses). NDN is based on address-less routing, where the search packet specifies the Interest and upon Interest/data discovery in its upstream path, returns the content to the seeker host. Such a paradigm accordingly fits with the IoT communications and is currently applied to cases searching data in buildings, vehicles [20][21][22] etc. We extend the vision with a proposal for more comprehensive search and discovery data items that would not only identify the value of the sensory or actuation equipment, but provide the geographical, network-based, ownership-based search and discovery of IoT devices and services in large scale deployment areas such as Smart Cities.

The following sections discuss two development paths for horizontal deployment of IoT devices. The first one discusses support of the standard Internet suite for accommodating large scale IoT populations. The second one summarises a long term vision [29] of extending NDN concepts to comprehensive discovery of IoT devices and services in large scale urban deployments.

III. FITTING IOT DEVICE POPULATIONS IN THE INTERNET

A. Example IoT Architecture Composition

Independent IoT devices are yet to be seamlessly integrated in the Internet protocol suite. Barriers are in the accommodation of large scale populations of IoT devices in the Internet as well as in the availability of practical skills for setting up of hardware and software. Our university setup strategy for IoT deployments and accompanying functionality is depicted in Figure 2 [23][24]. The development is carried as a test-bed platform termed as a multipurpose and generic IoT system. Its multipurpose is ascribed to the property that it can be deployed for various application scenarios from pure networking research to applications such as Smart Cities and Smart Agriculture. In addition, generic properties are associated with functionalities of devices that have functionality that can be adapted, usually with network administrator interventions, to many different purposes and configurations.

The protocol composition of the IoT network is given in Figure 3 [24]. In essence, the architecture consist of two types of IoT devices (*processing* and *gathering* devices with either sensory and actuation components) and two types of IoT Gateways (with standard and localised processing capabilities). Choices behind the features are explained in [24]. The network is setup to function using a model of data provisioning in two ways:

**1. Data provisioning using a centralised data storage and application delivery:** Data is gathered in a centralised platform and delivered in a manner suitable to application requirements. Ranges of applications are diverse subject to

data processing and extractions from the centralised platform. Such a model follows the vertical setups of IoT eco-systems.

**2. Deployment of IoT Devices as independent “tiny” IP hosts:** Provided by the features of CoAP [14] and CoAP/HTTP proxying [25], data collections from IoT devices are possible “directly” via standard Internet communications. The process can involve direct communications using an IP address of an IoT device (in case it is stand-alone) or via an IoT Gateway that connects its local cluster of IoT devices to the Internet.

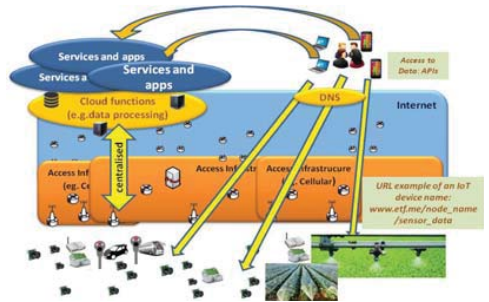


Figure 2: Basic configuration of a multipurpose and generic IoT device deployment strategy

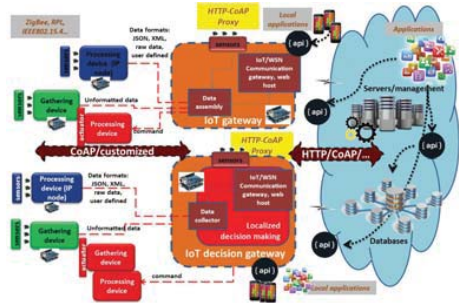


Figure 3: Composition of protocols for multipurpose and generic IoT networks [24]

B. Discussion on the Deployment Requirements

The architecture provides some options that are driven by considerations of uses, applications and commercialisation. These options can be discussed as general considerations for IoT deployments in the Internet:

- (*Horizontal*) IoT Service Provider: We note a necessity for formulating IoT Service Provider that allows its customers a generic and multipurpose service of IoT deployments, both using a centralised data platform and using independent IoT device/cluster deployment. The former is common nowadays, while the latter (i.e. Horizontal) is deemed needed due to skills required to setup and connect IoT devices as IP hosts, often not possible by customers themselves (this option is the focus below). This step largely differs from setting up of a

conventional IP host in an office or home. In addition, such a (Horizontal) IoT Service Provider might not be able to associate the connectivity of its IoT devices with the physical connectivity to its network (i.e. IoT Service Provider need not own an access network). This differs from the configuration of existing ISP providers for regular Internet hosts/customers. Rather, IoT devices by nature of the deployments are often scattered and connected over various access networks and physical location (e.g. in diverse locations in cities)<sup>4</sup>.

- *IPv6 Addressing*: It is realistic to assume that locations of IoT devices are not to be bound to deterministic network locations (subject to customers preferences) and that such configuration might not be static, i.e. IoT can be mobile (e.g. vehicular, re-deployments etc). Following this, sole assignment of IPv6 addresses might not suffice to enable the connectivity of large populations of IoT devices to the Internet. It is often the case in the current networks that IP hosts receive a temporary IP address (or private IP address due to Network Address Translation). A question might be how to bind an IP address of an independent IoT device into DNS. In other words, assuming a standard API such as REST is applied to pinpoint data into an IoT device, i.e. DNS needs to resolve the IoT device's URL/URI to the exact network location/address.

- *IoT DNS Extensions*: Limitations of DNS for IoT Networking via Internet (i.e. TCP/IP architecture) are already noted in [26] considering DNS-based service discovery (DNS-SD) [27] and CoAP based Resource Discovery [28]. However, these discoveries are relevant when IP address of an IoT Gateway or device is known and where a resource can then be appropriately queried. In the model we advocate, it is assumed that customers usually know the resources they want to fetch or control, hence, the priority is in establishing a simplified means of reaching the device. In the horizontal deployment discussed in this paper, we focus on resolving large scale URL/URI-to-IP address mappings to locate the IoT device(s). We foresee some options for modifications of the standard DNS mechanisms and extension of APIs (e.g. REST API) to accommodate for the specifics of IoT deployments:

a) *Extensions to DNS's name-to-address resolving steps*: Devices' names and query can be contained in the URL/URI description of the IoT. This step supplements and/or precedes the mentioned CoAP Resource Discovery and its Link Formats and assumes that users/customers will ping their devices with pre-set names and known resources. Such an assumption is based on the existence of the IoT Service Provider. An example URL/URI structure can include an IoT Service Provider's domain name (e.g. <http://coap://etf.me>) followed by URI/URN fields such as path and query that can contain the device's name and preset syntax describing its resource or a more complex query logic.

b) *Scaling DNS by localised Registries*: Following the described model for (Horizontal) IoT Service Providers and the global projections for billions of IoT connected devices it is reasonable to expect that DNS needs to be scalable to accommodate for the explosion of "tiny" IP hosts. We

foresee a two-way DNS URL/URI resolving where the first address resolving can point to a localised IoT DNS/Registry that keeps bindings of current IP addresses of IoT Devices/Gateways and redirects the application layer request (e.g. using HTTP/CoAP redirection) to provide the current address of the IoT device (analogous to dynamic DNS or mobility management agents). It is assumed that this step will contain the necessary authentication and protection against malicious attacks (e.g. by naming or authentication steps) before returning the address locations of IoT devices.

#### IV. VISION OF EXTENDED NAMED-DATA NETWORKING FOR IOT

NDN paradigm fits with the nature and requirements of IoT communications. As much of the IoT communication is proactive and concerned with data delivery of small chunks of data, using a search driven NDN routing can accordingly fit with distribution of data from IoT devices. There are already numerous positioning papers and trial solutions that couple NDN concepts [18][19] and IoT deployment [15][16][19][20][21][22]. As NDN is about using "data interest" as the search and routing pointer, the IP discovery methods such as the conventional DNS are obsolete. Examples of IoT NDN implementations include setting up of data repositories in buildings where sensory/actuation data is obtained using the data search, then, vehicular network etc. The search usually reflects the nature of sensory/actuation hardware (i.e. the name of the data) and it is fairly simple (e.g. room temperature in a building etc).

We extend the thinking behind the NDN application for IoT with a vision of a broader inclusion of search items, that is, data names that describe IoT devices. This leads to the meta-data that would present the comprehensive description of IoT devices and include some of the following items for its discovery (more is given in [29]):

- **Physical location**: a search can include the actual physical location as the main search criteria, e.g. a city region such as a square, geographical coordinates/tags, streets, buildings etc. Currently, users refer to prior knowledge of the web address/apps of the IoT provider in the relevant area. Running a web-level search using a city's region as the IoT-search-item would be rather inefficient in finding the local devices and services.

- **Timely dimension**: values, statuses and device populations are highly dynamic.

- **Ownerships**: knowing the responsible company, public body or individuals that own the IoT devices or services and certify the data, would make the search and data fetching more meaningful. It would define the trust and data integrity framework.

- **Connectivity**: a defining item of the search can include network or access location, such a seeking specific (cellular) network operators or wireless access points. This option can also apply addresses (e.g. IP address or subnet).

- **Data Type, Name, Status and Query Logic**: running a generic search such as "weather conditions", a URI path/query segment, or applying a simple query logic (e.g. seek only weather values from a public provider/owner in a city area)

<sup>4</sup> Emerging Low Power Wide Areas IoT networks aim to offer direct access to IoT provider networks

would render a targeted and relevant resolving of the information and is already contained in NDN solutions for IoT. **The search for IoT devices and their data and services would be an opportunistic and near-match action.** The vision is targeted as a future research direction for large scale implementation of IoT devices such as in Smart Cities. As shown in Figure 4, locations of data repositories can be distributed at various locations in networks, from gateways, to elements in the access infrastructure to public/private databases with web-level maps in cities.

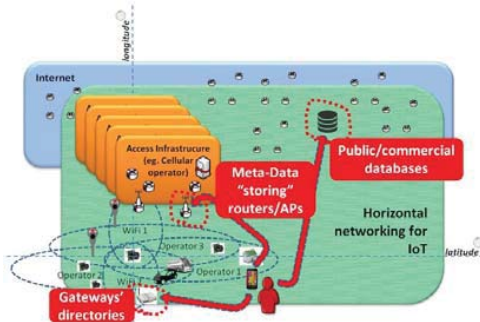


Figure 4: Examples of locations of IoT data for NDN-driven horizontal networking for IoT communications (from [29])

## V. CONCLUSIONS

This paper contains an analysis of deployment options for IoT termed as horizontal and discusses topics that diverge from the current typical configurations of IoT eco-system. The current eco-system generally fit into a structure that is considered as vertical, referring to the composition of the elements of the eco-system, from IoT devices to web-level applications. Such configurations are relevant to deployments of IoT in large scale urban environments, i.e. Smart Cities. We give two alternatives driven by the expectation that IoT will be deployed in a liberalised manner not bounded by provider services of today. The first alternative observes adaptations needed in the conventional Internet setups for supporting a large population of IoT devices. The other one considers extensions to NDN novel routing paradigms that can fit with the nature of IoT communications.

## Acknowledgment

Work is funded by European Union project Fore-Mont as a part of Seventh Framework Programme (Grant Agreement No. 315970 FP7-REGPOT-CT-2013) <http://www.foremont.ac.me>.

## References

[1] D. Giusto, A. Iera, G. Morabito, L. Atzori (Eds.), *The Internet of Things*, Springer, 2010.  
 [2] L. Atzori, A. Iera, G. Morabito, "The Internet of Things: A survey", *Computer Networks*, Vol. 54, Is. 15, pp. 2787-2805, October 2011.  
 [3] M.R. Palattella, et al., "Standardized Protocol Stack for the Internet of (Important) Things," *Communications Surveys & Tutorials, IEEE*, vol.15, no.3, pp.1389,1406, 3<sup>rd</sup> Quarter 2013.

[4] M.R. Palattella, et al., "Internet of Things in the 5G Era: Enablers, Architecture and Business Models", *IEEE Journal on Selected Areas In Communications*, 2016  
 [5] L. Mainetti, L. Patrono, A. Vilei, "Evolution of wireless sensor networks towards the Internet of Things: A survey", *IEEE 19th SoftCOM conference, Croatia, Sep. 2011*  
 [6] "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020", *Gartner*, 2013-12-12, <http://www.gartner.com/newsroom/id/2636073>, Retrieved 2014-01-02.  
 [7] M. Compton, et al., "The SSN ontology of the W3C semantic sensor network incubator group", *Web Semantics: Science, Services and Agents on the World Wide Web*, Volume 17, December 2012, pp. 25-32  
 [8] P. Barnaghi, M. Presser, K. Moessner, "Publishing linked sensor data", *3rd International Workshop on Semantic Sensor Networks (SSN), ISWC2010, Shanghai, China, Nov. 2010*.  
 [9] A. Zanella et al., "Internet of Things for Smart Cities", *IEEE Internet of Things Journal*, Vol.1 No.1, Feb. 2014.  
 [10] "IEEE Smart Cities" *IEEE, Web. 06 Sept 2015 smartcities.ieee.org*  
 [11] A. Sheth, C. Henson and S. S. Sahoo, "Semantic Sensor Web," in *IEEE Internet Computing*, vol. 12, no. 4, pp. 78-83, July-Aug. 2008.  
 [12] G. Mulligan, "The 6LoWPAN architecture", *4th Workshop on Embedded networked sensors, EmNets, ACM, Jun. 2007, Cork, Ireland*.  
 [13] T. Winter et al., "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks", *RFC 6550, IETF RFC 6550, Mar. 2012*.  
 [14] C. Bormann, A. P. Castellani, and Z. Shelby, "CoAP: An Application Protocol for Billions of Tiny Internet Nodes," *IEEE Internet Computing*, vol. 16, no. 2, pp. 62-67, 2012.  
 [15] W. Shang et al., "Named Data Networking of Things", *1st IEEE International Conference on Internet-of-Things Design and Implementation, Apr. 4-8, Berlin, Germany 2016*.  
 [16] M. Amadeo, C. Campolo, A. Iera and A. Molinaro, "Named data networking for IoT: An architectural perspective," *European Conference on Networks and Communications (EuCNC)*, Bologna, 2014, pp. 1-5.  
 [17] E. Baccelli, C. Mehlis, O. Hamm, T. C. Schmidt, and M. Wählisch, "Information centric networking in the IoT: experiments with NDN in the wild", *1st international conference on Information-centric networking (ICN '14) ACM, 2014 New York, USA*, pp. 77-86  
 [18] L. Zhang et al., "Named Data Networking", *ACM SIGCOMM (CCR)*, Jul. 2014.  
 [19] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggs, R. Braynard, "Networking named content", In *Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT 09)*, ACM, New York, USA, 2009.  
 [20] W. Shang, Q. Ding, A. Marianantoni, J. Burke, L. Zhang, "Securing Building Management Systems Using Named Data Networking", *IEEE Network*, vol. 28, no. 3, pp. 50-56, May/June 2014.  
 [21] J. Burke, P. Gasti, N. Nathan, G. Tsudik, "Securing Instrumented Environments over Content-Centric Networking: the Case of Lighting Control", *IEEE INFOCOMM 2013, NOMEN Workshop, Turin, Italy, Apr. 2013*.  
 [22] G. Grassi et al, VANET via Named Data Networking", *IEEE INFOCOM NOM Workshop, Toronto, Canada, April-May 2014*.  
 [23] A. Mihailovic, M. Simeunovic, N. Lekic, M. Pejanovic-Djurišić, "A strategy for deploying diverse sensor-based networks as an evolution towards integrated Internet of Things and Future Internet," *22<sup>nd</sup> TELFOR, Nov. 2014, Belgrade, Serbia*  
 [24] M. Simeunović, A. Mihailović, and M. Pejanović-Djurišić, "Setting up a multi-purpose internet of things system," *23<sup>rd</sup> TELFOR Nov. 2015, Belgrade, Serbia*  
 [25] A. Castellani, S. Loreto, A. Rahman, T. Fossati, E. Dijk, "Guidelines for HTTP-CoAP Mapping Implementations", *Internet Engineering Task Force (IETF) Interest-draft, draft-ietf-core-http-mapping-04, Jan. 2015*.  
 [26] W. Shang, Y. Yu, R. Droms, L. Zhang, "Challenges in IoT Networking via TCP/IP Architecture", *NDN Technical Report NDN-0038, Feb. 2016*  
 [27] S. Cheshire and M. Krochmal, "DNS-Based Service Discovery", *RFC 6763 (Proposed Standard)*, Feb. 2013.  
 [28] Z. Shelby, M. Koster, C. Bormann, and P. van der Stok, "CoRE Resource Directory", *draft-ietf-core-resource-directory-05, Oct. 2015*.  
 [29] A. Mihailovic, "Liberalising Deployment of Internet of Things Devices and Services in Large Scale Environments". *Springer Journal Wireless Personal Communications*, Nov. 2016.



# Cybersecurity Challenges for Public Authorities

## A Danish View

Samant Khajuria

Center for Communication, Media and Information  
Technologies, CMI, Electronics Systems  
Aalborg University, Copenhagen,  
skh@cmi.aau.dk

Knud Erik Skouby & Lene Sørensen

Center for Communication, Media and Information  
Technologies, CMI, Electronics Systems  
Aalborg University, Copenhagen  
skouby@cmi.aau.dk; ls@cmi.aau.dk

**Abstract**—The paper discusses aspects of cybersecurity and cyberattacks specific to public authorities and institutions. The European directives for prevention and protection on the area are introduced. The focus is on Danish examples illustrating specific challenges for public bodies and the Danish regulation trying to meet these. From the discussion some recommendations for prevention are derived.

**Keywords**— *Cybersecurity, cyberattacks, public authorities, EU regulation.*

### I. INTRODUCTION

Within the last years' cybersecurity has become a well-known concept also to most governments and public authorities. Increasing cyberattacks of various forms on governmental and public authorities have set focus on this. In 2007, the Estonian Government and parliament (as well as banks and political parties) were subject to cyberattacks from malware or spyware programs [1]. In 2015, two Danish municipalities were attacked also by a ransomware attack that meant that the municipalities got their data encrypted and they were asked for a ransom [2]. Neither of these examples are special, many more are to be found; in the UK, e.g. it is estimated that 40% of malware attacks are against public authorities [3].

The cybersecurity threat on public authorities is serious. Public authorities usually receive, store and manage a large number of data critical to the persons represented in the data as well as to the state and governments. However, in managing these, the public authorities face a number of challenges such as insufficient funds, lack of control, regulation and know-how [4].

With the EU Network and Information Security Directive [5], where EU member countries are obligated to secure a common level in net- and information security for all countries, there is placed a certain pressure on member countries to put IT security on top of political agendas. Most European countries have, as a consequence, formulated National Cybersecurity Plans.

The new General Data Protection Regulation [6] places any institution in the EU including Public Authorities under strict regulation to protect personal data from misuse incl. to prevent cyber-attacks. The Directive entered into force on 5 May 2016

and EU Member States have to transpose it into their national law by 6 May 2018, so the specific effect remains to be seen.

The Scandinavian countries are in a high risk of cybersecurity attacks due to the high level of IT in public services and in the use of IT in communication between citizens and public authorities. A security company, FireEye, concludes in a report [7] there are increasing opportunistic cybersecurity attacks across the Nordic Region. They have estimated that the Nordic countries have a rather high share of advanced persistent threats (APTs) and malware alerts observed in Norway (47%); Denmark (36%); Sweden (14%); and Finland (3%). As one of the most digitalized nations in the world, threats are considered immensely high against the Danish authorities and companies as it is stated in a threat assessment made by Centre for Cybersecurity [8].

The purpose of this paper is to discuss and identify recommendations that can support public authorities and governments in search for better solutions against cybersecurity attacks. The paper is, methodologically, done as a desktop study of national strategies within cybersecurity and papers providing an insight into the IT infrastructure landscape of the future. In combining these elements, recommendations are found to address the cybersecurity challenges in public authorities. The paper builds on a case study of the Danish National Cybersecurity Strategy [9]. The Danish Strategy is in focus because it is relatively new and will have an update by the end of 2016.

The paper is organized as follows: Section 2 presents the general set of cybersecurity challenges. Section 3 includes a literature overview of similar studies. In section 4, challenges for public authorities are discussed with an example from Denmark. Section 5 looks at the learnings from the threats facing public authorities and the Danish National cybersecurity strategy as a basis for identifying suggestions on how to deal with the challenges in a more holistic format. Finally, conclusions are presented in section 6.

### II. GENERAL CHALLENGES

The general challenges in cybersecurity for public authorities consist of a variety of targets, tools and actors.

The International Standards Organization (ISO) defines a threat as: "A potential event. When a threat turns into an actual

event, it may cause an unwanted incident. It is unwanted because the incident may harm an organization or system” [10]. In itself, this definition does not say something about the threat tools or provide more detail into which this threat will harm (organizations or system).

The RAND Europe Study Team [11] has identified four basic targets that a perpetrator would have an interest in influencing. These are:

- Individuals who can be threatened on privacy, freedom of speech, access to services and physical security
- Organizations that can be threatened on their products and services, their production line (including money and patents), reputation and trust levels
- The supply chains that can be threatened on the responsibilities for information on individual, civilian persons or clients, on control of general facilities and systems and their interdependence between organizations
- Societal can be threatened on the availability of essential services, protection of the democratic legal order and national security, the infrastructure of the Internet, free movement of services and digital security.

This definition does go across the more traditional definition of targets which include individuals (consumers and citizens) and governments and corporations [11]. This paper focuses on public authorities and therefore lies across the three of the four categories mentioned above – the organizations, supply chain and societies categories.

Cybersecurity threats come from a variety of cybersecurity actors. The actors can be a variation of ([12], [13]):

- Nation states with sophisticated cyber programs using these for espionage and cyberwars (where computer technology is used for disruption of activities of state of organization by attacking the information system for strategic or military purposes
- Nation states with less sophisticated cyber programs with the purpose of disrupting
- Profit-driven criminals such as organized cybercrime where targeted attacks are placed on special targets either via a special tool or via traditional financial cybercrimes via the Internet
- Ideologically motivated hackers or extremists.

On top of the identified targets, the cybersecurity threats are defined also via the threat tools (malware and its variants) and threat types (such as unauthorized access etc. – see more below) [11]. In this paper, we operate with the following cybersecurity threat tools

- Malware which essentially is malicious software which can be rather similar to the software code it inflicts so it in that way can be difficult to find. This is often used by profit driven cybercriminals

- Trojans which can harvest login-in credentials on banking information and by that misuse the information for own profit
- Ransomware infects computer systems and take over control so that the normal user of the system can be denied access before a ransom is paid
- Point of sale (PoS) malware can scan, copy and send credit card information through check-out systems using the shops. Receivers of the information (often referred to as a command and control center) can hereby track name and pin code for the credit card and in that way misuse the credit card
- Botnets are remotely controllable malwares that are installed without the knowledge of the user, and provide unauthorized access to systems and information
- Exploits which are data of series of commands that can exploit a software or hardware in terms of its vulnerabilities with the purpose of creating undesired behavior. These are used also by security personnel and are therefore not only threats.

The above mentioned tools are often divided into categories explaining their threat type such as: Unauthorized access, destruction, disclosure, modification of information and denial of service.

### III. CYBERSECURITY AND PUBLIC AUTHORITIES

Searching for related research on cybersecurity and public authorities, some documents can be found addressing public authorities (or local authorities) and their web services. Deaking and Dinnon [14] look at e-government for local authorities in New Zealand, and Choudrie et al [15] who carries out a similar study in the UK. Both of these studies look at the single web-service as an entry to assessing the cybersecurity. Other papers (such as for example [16]) suggests a Smart Grid Infrastructure as a key element in dealing with Cybersecurity.

A search addressing Governmental and political related issues gives many results. For example the “Cyber Security Strategy for Norway” [17], where there it is mentioned that local governments shall establish information security management systems as all state-owned companies must do. There is here a tight focus on the physical aspect of a cybersecurity infrastructure linking local authorities within this infrastructure. Another examples is the “UK Security and the UK’s Critical National Infrastructure” [18]. In this document there is a focus on building up a cybersecurity culture providing the backbone of managing the security threats facing UK institutions and Government. Furthermore, Lipman [19] suggest that local Government organizations should take a cloud-based approach to enable security policies.

In Denmark, the “National Strategy for Cyber- and Information Security [9] acknowledge a strong and coherent cybersecurity strategy but they do not mention public authorities in their plan.

IV. CHALLENGES FOR PUBLIC AUTHORITIES – A DANISH EXAMPLE

A public authority will in this paper be an agency or institution with primary functions to produce non-marked related services [20]. The authority is usually paid by taxes or by subsidies from the government, they are not necessarily integrated but often have their own accounts. Citizens can be services by the public authorities free of charge (at least in Denmark).

One example of a public authority is the unemployment service called Job Center, in Denmark. The Job Center supports unemployed citizens in finding jobs, counselling, making job-plans etc. [21]. The Job Center is situated in the local Danish municipalities and answers to the Government institution Danish Agency for Labor Market and Recruitment. In Figure 1, the Job Center is shown in relation to the IT logistics and its related partners through this IT infrastructure. It shall be mentioned that the figure is a simplified version of reality. The figure has been verified by a person working in the local Job center in Roskilde, Denmark.

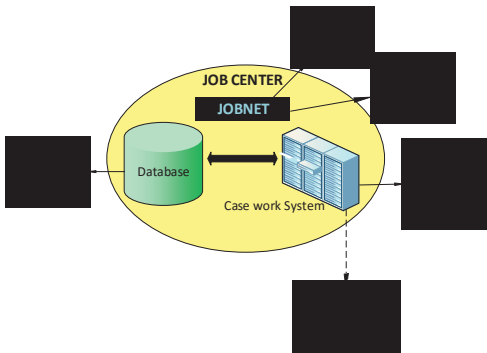


Figure 1 A simplified view of a Danish Job Center and its IT connections to other entities outside the Job Center

Citizens and employers have access to the IT system Job Net where jobs can be announced and found and the citizens must document that they are active in searching jobs. Job banks are both private and public banks that feed into the Job Net system. The Case Work system includes case files of the unemployed citizens with plans for how to get a job or improve chances for a job. The private actors linked to this system are mainly providing services relating to the unemployed citizens – such as courses etc. As mentioned the Danish Agency for Labor Market and Recruitment uses data from the database in the Job Center for further analysis and for political documents. There are other IT systems in a Danish Job Center—for example economy systems—but the above-mentioned provides a short insight into parts of the IT infrastructure for a local public authority.

In relation to the cyber security challenges in public authorities, the Centre for Cyber Security [8] has assessed the threat against these to be very high. They point towards hackers and cyber espionage as main threats and say that the attackers all the time continuously are perfecting their technical skills

and capabilities which must force the public authorities to raise security levels.

The Data Protection Authority in Denmark has identified three categories relating to security flaws and errors they have found relating to the internal environment in public authorities. These are [22]:

- Organizational errors such as missing security protocols, differences in agreements between employees working with data and those at the top level, missing control with security measures ordered by the IT service providers, lack of use of anonymous user IDs
- Missing or flawed IT solutions relating to lack of use of encryption when confidential or sensitive data is sent through Internet, lack of control so employees accidentally can enter areas of the IT system not intended for him/her, no logging of information on what the employee has been doing with the systems
- Human errors such as making sensitive or confidential information public online due to lack of knowledge or skills in terms of how to handle data or the IT system; errors in connecting with sending e-mails – to wrong addresses and un-encrypted; mix up of documents in different cases when collecting the case via post and digital post.

However, as already mentioned, the above categories all relate to the intra-organizational environment and how employees and the organization generally are not structurally geared for dealing with security elements. Another perspective of these internal security threats is the general lack of knowledge about the threats and how to deal with them and the lack of resources to set-up security measures. Security has become a complex area to deal with: there are no clear cut cybersecurity solutions that can be installed, the public systems often require many different products from multiple vendors and there is the non-trivial problem of integration with existing public IT systems. Often this is so expensive and technically heavy that there is no money or other resources available to these kinds of challenges [23].

The threats mentioned all relate to public authorities. However, it also seems that the threat actors have a preference for use of ransomware tools towards public authorities. There are a number of cases in which for example Danish municipalities have been hit by ransomware and have been forced to pay to get access to data files with private data [24].

Another aspect related to public authorities' vulnerability is their connection with other private, small and bigger private companies (for sub-contracting, e-commerce and the like) as well as other public authorities –each public authority is part of a logistics supply chain in a larger infrastructure [24]. It is well known that a hacker only needs access to one entity in a chain to also gain access to the others.

Worldwide, nations are trying to comply with the cybersecurity threats through creating national cyber security strategies. Denmark also has a cyber security strategy from 2014.

## V. MEETING THE CHALLENGE

First and foremost, it is a bit unclear what a cybersecurity strategy should be able to do. In the Directorate General [25], it is discussed that in spite of the lack of territoriality and borders in cyberspace nations usually look at the cybersecurity as a part of their national security agendas. National strategies often define cybersecurity issues across national context, private industry and civil societies and with a lack of the international element in this [25]. There is no clear definition of responsibilities with the different stakeholders in for example EU institutions, national governmental bodies and the private sector.

Within the last years, the EU has taken some initiatives to place cybersecurity on the agenda. These initiatives are [25]:

- EU – NATO collaboration about cybersecurity: the Computer Emergency Response Team – EU (CERT-EU) started collaborating with NATO Computer Incident Response Capability (NCIRC) in 2010 but has just recently (February 2016) signed an agreement on information sharing on best practices in cyber incident prevention, detection and response
- “The Cybersecurity Strategy of the European Union – An Open, Safe and Secure Cyberspace (February 2013) was EU’s first attempt to put together a policy document related to cybersecurity issues. EU prioritized several policy areas for the EU’s international cyberspace and gave priority for developing cyber defense policies
- The “NIS Directive” (Network and Information Security) was adapted to the EU in March 2014. This was an effort to harmonize the creation of standards and levels of cybersecurity across the EU. As part of this each country should create a Computer Emergency Response Team (CERT) that could provide cooperation, and information exchange between members’ states and the Commission. The NIS directive has just come into force (August, 2016).

As a result of the increasing cyberattacks and the focus on establishing strategies also Denmark has published a national cybersecurity strategy.

### A. *The National Strategy for Cyber- and Information Security*

In 2014, Denmark formulated the National Strategy for Cyber- and Information Security [9]. The strategy was made by the Government representatives only with no interest groups involved (Foley, 2015). The strategy operates with 27 initiatives distributed over 6 different focus areas. The focus areas are [9]:

- Professionalization and a stronger IT inspection
- Clear demands for IT service providers
- Stronger cybersecurity and knowledge within the area
- A robust infra structure within the energy sector and the telecom structure

- Denmark as a strong international partner
- Strong research and clear information to citizens, companies and authorities

The initiatives mean amongst others that persons with responsibility for IT within the ministries must make sure to use the International IT security standard ISO27001 [26]. The use of the ISO standard is considered to take place within the Governmental institutions and with resources exchanged within the ministries themselves [26]. The so-called National Center Cyber Crime Center (which lies under the police) must be build out and they must establish a unit to assess cyber threats and incidents. The police must open a platform where citizens can report on IT criminal incidents online. And there must be made an information campaign for citizens and companies to strengthen these groups’ knowledge about data protection. This must be done by the Data Protection Agency.

The strategy is mentioning these initiatives and their implementation by the end of 2016. After this period, there must be made an attempt to involve other partners (the private industry and research) in making an update of the strategy.

Local authorities are not mentioned within the strategy. The strategy is made without any extra resources given for implementation of the initiatives. At the same time, the Military Intelligence was given 465 mill Danish kroner to develop defense and attack weapons on different cyber threats. Details on the initiatives can be read in the strategy document [9].

## VI. RECOMMENDATIONS

This paper focuses on a single public authority in Denmark and can as such only be viewed as a case study. However, the points addressed here can be viewed in a larger perspective as part of the Danish Strategy for Cyber Security and as a representative of a part of this strategy which is not dealt with.

As mentioned above, the Danish Strategy for Cyber Security does not address the local authorities as part of the public infrastructure. As discussed above attacks on one element in an IT data supply chain can create access to the whole chain and a complex connected to the element (cf. Figure 1). This implies that complex systems by their nature have a high cybersecurity risk. This also specifically implies that the cybersecurity of the cooperating and communicating public sector is dependent on each of the many elements in the increasingly connected structure. This is a result of the development where IT is seen both as mean to achieve rationalizations and budget savings and as a goal to achieve better services. The first part is, e.g., seen in an announcement in October 2016 that all Danish state institutions will be forced to join a single, unified IT-system [27]. Along with this there has been a strong drive to secure the state-network including an advanced firewall with pre- and post-surveillance of traffic.

The drive to achieve better services through the use of IT is in many ways challenging the cybersecurity. It involves connection and communication with citizens giving them access to the network – an issue that increasingly is seen as demanding a more sophisticated solution than the existing

(NEM-ID) – and with local authorities. A strategy focusing on the state sector, involves a risk for insufficiency. The local authorities are centrally placed, c.f. the discussion above and therefore a middle player and access point for cybercriminals. This is partly addressed in the Danish pre-GDPR, but there is a need for full alignment with the GDPR Directive.

The ideal cybersecurity strategy addresses the full infrastructure in a nation. This does not necessarily mean one unified system. However, when all the state institutions are included in one IT-system, there has to be one cybersecurity strategy covering them. As argued above, the IT-systems of all public institutions are increasingly communicating, meaning that there has to be a cybersecurity system addressing this communication situation between the state and the local authorities. Finally, the communication between citizens and public institutions; this is partly addressed by the Danish NEM identification system – a system that needs updating.

VII. CONCLUSION

Cybersecurity is an area of increasing concern in the public sector – as in all other walks of life. In the public sector there are daily reports on fear of or actual hacking activity at all levels from internal, daily activities to top-level secret government negotiations which has, e.g., prompted the UK government to prohibit smartphones and smartwatches in cabinet meetings [28].

With the EU directive on GDPR [6] which focuses on protection of personal data to the Danish National Strategy for Cyber- and Information Security [9], cybersecurity is set on the agenda. However, the link between the Directive and the Nation al Strategy mentioned is not clear. Local Authorities fall between the document and there is a need to address these as a part of the national infrastructure mapping the entities and their linkages as discussed above nations such as Denmark will be one step closer to dealing with the cybersecurity threats.

As cybersecurity is clearly emerging as a global concern, the need for learning from different national experiences and for alignment with EU directives under development seems obvious. Currently, the individual nations are generally seeking their own solutions and precautions.

REFERENCES

[1] Pawlak, P. and Wendling, C. Trends in Cyberspace: Can Governments Keep Up? *Environ.Syst.Decis* (2013), 33, 536-543

[2] Olsen, T. (2015): It-Sikkerheden skal forbedres I Kommuner og Offentlige Institutioner.. <http://website.willis.dk/nyheder-pr/nyheder/it-sikkerheden-...>

[3] Gajli, Agnieszka (2016): Time to face up to cyber risk. *PublicFinance* 31 March 2016, <http://www.publicfinance.co.uk/opinion/2016/03/time-face-cyber-risk>

[4] Infosecurity magazine (2016): The Cybersecurity Challenges Facing State and Local Governments. <http://infosecurity-magazine.com/opinions/cybersecurity-challenges-state/>

[5] EU Network and Information Security (NIS) Directive (<https://ec.europa.eu/digital-single-market/en/news/network-and-information-security-nis-directive>)

[6] EU General Data Protection Regulation (2016): <http://eur-lex.europa.eu/legal->

content/EN/TEXT/?uri=uriserv:OJ.L\_.2016.119.01.0089.01.ENG&toc=OJ.L.2016:119:TOC

[7] FireEye (2015): Cyber Threats to the Nordic Region. <https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-nordic-threat-landscape.pdf>

[8] Centre for Cybersecurity (2016): The cyber threat against Denmark. <https://fe-ddis.dk/cfcs/CFCS/Documents/Threat%20Assessment%20-%20The%20cyber%20threat%20against%20Denmark.pdf>

[9] Regeringen (2014): National Strategi for cyber- og Informationssikkerhed. In Danish. (The Danish Government:Strategy for cyber and information security) [https://www.google.dk/search?client=safari&rls=en&q=informations-+og+cybersaerkerhet+i+Sverige&ie=UTF-8&oe=UTF-8&gws\\_rd=cr&ei=uVRyV5j0EMGfAGJyK2ADA#](https://www.google.dk/search?client=safari&rls=en&q=informations-+og+cybersaerkerhet+i+Sverige&ie=UTF-8&oe=UTF-8&gws_rd=cr&ei=uVRyV5j0EMGfAGJyK2ADA#)

[10] Praxiom Research Group Limited (2013): "Plain English" ISO IEC 27000 2014 Information Security Definitions.

[11] European Parliament (2015): Cybersecurity in the European Union and Beyond: Exploring the Threats and Policy Responses. Policy Department on Citizens' Rights and Constitutional Affairs. <http://www.europarl.europa.eu/supporting-analyses>

[12] Khajuria, S., Skouby, K.E. and Sørensen, L. Cybersecurity Challenges for SMEs – A European Perspective. Presented at Global Wireless Summit, 2016, Aarhus, Denmark.

[13] Clapper, J. (2015): Worldwide Threat Assessment of the US Intelligence Community – Statement for the Record. Senate Armed Services Committee, February 26. As of 12 October 2015: [http://www.armed-services.senate.gov/imo/media/doc/Clapper\\_02\\_26\\_15.pdf](http://www.armed-services.senate.gov/imo/media/doc/Clapper_02_26_15.pdf)

[14] Denkins, E. and dilling, S.M. E-Government in New Zealand: The Local Authority Perspective. *The International Journal of Public Sector Management*: 2002; 15, 4/6, pp. 375-398

[15] Choudre, J., Ghinea, G. and Songonuga, V.N. Silver Surfers, E-government and the Digital Divide: An Exploratory Study of UK Local Authority Websites and Older Citizens. Oxford University Press, 2013.

[16] Ericsson, G.N. Cyber Security and Power System Communication – Essential Parts of a Smart Grid Infrastructure. *IEEE Transactions on Pwer Delivery*, vol. 25, No 3, July 2010.

[17] Norwegian Ministries (2012): Cyber Security Strategy for Norway. [https://www.google.dk/search?client=safari&rls=en&q=informations-+og+cybersaerkerhet+i+Sverige&ie=UTF-8&oe=UTF-8&gws\\_rd=cr&ei=uVRyV5j0EMGfAGJyK2ADA#](https://www.google.dk/search?client=safari&rls=en&q=informations-+og+cybersaerkerhet+i+Sverige&ie=UTF-8&oe=UTF-8&gws_rd=cr&ei=uVRyV5j0EMGfAGJyK2ADA#)

[18] Cornish, P., Livingstone, D, Clemente, D. and Yorke, C. Cyber Security and The UK's Critical National Infrastructure. Chatham House Report, 2011.

[19] Lipman, P. Four Critical Cyber Security Issues Facing Local Government. <https://www.publictechnology.net/articles/opinion/four-...>

[20] Den Store Danske (2016): [http://denstoredanske.dk/Samfund\\_jura\\_og\\_politik/Samfund/Samfund\\_og\\_statistik/den\\_offentlige\\_sektor](http://denstoredanske.dk/Samfund_jura_og_politik/Samfund/Samfund_og_statistik/den_offentlige_sektor)

[21] <https://www.borger.dk/sider/jobcenter-og-jobnet.dk.aspx>

[22] Data Agency (2015): <https://www.wen.dk/sider/jobcenter-og-jobnet.dk.aspx>

[23] Tomassen (2016): <http://www.cothomasen.dk/Default.asp?locale=da>

[24] Effektivitet (2015): Pas På Hackerne Kommer. <http://scm.dk>

[25] Directorate-General for Internal Polities. Cybersecurity in the European Union and Beyond: Exploring the Threats and Policy Responses. European Parliament, 2015.

[26] Hansen, J. and Kjær, J.S. (2014): Regeringen hæbes: Ny National It-Sikkerhedsstrategi Mangler Penge og Ekspertise (In Danish). <http://politiken.dk/inland/ECE2486408/>

[27] Computerworld(2016),[http://www.computerworld.dk/art/238214/sparepl-an-til-kvart-milliard-kroner-alle-ministerier-tvinges-ind-i-staten-it-fremmod2021?utm\\_source=newsletter&utm\\_medium=newsletter&utm\\_campaign=dispatch+15953](http://www.computerworld.dk/art/238214/sparepl-an-til-kvart-milliard-kroner-alle-ministerier-tvinges-ind-i-staten-it-fremmod2021?utm_source=newsletter&utm_medium=newsletter&utm_campaign=dispatch+15953)

[28] Spamfighter(2016)<http://www.spamfighter.com/News-20541-UK-Cabinet-Ministers-Not-Allowed-Smart-Watches-during-Meetings.html>

# Cybersecurity Challenges for SMEs

- A European Perspective

Samant Khajuria

Center for Communication Media and Information  
Technologies (CMI), Electronics Systems  
Aalborg University, Copenhagen, Denmark  
skh@cmi.aau.dk

Knud Erik Skouby, Lene T. Sørensen

Center for Communication Media and Information  
Technologies (CMI), Electronics Systems  
Aalborg University, Copenhagen, Denmark  
{Skouby, ls}@cmi.aau.dk

**Abstract**—Cybersecurity is a complex and multifaceted field that needs a multifaceted approach especially in the case of small and medium size enterprises (SMEs) where their size and budgetary constraints often prevent them from developing a fine-grained organizational structure such as policies and procedures. The European Union has with its Digital Single Market Strategy put emphasis on it within the union. This is followed by a number of cybersecurity initiatives to work towards more secure online environments. However, for SME's this is a challenge in terms of knowhow, economy and maturity. The purpose of this paper is to discuss cybersecurity challenges European SMEs face today, and to hold this towards the European initiatives on cybersecurity. The paper concludes that an effective SME cybersecurity framework must be composed of a combination of technical knowledge, leadership management and standardized processes to cope with the European Union's cybersecurity initiatives.

**Keywords**—SMEs; Advanced Persistent Threats (APTs); Attacks; Cybersecurity Standards.

## I. INTRODUCTION

Today's marketplace demands an efficient Information and Communication Technology (ICT) framework to facilitate business operations between businesses and their clients, employees and business partners. This helps businesses to stay competitive in an increasingly evolving technological marketplace. Around the world, governments put more emphasis on the use of ICT in different sectors, including SMEs. Additionally, with the wide adoption of internet, reduced prices of technology and hardware and a growing need of applications have compelled businesses of every size to depend on computers and internet to store, manage and transmit vital information of an enterprise or its clients. This information has value. Whether it is a clients' personal or confidential information or assets belonging to service providers all these entities have a strong need to protect their information from hackers, cyber criminals and even insiders who seek to steal from or damage an enterprise. According to the European Commission, SMEs in Europe today face particular challenges in addressing basic cybersecurity threats [1]. These small and medium enterprises are more vulnerable to the whole spectrum of attacks because of their size and budgetary constraints as well as their general lack of being security aware and that they serve as a link to bigger

companies and public authorities [2]. This often precludes them from developing a fine-grained organizational security structure to include such as policies and procedures. Lacking in such, leaves enterprises susceptible to attacks from insiders and intruders who can gain direct access to a company's computer systems.

The European Union has issued its European Digital Single Market Strategy in which there is a vision of the EU to become fully digital and hereby become "world leader in information and communication technology" [3]. Such a vision is foreseen to involve all sectors of the European economic system and therefore also SMEs. Such a vision puts emphasis on how to implement information and communication technologies in all business processes but also the challenge is to understand how these processes can be secured and become less vulnerable towards cybersecurity attacks.

The paper is organized as follows: Section 2 describes some of the general cybersecurity challenges seen today. Section 3 identifies cybersecurity challenges SMEs faces today, ex., data handling and storage. Section 4 presents the European Union documents relating to cybersecurity. Section 5 presents the best practices for SMEs to comply with the European Union strategies and initiatives. Finally, Section 6 presents the conclusions.

## II. DEFINING THE CYBERSECURITY CHALLENGE

In the past, implementing cybersecurity was quite simple. Most of the cybersecurity threats were limited to viruses and worms. These threats were mostly considered an act of vandalism because the attacks were random to the computers that were connected to the internet. With the help of firewalls on the outside and anti-virus protection on the inside, the enterprises seemed protected and safe. But slowly the attackers moved from random attacks to targeted attacks and started to get inside the enterprise networks. Attackers started infecting vulnerable machines and took control via command-and-control (C&C) systems. Vulnerabilities were exploited to move laterally in the network to gain more and more credentials in the enterprise. Ultimately the goal of an attacker is to gain administrative privileges and get complete control of enterprise systems. These kinds of attacks are known as Advanced Persistent Threats (APTs). In these attacks an

attacker is particularly skilled and effectively uses IT technologies to breach and by-pass all layers of protection. As compared to conventional attacks i.e., arbitrary release of viruses and worms, the persistence of the attack makes APT different. An APT attacker constantly adjusts the attacks based on the defenses laid in front. A new kind of defense approaches such as Continuous Persistent Monitoring (CPM) are required that adapts to the attack as quickly as the attack adapts to the defenses [4].

Behind every attack on the enterprises, attackers have different motivations and objectives. These attackers are broadly classified into following five categories.

**Commodity Threats:** These threats are random and non-targeted attacks to gain access to enterprise networks through malware, viruses, worms and techniques widely available across the Internet. These undirected attacks can be automated and are intended to exploit vulnerabilities or other cyber-defense weaknesses. Countermeasures for these kind of threats in enterprises are usually taken care off by tools that are focused on detecting specific malwares, known bad traffic and other patterns that can be codified into signatures [5].

**Hacktivists:** The activist of the digital landscape. Hacktivists use hacking to make a public or political statement [6]. These groups usually have different motives behind different attacks and use similar tools that penetration testers and other security professionals use. The defenses to protect against these tools and techniques are widely available. The hacktivists take advantage of vulnerabilities that are unpatched or open of exploitation.

**Organized Crime or Cybercrime:** Are also targeted attacks. Factors like speed, convenience and anonymity makes the Internet particularly attractive to criminal elements. There is no specific definition of cybercrime but law enforcement agencies generally make a distinction between two main types of Internet-based crime i.e., Advanced cybercrime; related to sophisticated attacks against the hardware and software and Cyber-enabled crime; where Internet is used in traditional crimes like financial crimes [7].

**Espionage:** Unlike cybercriminals, where the motive or goal is to gain access to the computers and accounts and steal money or steal data than can easily turned into money; cyber espionage is a little more complex in its objectives. The attacks centers on stealing trade secrets also known as “crown jewel” of the enterprise or national secrets for political or military advantage. APT-style methods are frequently practiced by the attacks.

**Cyberwar:** As defined by oxford dictionary, cyberwar is “the use of computer technology to disrupt the activities of a state or organization, especially the deliberate attacking of information systems for strategic or military purposes”. The attacks are enabled by overwhelming, overloading, disabling or destroying the IT systems to malfunction and damage themselves or their operators.

The complexity and sophistication of cyberattacks continue to grow. In most of the cases the main purpose of these attacks are to steal information/ money, disrupt or deny information/capability and take over the machines [8]. These

sequence of advanced cyberattacks has been given many labels, including kill chain and attack life cycle, see Figure 1

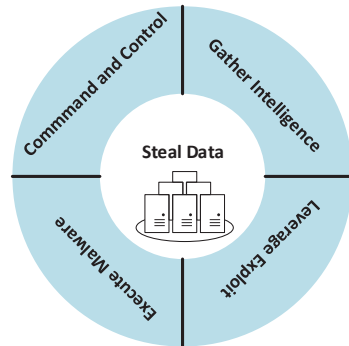


Fig. 1. Attack Life Cycle

There are multiple methodologies to create and deploy these series of stages but the main characteristics and most common are the following:

**Gather Intelligence:** Research, Identification and Selection. At this initial stage, information about the target is gathered ex., e-mail addresses, social relationships or information on specific technologies to detect weaknesses and exploits for later stages.

**Leverage Exploit:** At this stage, the information gathered about the discovered weaknesses and exploits are used for an initial infection with malware.

**Execute Malware:** The malware runs within the organization to expand control to other workstations, servers and infrastructure elements. This is done by escalating privileges i.e., acquiring administrator privileges.

**Command and Control:** Attackers take action to ensure continued control over the network from outside the network. Family of malware are installed on different computers to establish C&C channel.

**Steal Data:** The main goal and the final step of the attack is to steal data or any other “crown jewels” of an enterprise.

### III. SME CYBERSECURITY CHALLENGES

Most of the businesses- especially the SMEs- today have to be mobile and flexible. A critical non-negotiable element of an SME is that the business should take place across platforms and geographies [9]. Architectures and technologies like thin clients, mobile devices, cloud storage, apps and remote processing come together in different constellations to enable the modern workforce and customers. However, investigations and surveys have shown that these advancements present SMEs with a whole new set of challenges [10]. Computer networks can be breached, personal data can be compromised, critical confidential corporate information or classified commercial secrets can be stolen by employees, Web sites can

be hacked and keystroke loggers can be surreptitiously installed [11].

According to Gartner 40% of large companies will have adopted appropriate security systems in order to defend themselves against cyber-attacks by 2018 [12]. However, SMEs will still represent easy targets because of budgetary constraints, lack of awareness and are often not able to detect the extent of theft which has taken place. This leaves SMEs as easy targets for cyber criminals, especially the ones that collaborate with major corporations, producing individual parts or who play a role in complex projects. These SMEs provides a gateway of cyber criminals that wish to enter the systems of large corporations via backdoors.

In terms of implementation of cybersecurity countermeasures, SMEs can be broadly classified into three categories.

*First category of SMEs* is the ones with lack of awareness and knowledge. Here the company thinks that they are too small to be of interest to the cybercriminals or they're not targets for attackers as they don't have anything worth stealing. However, most businesses have information that needs to be kept secret ex., customers' credit card numbers or employees' personal data etc. According to FireEye [13], this is one of the core reasons why small and medium size businesses especially become inviting targets i.e., "Your data is more valuable than you think". In situations like this most of the SMEs rely on security through obscurity and do not really implement any security. In addition to having valuable data of their own, most SMEs do businesses with larger companies. This often means trust relationship between businesses, computer systems as part of an integrated supply chain or access to their sensitive data and intellectual property. Even though larger enterprises are capable of protecting their own businesses from adversaries with state-of-art technologies, cybercriminals exploit the trust relationship by hacking into SMEs and get access into the bigger company's networks.

*Second category of SMEs* is the ones that underestimate the threat of cybersecurity. Here the company does not prioritize improvement to security for their future business growth. The company sees security as more of a hassle than anything else, so a bare minimum is implemented to comply with mandates and government regulations. The security architecture does include most of the cyber essentials ex., boundary firewalls and internet gateways, secure configuration, access control and malware protection but if patch management and software updates are not in place, satisfying bare minimum requirements will not offer much in terms of security. Any hardware equipment and software need regular updates to keep it running smoothly and to fix any security vulnerabilities. Anti-virus, Anti-malware, Intrusion Detection/Prevention Systems (IDS/IPS) and firewalls needs regular updates in order to continue to provide adequate protection, failing to do so will vitiate the security measures company had in place. Tactics used to exploit above mentioned vulnerabilities in both the categories are for example: *Zero-day vulnerability* [14], *Spear phishing attacks*

[15], Water Holing, Ransomware, Botnet or Denial-of-Service Attack.

*Finally, the third category of SMEs* are the ones that comply fully with regulations and standards. The company has taken a more proactive approach towards cybersecurity. All cybersecurity essentials are in for example, Boundary firewall and internet gateways, secure configuration, access control, malware protection, patch management and software updates.

In conclusion, SMEs can come over most of the above mentioned threats by simply employing good IT security practices like patching vulnerabilities in IT systems. Knowing the value of data is paramount whether its employee's/ enterprise personal data or customer credit card information. SMEs need to recognize that the data may be at risk and some appropriate technical measures need to be taken in order to secure the data. These measures should fit the needs of SMEs particular business. Following are the largest threats to SMEs: *Adversaries attack against infrastructure*, a continuous refining of tools and techniques for malicious exploits for gaining access to web servers and databases

*Malicious encounter towards electronics manufacturing*. The supply chain is vulnerable to attacks and corruption.

*Insecure Passwords*; adversaries make use of employee's email to gain access to personal accounts leaving enterprise vulnerable to theft and fraud. In most of the kill chains or attack life cycles an adversary tries to escalate their privileges in enterprise's system after establishing foothold. Escalating privileges involves collecting items that will allow more and more access to the resources within victim's enterprise. In most of the case this is done by procuring usernames and passwords. Surveys have shown that 56% of employees reuse passwords for the personal and corporate applications, 14% of employees use same passwords across all applications and 20% of the employees share their passwords with their team members [16].

*Phishing and Spear-phishing emails* are still the popular choice of weapon for an attack on a SME.

*Lack of fine gained control of policies in the enterprise*. This factor comes from either lack of awareness or budgetary constraints. Liberal policies for internet usage invite more attacks and intrusions.

*Use of trusted applications by adversaries to exploit gaps in perimeter security*. Attacks like water hole targets specific industry related websites to deliver malware. In addition to that, social networks pose even more dangers to an SME.

#### IV. EUROPEAN CYBERSECURITY STANDARDS

The European Union has around 31.5 million Internet users [3] however, only 15% of the Europeans shop from another EU country and furthermore, only 7% of SMEs sell cross border. The overall reason for the Digital Single Market Strategy [3] is to explore the potential of becoming a digital market and taking away the US monopoly ([3], 54% of EU citizens use US based services today). Some of the central elements in the strategy are: removal of digital barriers and a guarantee on portability across countries in EU, focusing on the different platforms in the EU, and at the same time making



sure that this can take place in a secure and trustworthy environment. In particular, the EU looks at data protection, telecoms rules, and obligatory ICT standards (including security) in all EU countries. Security has come on the European Union's agenda and a number of cybersecurity frameworks have been issued:

- The European Telecommunications Standards Institute (ETSI) [17]
- The International Organization for Standardization (ISO) – 27001, 27002, 27032 [18]
- The International Telecommunications Unit (ITU) – National Cybersecurity Strategy Guide [19]
- The Council on Cyber Security Critical Security Controls [20]

One of the more important documents is the Cybersecurity Strategy of the European Union [21], and the Information Security and Privacy standards for SMEs [22]. The Cybersecurity Strategy for the European Union [21] presents the vision of a European Union and put emphasis on national security strategies that shall comply with the digital vision. Following are the objectives of National Cybersecurity Strategies which are also in line with European Cybersecurity Strategy [21].

- Cyber Resilience; efficient cooperation within the public and private sector.
- Securing critical information infrastructures
- Reducing cybercrime.
- Development of the industrial and technological resources for cybersecurity.
- Contribution to international cyberspace policy.

The European Union Agency for Network and Information security (ENISA) presents recommendations to improve the adoption of information security and privacy standards for the SMEs within EU. The standards look through specific barrier and provides guidelines on how the SMEs should comply with the barriers for cybersecurity management.

As presented in the previous section, barriers are related to lack of awareness and engagement from the enterprises, limited resources and budgetary constraints and also shortage of standards, as most of the standards are driven from larger size organizations and many standards are not easily scalable for use by SMEs.

#### V. DEALING WITH CYBERSECURITY: GOOD PRACTICES

Good practices or a good program in cybersecurity is not only about compliance and checklists. Nor just about technology, defenses against malicious activities or just protecting people and assets. An optimal program protects the enterprise by balancing all the variables in a cost-effective manner while supporting the business mission as much as possible. A cybersecurity program needs to be implemented to proactively defend the networks. This program must be composed of people, technology and standardized processes as it requires a combination of business savvy, technical knowledge, leadership, management and common sense. Following figure 2 shows the intersection of four factors that must be considered for an effective cybersecurity framework.

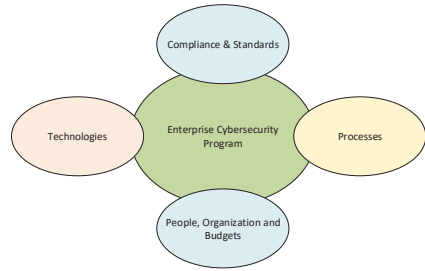


Fig. 2. The Cyber Defence Framework; figure shows the intersection of people, organization, budgets, technologies, processes and external or internal compliance.

As mentioned in section 3, in most of the cases SMEs work together with large corporations. In order to do that SMEs, need to fulfill requirements for external validation ex., External Audit that the enterprise security measures are in place and functioning properly. These can be enterprise's own requirements, from external entities or from regulators/ auditors. However, an enterprise can be secure without being compliant and compliant without being secure [4]. Even with the most security management systems implemented and compliant with the standards, someone still needs to be there to investigate the alerts and interpret the reports. Compliance must not be treated as the only cyber defense objective. Process and Technology goes hand in hand and there must be a fine balance so technologies are effective and the processes are manageable. The most fundamental factor in the framework are people, as they are the one that make the program succeed or fail. In order to execute an effective program people must have a clear understanding of what to expect from their enterprise's information security program. They need to know how to direct the implementation of an information security program, how to evaluate their own status existing security program and how to decide the strategy and objectives of an effective security program.

The goal of the above mentioned frameworks is to provide a methodology for cybersecurity in the enterprises and making sure that the methods comprises the most important elements of protection and defense. Following is the consolidated list of good practices and recommendations an SME can do to protect against threats:

- Assessment of threats and risks to specific SMEs. Consider how valuable, sensitive or confidential the information is and what damage could be caused to employees and clients if there was a security breach.
- Getting the basics right [13] i.e.,
  - Download software updates and Patch Management: Security software like anti-virus and anti-malware needs regular updates in order to continue to provide adequate protection.
  - Uses of strong passwords: Encourage employees to use strong passwords, refrain from using the same

passwords for multiple logins or even use two-factor authentication ex., FIDO Alliance [23].

- Delete suspicious emails: Know how to recognize phishing.
- Awareness and Training: Most of all security incidents involved human error and employees pose the biggest vulnerability to the IT system. Educating employees on how to protect sensitive data using security best-practices is crucial to safeguarding your business.
- Encrypt sensitive data such as employee details and financial accounts
- Control all traffic though the network by granting access to certain IP and e-mail addresses by creating “whitelists” to prevent employees from visiting compromised or receiving malicious emails. This can be executed by deploying Security Information and Event Management (SIEM) solutions
- Administrative accounts should not have access to email or internet accounts of other employees to prevent attackers entering the system through these channels. Ideally two-factor authentication can be implemented if the administration user needs access.
- Implementation of fine grained control of policies. Good policy will enable enterprise to address the risks in a consistent manner.
- Secure Enterprise data in the cloud.
- Implement cybersecurity controls to reduce confidentiality, integrity and availability risks [4] ex., Preventive, Detective, Forensic and Audit controls.

## VI. CONCLUSION

In this paper we have addressed some of the main cybersecurity challenges related to SMEs. Generally, the challenges facing SMEs are not very different from the challenges facing other/ bigger enterprises. However, the ability and resources to face the challenges turn them in to different problems. We categorized SMEs in three segments, where one part of the segment has started to address the threats and risk in their enterprise and place defensive countermeasures. Some are hesitant to implement any cybersecurity program mostly due to budgetary constraints. However, the cost of inaction potentially far outweighs the upfront cost of cybersecurity [13]. SMEs evidently need to be reminded what is at risk. SMEs tend to store confidential information such as client lists, customer databases or financial details which are highly prized assets for criminals. The attackers make money through sale of stolen information, holding data to ransom or selling product designs or manufacturing processes to competitors. According to Symantec estimated cost of stolen information on the black market in 2014 was around € 0.30 - € 13 for credit card details, € 10 - € 3000 for custom Malware, Drive-by downloads Web Toolkit (1-week rental) € 78 - € 590 and online banking malware € 124 - € 1050 [24].

An attack can damage a business’s financial health due to stolen money and the recovery process can be lengthy and

costly. In conclusion the longer SMEs wait to take appropriate measures, the risk they will be exposed to develop and grows. The only option is to treat the risk by mitigating the threat.

## REFERENCES

- [1] Secure Societies, Protecting freedom and security of Europe and its citizens. Work Program 2016 – 2017. Horizon 2020.
- [2] B. Ommen, “IT Security in SMEs: Necessary or Irrelevant?” 21st Twente Student Conference on IT, Enschede, Netherlands, 2014
- [3] A Digital Single Market for Europe, [http://ec.europa.eu/priorities/digital-single-market\\_en](http://ec.europa.eu/priorities/digital-single-market_en), 2015
- [4] S.E. Donaldson, S.G. Siegel, C.K. Williams, and A. Aslam, “Enterprise Cybersecurity – How to build cyber defense program against advanced threats,” Apress, May 2015.
- [5] J. Jaccard, Julian and S. Nepal, “A survey of emerging threats in cybersecurity.” *Journal of Computer and System Sciences* 80.5, 2014, pp.973-993.
- [6] B.B. Kelly, “Investing in a Centralized Cybersecurity Infrastructure: Why Hactivism Can and Should Influence Cybersecurity Reform.” *BUL Rev.* 92. 2012.
- [7] Europol: European Cybercrime center (EC3), <https://www.europol.europa.eu/ec3old>
- [8] Architecture Matters, Beyond Layered Security Approaches, White Paper. Paloalto Networks: 2016
- [9] G.K. Trellis, “IT Security: It’s Everybody’s Business (Insight from 31 CIOs from the private sector in denmark).” *KMD White Paper*, 2nd Edition (2016).
- [10] Business Crime Reduction Center: Cybercrime: SME ICT Security Programmes. <http://www.bccr-uk.org/>
- [11] N. Malgorzata, M.J. Martin and Z. Polkowski, “The cyber security in SMEs in Poland and Tanzania.” *Electronics, Computers and Artificial Intelligence (ECAI)*, 2015 7th International Conference on. IEEE, 2015.
- [12] “Attack on Sony Pictures Is a Digital Business Game Changer.” *Gartner Press Release*: <http://www.gartner.com/newsroom/id/2990717>
- [13] “Big threats for small businesses – Five reasons your small or midsize business is a prime target for cybercriminals” *White Paper*. FireEye, Inc
- [14] D.R. Han, “SME Cybersecurity and the three little pigs” *ISACA Journal* Vol. 6, 2012.
- [15] “Spear Phishing Attacks – Why they are successful and how to stop them” *White Paper*. FireEye, Inc
- [16] Survey Sail Point. 2014 Market Pulse Survey: Employees Going Rogue with Corporate Data. <https://www.sailpoint.com/news/2014-market-pulse-survey/>
- [17] C.B. Rizzo, “Security for ICT-The work of ETSI. ETSI White Paper No.1. ISBN No: 979-10-92620-05-4. 2015
- [18] ISO/IEC27001Information Security Management. <http://www.iso.org/iso/home/standards/management-standards/iso27001.htm>
- [19] Dr. F. Wamala, “The ITU National Cybersecurity Strategy Guide”.
- [20] Council on Cybersecurity, “The Critical Security Controls for Effective Cyber Defense”. Version 5.0
- [21] European Union Agency for Network and Information Security, “An Evaluation Framework for National Cyber Security Strategies”, 2013.
- [22] European Union Agency for Network and Information Security, “Information Security and Privacy Standards for SMEs”, 2015.
- [23] S. B. Srinivas, D. Tiffany, and E. Czeskis, “A. Universal 2<sup>nd</sup> Factor (U2F) Overview”. FIDO Alliance, 2015.
- [24] C. Wueest, “Security Response, Financial Threats 2015”. Version 1.0. Symantec.2016

# **Sponsors**



SCHOOL OF BUSINESS AND SOCIAL SCIENCES  
AARHUS UNIVERSITET



*Published, sold and distributed by:*

River Publishers  
Alsbjergvej 10  
9260 Gistrup  
Denmark

River Publishers  
Lange Geer 44  
2611 PW Delft  
The Netherlands

Tel.: +45369953197  
www.riverpublishers.com

ISBN: 978-87-93609-30-3 (Hardback)  
978-87-93609-29-7 (Ebook)

©The Editor(s) (if applicable) and The Author(s) 2017. This book is published open access.

### **Open Access**

This book is distributed under the terms of the Creative Commons Attribution-Non-Commercial 4.0 International License, CC-BY-NC 4.0) (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated. The images or other third party material in this book are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt, or reproduce the material.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper.



# Proceedings

Global Wireless Summit-2016

November 27-30, 2017

Aarhus University, Aarhus, Denmark

ISBN 978-87-93609-30-3



9 788793 609303