altice labs

InnovAction

#2 | 2017

Creativity is thinking up new things.

Innovation is doing new things.

Theodore Levitt

InnovAction

#2 | 2017

The needed transformation

Dear Reader,

Telecommunications and Media Industries have been facing major disruptive and unpredictable forces for the last decade, having to rapidly adapt and rethink their strategies while still delivering top quality service to their clients and value to all stakeholders. Competition nowadays is not only with traditional players, with whom network-based Communications Service Providers (CSP) will keep fighting, but the greatest threats come from global players like the so-called GAFA: Google, Apple, Facebook and Amazon. There is a need to create a new offer of digital services, with a unique, outstanding and differentiated customer experience, associating the brand of the CSP to innovation and exceptional positive emotions, in order to acquire new customers, most of them digital natives.

The needed transformation must address several dimensions along the journey to become an agile company, in order to compete in the digital arena and to lead in customer experience and digital services, namely:

- The need for an extensive cultural and organizational mind shift – the CSP becoming a Digital Service Provider (DSP) focusing on new platform-business models;
- The transformation of the network into a fully virtualized and programmable platform;
- The journey to become a data-driven company;
- The exploration of new revenue opportunities, in the areas of digital media, security, Internet of Things (IoT) and IT services, and the focus on B2B products and services;
- The impact of technologies now going mainstream, like artificial intelligence (AI) and blockchain.

For the network-based CSP, the foreseen challenges are huge: the more people connect to all kinds of broadband access points, the more successful the GAFA will be. Giant Internet companies will continue to drain a significant revenue stream from CSP, content and other utility providers.

On the other side, CSP will have to redirect their investments accordingly to a well-defined technology strategic plan supporting competitive value propositions for their customers as otherwise, they face the risk of losing a significant amount of communication and entertainment business or, in the worst scenario, becoming totally irrelevant.

For a group like Altice, which has a very clear investment strategy in telecom and media, a technological publication such as the present one, is particularly relevant, firstly as a vehicle to share the most relevant technologic developments and knowledge throughout the group's operations all over the world, and, secondly, as a way to reinforce Altice's firm commitment to technological innovation and the future.

I hope you enjoy reading it, as much as we enjoyed writing it.

Alcino Lavrador
General Manager of Altice Labs

Editorial note

With a 2020 horizon, this year publication is not so much about technology per se, since this is no longer a key differentiator factor, but more about what technological evolution has caused and may still cause in society. Actually, one of the main affected actors are communication service providers (CSP) that should totally embrace the digital revolution, becoming the digital service providers (DSP) of the (near) future.

Presenting a total of fifteen articles, the second edition of InnovAction intends to summarise the main digital trends that these new providers need to somehow address in order to fit the current market context, more and more digital:

- **Prepared for a digital world you must be:** summarizes what authors believe to better outline the near future, what society in 2020 will look like according to several senior consultants that attended an ideation session held at Altice Labs;
- **Smart living powered by Digital Service Providers:** presents a reflection on the demands of future customers in a digital society and their impact on the transformation of CSP strategy to answer the new smart living requirements;
- **Smart cities - the importance of sharing:** provides not only a perspective on how technology advances and data availability are reshaping long-time established social and economic relations among people and between people and companies, but also how they are transforming the way cities are governed, both at the operational and strategic planning levels;
- **Digital innovation labs and the digital CSP:** explains what digital innovation labs are and what they aim by briefly analysing relevant experiences, pitfalls as well as suggesting possible approaches to take the most out of them;
- **Merged reality for everyone:** besides connecting the current status of merged reality with exploratory research and development work carried out by Altice Labs, the article addresses some challenges and business opportunities within the promising merged reality ecosystem;
- **Industry 4.0 - The telco perspective:** synthesises how telcos may be part of the huge transformation the industrial sector is facing in order to become a smart factory;
- **Blockchain – a brief introduction:** explains in a high-level manner the blockchain theme, aiming to provide readers with a basic understanding which they can use as a basis for exploring further resources to dive deeper in the technical aspects and potential applications;
- **Big data and recommender systems:** intends to present relevant techniques and algorithms for user profiling and how they can be used for analytics and recommendations purposes;

- **AI: overview and applications:** This article aims to introduce AI technology and applications, beginning with a brief historical overview and an approach to the multiple ways how it is defined and put under perspective by different protagonists of its evolution till nowadays.
- **Tackling security and privacy in a digital world:** analyses the European Union general data protection regulation in order to provide a list of technical security and privacy recommendations not only for systems but also for the IoT world;
- **A holistic study of cable and mobile identity credentials:** explores, through a joint study of Altice Labs and CableLabs, several scenarios regarding the evolution trends and convergence of digital identity management on future ecosystems of network service providers;
- **OTT multimedia content delivery: a study:** presents an analysis of the state-of-the-art and trends of TV/Multimedia services, proposing an OTT-based strategy in order to take advantage of the technologies used by global competitors;
- **From TV Flow to TV Concierge: a new TV experience:** provides an overview about the work that is being done, in a consortium where Altice Labs plays a major role, towards a completely personalized TV experience that savvy users expect;
- **Demystifying truths and myths of 5G technology:** introduces briefly 5G and summarises its current status, with a focus on standardization. It also sums up a number of questions and respective answers aiming at clarifying what 5G is (and what it is not).
- **Demystifying truths and myths of 5G technology:** This article describes the edge computing technology, the prototype under development by Altice Labs, its proof-of-concept integration and its evaluation results.

Finally, as a special gratitude and acknowledgement statement, it is important to highlight that InnovAction is the result of a strong teamwork were the identified authors shared their valuable knowledge and pertinent perspectives; technical and editorial reviewers gave their relevant contributions in order to improve the insights readers will receive; editorial team along with graphic team granted a coherent, rich and influent publication from Altice Labs to the Altice Group and all their customers and partners.

Pedro Carvalho

Ana Patrícia Monteiro

Arnaldo Santos

Altice Labs

Contents

01

Prepared for a digital world you must be!

Ana Patrícia Monteiro, Altice Labs
ana-p-fonseca@alticelabs.com

Nuno Seixas, Altice Labs
nuno-a-seixas@alticelabs.com

This is not the time for just strategic planning. This is the time for addressing strategy by defining a tactical plan. Let us engage digitally with 2020, let us act!

Keywords

2020; Digital Transformation; Telco; Contents Convergence; Business Models; Society; Environmental & Sustainability; Consumer

Introduction

The ability to predict the future has always been a desire of mankind. Business (and technology) aspire the same and so companies try to foresee what is coming by looking into the present and finding signals or leads that will reveal to them what the future will look like.

Aligned with this desire and wanting to outgrow their customers and stakeholders' expectations, Altice Labs, as a major enabler of innovation through technology, promoted an initiative for assessing the impact of digital convergence and transformation in 2020. How will the consumer behave? How about companies, what will change? What business models will prevail, the most disruptive ones, the ones focused on digital services? And the impact on privacy and security, what will cause, what will enforce? How will the environment be affected? How can technology foster sustainability?

These are only a few questions to help outlook the (near) future and were the main drivers in the above-mentioned initiative. As this article will show, many things are changing and many others will change and (digital) society must be prepared for it, to adapt itself. 2020 is just around the corner and digital providers must prepare themselves for it not tomorrow but today!

Methodology

Under the title "The impacts of telco, IT, contents convergence and digital transformation in society by 2020", Altice Labs conducted an ideation session on February 2017 - see **Figure 1** - to predict the major cha(lle)nges companies will face by 2020.

Distributed by five major work groups, several senior consultants from different areas, namely: telecom operators, media, technology companies, universities and research and advisory, gave their contributions based on their level of experience and knowledge about the following topics:

- Consumer in 2020;
- Companies in 2020;
- Business models for digital offer in 2020;
- Societal impact in 2020;
- Environment and sustainability in 2020.

The results of this work and the main ideas that Altice Labs believes to better outline the near future, what society in 2020 will look like, are shared in section "Impact in 2020".

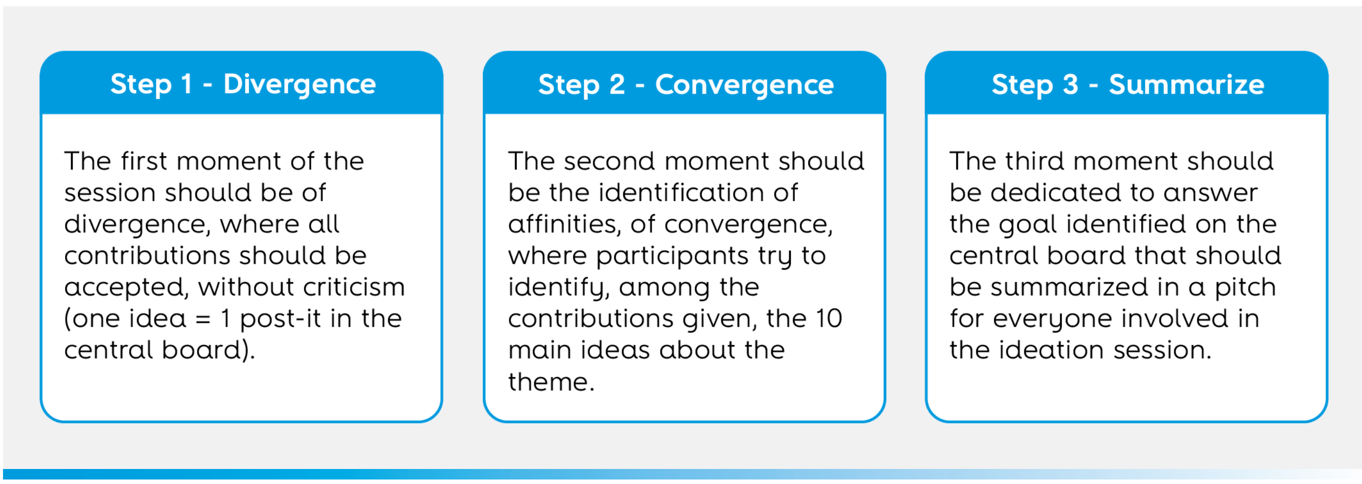


FIGURE 1 – Ideation session alignment

Digital convergence and transformation

What is digital convergence

There is a current trend in the telecom world. IT, telecom and contents, once three different worlds, with different technologies and even different approaches, are now being converged into one single digital experience.

Following the overall technical evolution, these three different aspects of how users interact with technology are making each one of them more reliable on the others, which translates into higher value for everyone involved, the user included.

For the user, it's simpler to use a similar interaction paradigm whether he is dealing with a telecom service, like data access, an IT service like an application for social media or even if he is consuming contents produced specifically for a given platform. Since users access is now more ubiquitous, they also expect their services to be coherent and convergent.

What first started with the mobile and landline convergence, making possible to customers to use both kinds of communication support without any difference, began also to incorporate IT convergence, since almost every aspect of a telecom operation is now controlled by IT systems. Lately, even contents production and usage are being considered as a part of the whole digital experience.

If this is already something fully verified in every modern operator, society will see an even greater convergence between these three aspects for 2020. Therefore, digital companies need to embrace this convergence into their own products and services portfolio.

What is digital transformation

"The world is being transformed by new technologies, which are redefining customer expectations, enabling businesses to meet these new expectations, and changing the way people live and work. Digital transformation, as this is commonly called, has immense potential to change consumer lives, create value for business and unlock broader societal benefits. Every industry has its nuances and contextual differences, but they all share certain inhibitors to change. These include the innovator's dilemma (the fear of cannibalizing existing revenue models), low technology adoption rates across organizations, conservative organizational cultures, and regulatory issues. Business and government leaders should continue to work towards addressing these challenges." [1]

Digital transformation is happening everywhere and it will keep on its path to globalisation in the following years. Software is now being put at the centre of every technological company, assuming its role as a strategic asset and not just as a support tool. In order to fulfil the promise of increasing the overall value delivered by digital services, companies need to undergo their own digital transformations. But this transformation cannot happen just by taking the way people used to do things and make it, following the same steps, but now on a computer. Instead, companies need to rethink their approach to their customers' needs and how they can fulfil them.

Digital transformation can be achieved by implementing user experience principles, making the whole user journey a more meaningful and valuable one. By doing this, companies will bring new digital services fully adapted to the new world, to new technologies and new generations.

Impact in 2020

The impact caused by digital transformation and convergence crosses and affects all societies, all its characters.

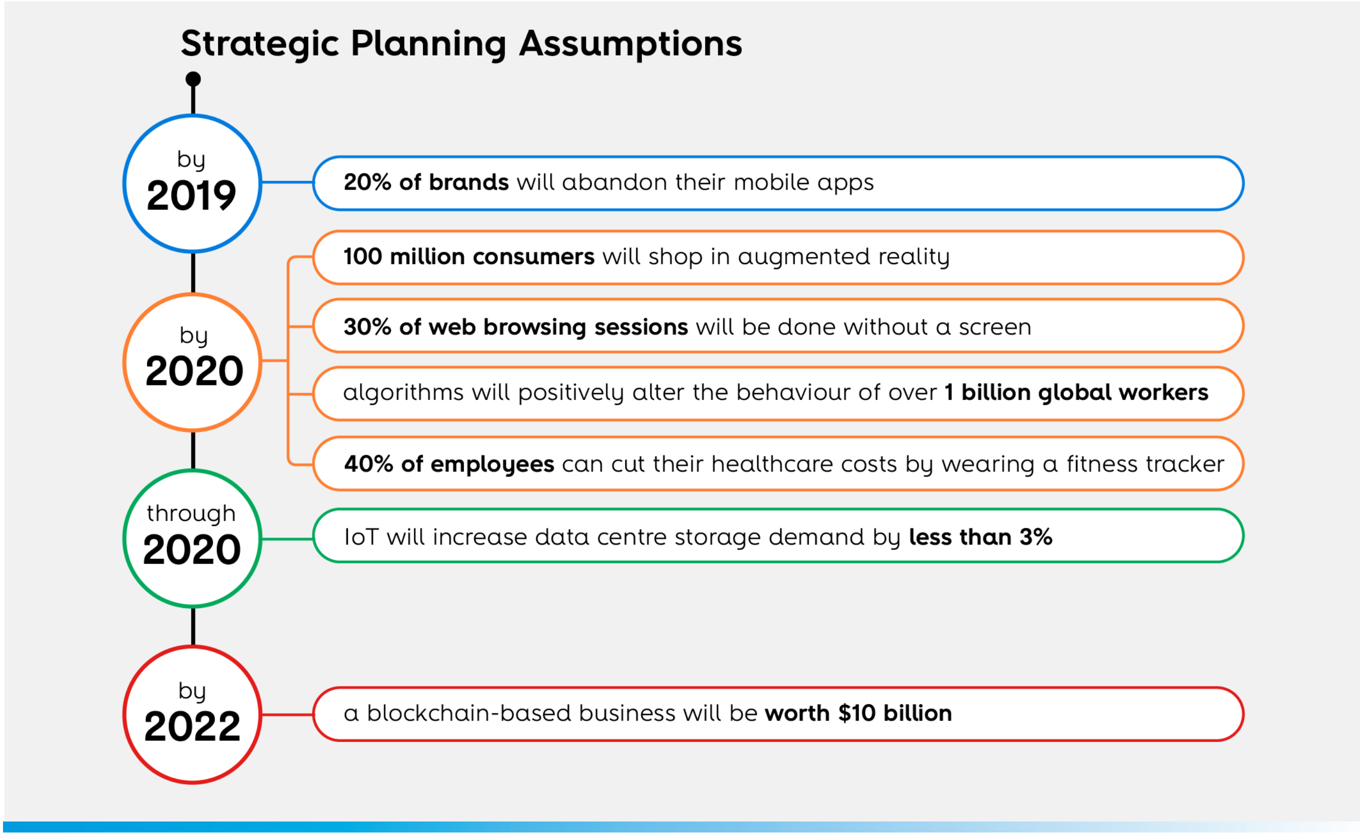


FIGURE 2 – How 2020 will look like [2]

As **Figure 2** shows, many things will change and all players must be prepared for them, prepared to adapt themselves. Therefore, and because 2020 is just around the corner, five of the foremost themes were selected and will be detailed in the following sections (Consumer in 2020; Companies in 2020; Business models for digital offer in 2020; Societal impact in 2020; Environment and sustainability in 2020) in order to help outline not just a long term strategy, because time urges, but an action plan for the next two years.

Consumer in 2020

Before going into further details it is important to note that during this article the authors will use consumer and customer as an interchangeable definition since it is intended to represent the end user for digital services, both from companies and personal perspectives. Hereupon meet consumer 2020 (C.2020), an always-connected persona that is simultaneously customer, user, buyer, renter and, especially, an influencer (positive or negative), all of them empowered by the technology ubiquity (see **Figure 3**). C.2020 is, or expects to be, the centre of the world for companies, for service providers and for all those who want to sell him something. Provide him with the best-tailored experience, the best product for whatever he wants, and you will conquer him. On the opposite side, provide him with just regular experience or a product without differentiation and a "please, do not disturb" is all you will receive next time you contact him. *"The customer of 2020 will be more informed and in charge of the experience they receive. They will expect companies to know their individual needs and personalize the experience. Immediate resolution will not be fast enough as Customers will expect companies to proactively address their current and future needs."* [3] Consumers in such a digital time expect machine learning, recommendation engines, artificial intelligence (AI) and foremost technological trends to work for them (and not only for marketers) in order to create personalised recommendations

Consumer in 2020

- Connected consumer
- Digital consumer/customer
- Empowered customer/consumer (with more access to information and more control over their experiences)
- Financially conscious (due to the economic crisis)
- Everyone is exposed
- Very knowledgeable
- No time available/impatient/everything NOW
- Lots of info to make decisions
- Doesn't accept mistakes from providers
- Wants tailored, personalized, integrated services and experiences
- The centre of the world
- Social conscience
- Buy/use/rent everything digital
- Wants to buy experiences, to be surprised and delighted
- Wants to try before buying
- Wants full freedom to choose and change

FIGURE 3 – Meet C.2020

and experiences that fit exactly their desires, saving their precious time. *"Consumers expect their experience to "automagically" adapt whenever they engage physically, digitally and emotionally. And now advanced computing techniques can harness expanding volumes of personal data (e.g. search, social, geo-tagged sensors, payments, shopping carts, speech) to create the magic behind new hyper-personalized experience."* [4]

By having huge amounts of information available, C.2020 is not naïf, he is knowledgeable and aware of the best offers. *"Savvy and highly informed: More and more Customers are turning to the Internet to find the latest information on products, solutions, and best practices. In 2020, you'll be dealing with Customers who are more informed and operating from a far superior base of understanding. The customer will know more and will expect you to know more too."* [3] Thus, this consumer seeks online recommendations, analyses ratings from different web pages and, after consulting his (e-)friends (those belonging to his social networks, those who can destroy the reputation of a brand!), he will make his decision expecting nothing but reliability and to be delighted (by operational simplicity, quickness or by the availability of multiple digital and contextualized interaction channels and support) at any stage of the relation with the seller. *"More stakeholders – a buying "ecosystem" – are becoming involved in purchasing decisions (…). The companies that will be most successful in this new arena are those that understand each customer's ecosystem and cater to the various needs represented."* [3]

Finally, no provider should expect loyalty as given because C.2020 will want to use what was tailored for him but will not commit himself until he realizes the value of the offer, until he feels he really is someone special for the seller. Most of the times, C.2020 will prefer to rent something instead of owning it since this will give him the freedom to change more easily from one provider to another, from one offer to another he believes it is best for him at a given moment. Approaches like "try before you buy", "limited access to premium" or "share between the community" may be good approaches to seduce him.

Companies in 2020

"Darwin at the speed of light: change or die" [5] was one of the most assertive sentences used to describe how companies must behave not only in 2020 but as soon as possible – see **Figure 4**. In fact, *"adaptability is more vital to success than ever: Change is going to happen whether you pursue it or not."* [6] Having a heavy hierarchical structure, outdated processes and decision workflows, that don't take the most out of the companies' data (use AI-based applications in your daily operations instead of only trying to sell it), will only stop companies to fast adapt to new regulations and law enforcements, new policies, political instability and, most of all, new demands from customers and from markets. Thus companies, brands, providers must forget straight-forward paths, business-as-usual postures in order to quickly resolve the several needs, aggressive strategies from competitors and several trends that emerge constantly and rapidly.

See, for example, how Industry 4.0 [7] is affecting not only the way physical facilities are built/shaped - *"Industry 4.0 introduces what has been called the "smart factory," in which cyber-physical systems monitor the physical processes of the factory and make decentralised decisions. The physical systems become Internet of Things, communicating and cooperating both with each other and with humans in real time via the wireless web"* [8] - but also, and not always highlighted, the way companies act and evolve. *"Industry 4.0 means getting to grips with radical new approaches to business rather than merely making incremental improvements to established business models. To achieve this, companies need to develop new skills, both at individual employee level and within the organisation as a whole. A solely top-down approach will create resistance in the organisation while introducing pockets of innovation within traditional business will provoke a reaction from less engaged employees.* [9]" In fact, in order to be ahead in this fierce competition, companies in 2020 will need not only to have access to skills and knowledge but to have access to the best skills and to the right knowledge at the right time. How companies source, how companies build their workforce will also be critical and differentiator for a place in the pole position; cooperation and exchange, even if temporarily, between companies inside an Enterprise Group or even between partners

Companies in 2020

- Innovation-driven
- Fully digital
- Agile and highly flexible
- Capable of cope with regulation, data privacy, political instability
- Main areas to provide: knowledge, structured data, processing capacity, security, virtual and augmented reality, e-health solutions, gamification
- Will buy mainly: as a service, on a global scale, with risk sharing models and via web/complete digital

FIGURE 4 – This is how companies should be (in 2020)

should be considered as an option to be agile in this matter.

Finally, see also how digital transformation, a demand (or reality, if preferred) from the current era, is also affecting the context of the companies. *"Modern enterprises succeed when they adapt to industry and marketplace shifts and incorporate new technology into company culture and regular operations. However, digital transformation isn't only about technology, it's about bringing together the power of technology with a culture that embraces the change that it can lead for the organization."* [6] Actually, companies wanting to survive and to best position themselves in the market must not only be technological/digital-driven, by embracing technology and all its benefits but, most of all, be innovation-driven, truly innovation-driven. Companies achieve this state by focusing, on one hand, in truly create and capture value in new ways as mentioned in Oslo Manual, (*"a common feature of an innovation is that it must have been implemented. A new or improved product is implemented when it is introduced on the market"* [10]), and, on the other hand, in providing tailored experiences, above the average products and services that leverage their customers' experience. As mentioned above in the "Consumer in 2020" section, these companies need to make them feel special, unique and valued at any point throughout the interaction with the vendor. Competition based only on price, on the lowest price, regular offers, with "more of the same" and no added-value, are no longer what users are looking for.

Business models for digital offer in 2020

Having in mind what was mentioned in the above sections, namely that customers are going to behave differently and companies must adapt themselves to that and to the fact that digital transformation and convergence will impact them (positively if not ignored), there will be no space for traditional business models, based on one or few interaction channels, built on linear workflows and without integrated data - see **Figure 5**. Hence, some aspects need to be assured:

- Provide contextualised services, where customer only needs to tell their stories once, no matter the interface or channel, in order to leverage their satisfaction, increase problem-solving accuracy and reduce churn;
- Optimized analytical algorithms over customer's data will not only reveal what they most want, don't want or don't value but also will identify trends and buying preferences

Business models for digital offer in 2020

- Disruptive business models
- Peer-to-peer economy
- Circular economy
- Multiple (communication) channels
- The Non-Stop customer (purchases models are now dynamic, driven by digital technologies and online platforms)
- Vendors' offer possible to easily adapt to popular trends and customization

FIGURE 5 – How to sell in (the digital) 2020

which will allow companies to define different and more effective business models, more accurate and, also important, to foresee which partnerships need to be established in order to fulfil customers' needs;

- There's more than one interaction channel and at least one takes into account users' impatience and lack of time - apps, for example, make interaction with users easier and possible at any time.

"Businesses aiming to generate new customer value propositions or transform their operating models need to develop a new portfolio of capabilities for flexibility and responsiveness to fast-changing customer requirements: deliver business model innovation, drive customer and community collaboration, integrate cross-channel, get insights from analytics, optimize the digitally enabled supply chain and enable the networked workforce." [11] Customers with linear and constant behaviours are endangered and what was a need today may not be important tomorrow; what was the main interaction channel may be the one that nowadays stops them to buy. They are ever changing users requiring flexible selling approaches that most of the times don't require huge efforts or resources from the providers' side.

Customer-driven business models, with combined products and services (even if through partnerships with unlikely players - you provide video on demand (VOD) services and offer a pizza as optional service) co-created with customers will be the smart and disruptive ones, the ones that will delight and engage them. *"Customisation is a global trend and is likely to spread even more rapidly across manufacturing industry in future. Customers increasingly want to determine how their products are designed and made and will be having an input into development and production processes at an early stage."* [9]

Societal impact in 2020

The societal impact of digital products and services are immense and trying to discuss all of them in one ideation session would be a herculean task. We could talk about the way politics is being influenced by big data techniques, or how social networks are influencing the way humans relate to each other, or how our sense of security has changed in the last few years and how it will be changed for the next ones.

So, this ideation session focused on security and privacy issues and how they have influenced and will influence society in the next years. Security and privacy are now being transformed from an almost exclusively technical discussion to a larger and overall issue. Data has grown to be one of the most important factors in every discussion, in every political election or even in a simple buying decision, as Ovum signals it: *"Ownership of and access to data will be the basis of a key political debate (...)."* [12]. And, at the centre of this data collection, analysis and access are, of course, digital products and solutions. So, it shouldn't be any surprise for anyone technically savvy that society needs to understand and discuss the impact of security and privacy in the future of digital products and services, as pointed out in **Figure 6**.

One of the most important factors to analyse in this domain is the new General Data Protection Regulation (GDPR), published on 14 April 2016 and with its enforcement data set to 25 May 2018 [13]. Because of this new regulation which is commonly referenced because of its heavy fines, digital companies started to look with extra care into the personal data protection mechanisms. But one of the major impacts this new regulation will have is not directly related to fines. The key aspect is the reputation of a company and how data protection incidents will affect it. More and more users are sensible to questions related to data breaches and loss of their private data. These users, which will use more digital services and therefore, generate more private data, will be more sensible to the companies that demonstrate an adequate data management system. In order to address this point, digital products and services must increase their resilience and security, making this a key aspect right from the first design moment. Users expect their services to be secure and trustable with their data. And that's where

Societal impact in 2020

- Regulation
- Security and privacy by design
- Cyber-hygiene for everyone

FIGURE 6 – Main view on digital services security impacts

the new GDPR will play a key role. With this new regulation, users now have a basis for demanding not just reputation but also to claim from their digital service providers objective processes for managing their data.

In order to address these new demands, digital services must be built with this new set of needs in mind, and therefore, concepts like privacy by design and security by design will have an exponential growth in the following years. Companies developing digital services need to understand this shift and make sure their products and especially their product development processes have security and privacy by design in place.

Another important aspect and also related to security issues is how will the already present growth of devices being deployed address these security and privacy needs. In areas like the Internet of Things (IoT) and automotive connectivity, companies need to integrate different levels of security and privacy mechanisms, in order to deliver a solution that users can trust on. If not, these same users will abandon that solution in favour of something else that will fulfil not just the functional need but also the security and privacy ones.

Finally, for this domain of security and its impact on society, the ideation identified one last pressing concept - the cyber-hygiene for everyone. Society, as a whole, needs to understand the implications of using the different digital services and products, possible security threats on that use, how their privacy is being addressed or how their data is being gathered, analysed, processed and even shared.

Environment and sustainability in 2020

Environment and sustainability are currently two of the most used terms in the world. Based on climate change effects, which for the past few years came to grow as a comprehensive worry for everyone, along with the recently signed Paris Agreement, this domain is also impacting on technology leaders, providers and users.

From several different publications and analyst reports, it is clear that this sustainability demand will also affect technology providers. For example, from Ovum [12], we get the information that reducing power consumption will be a pre-requisite for almost every part of technology, whether we're talking about IoT devices, large or small data centres or even everyday consumer electronics. Another significant signal, highlighted by Gartner in its hype cycle for sustainability report [14], is that these environmental concerns are growing inside most of the companies (even if the acceptance is still very slow) and that stock exchanges have incorporated some sustainability indexes for most of the markets.

One important question from this challenging context is how can the technology-based companies incorporate these principles into their products and production lines while improving their profitability or, at least, without reducing it.

From the ideation session, the group identified reducing digital services footprint as one key aspect of this domain (see **Figure 7**). *"Throughout all the digital industry, companies need to rethink the way they design and offer digital services so that it saves energy and resources."* [14] In order to achieve this improvement, companies need to analyse the problem as a whole, i.e., looking at the complete product and service lifecycle. Being able to think about energy footprint right from the conceptualization moment and not just when the service is being delivered is something companies need to be able to do. But, in order for them to be able to do this, two key aspects need to be addressed: (i) necessary state incentives must be put in place and (ii) companies need to integrate different mindsets, skill sets and tools into their product development cycles.

For the first one, state incentives, it is now widely accepted that in order to have quicker results in the environment domain, states must provide incentives, whether positive (fiscal incentives) or negative (special regulation and specific fines) ones [12]. The states must, therefore, take this important task in hands and start developing this kind of mechanisms that will accelerate the overall usage.

For the second one, the new mindset, skills and tools, companies need to start building and integrating them into their teams. First of all, companies need to look at their products and services production lifecycle as a whole, incorporating sustainability concepts right from the design process, improving their production processes to be more sustainable and eventually, incorporating circular economy principles. Even for digital products and services, companies need to think how they can improve their energy footprint, if not just for the sake of the planet, but also because the energy cost will grow. Being able to produce the same kind of digital products and services with a fraction of energy

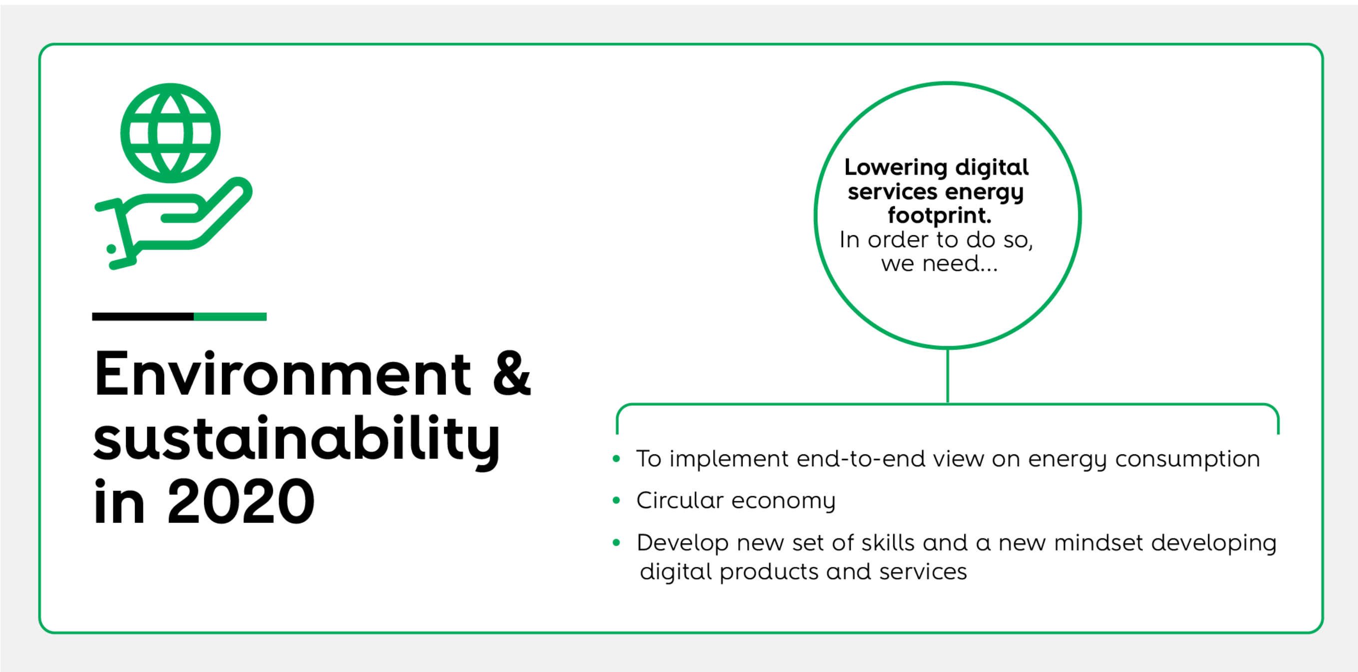


FIGURE 7 – Reducing digital services footprint as one key aspect

and raw materials is, therefore, one key element for every technological company continuous improvement efforts. Also and closely related to this organisational transformation, teams need also to acquire a new set of skills that will allow them to analyse different energy, environmental and sustainability approaches. For example, design teams for a telecommunication product can no longer discard the energy used in order to maintain the equipment at work.

Finally, from the ideation session also emerged a very interesting view on the sustainability issues. When one talks about smart-everything (cities, automotive, things, and so on), we must also look into our own production processes and make them "smart" too, in a way that makes them consume less energy, use fewer resources, reuse more materials and produce zero or close to zero residues. If we think about these processes with this mindset, digital products and services will for sure be more sustainable.

The other side

Digital services and products are currently one of the most ubiquitous aspects of western societies. One can find them in some day-by-day activities like e-commerce, on basic infrastructural services like e-mail and messaging, on social interaction like social networks and retail stores but also on very important aspects of healthcare management, justice or even politics. Each one of this kind of use brings different challenges, but they all have something in common - digital services are now one of the most influential aspects of human life. Digital services can now, like never before, influence how humans look and interact with each other, what and how they buy, how they communicate between them or even how they choose their leaders.

Another important aspect of this ubiquitous use of the digital is the amount of data that is collected. Estimates are that *"(...) the total amount of data in the world was 4.4 zettabytes in 2013. That is set to rise steeply to 44 zettabytes by 2020 (...)"* [15]. This data can have multiple usages and may lead to very different analysis and interpretations from the ones performed before.

These two aspects can get together in two different approaches - one that sees it as an opportunity and another, rather darker, that sees it as a threat. Probably they need to be looked at as a mix of both.

Society needs to understand some threats that are currently developing. In politics, based on personal data and computational power, there is now a more detailed view on how each individual can be influenced to make a given choice. Even worse, with this kind of knowledge, the very thin line between influence and manipulation can be crossed. In healthcare, people can be explored because of their own habits, like for example, having higher insurance premiums because of not doing enough exercise or not having a healthy diet, information that - we all know - is now widely available. Even in the military domain, there is now a larger arsenal of cyberweapons that can and ultimately will be used in geostrategic disputes. These weapons, once considered highly reserved, are now available to a large number of people and can therefore be used without any control, causing devastating effects.

Therefore, as mentioned before, the digital services amazing growth can be used to make humankind progress, but it can also lead to injustice, to manipulation and even to attacks against the human nature, and so, it needs to be monitored and consciously used by everyone. And this is a key aspect of everyone's education - people need to learn to be vigilant about their data, their digital footprint and their digital rights. If not, people's rights can be restricted, or worse, manipulated through what could be called a state of digital dictatorship!

Wrapping up

With this canvas in mind, and more than doing a wrap up of everything said in the above lines, we think it is important to point out that 2020 is not at a long distance away but is rather right next to us. Some things in our society have already started to change while others are starting to impact several social aspects. Thus, digital providers need to acknowledge this shift and, as soon as possible, start to walk the walk towards providing valuable services to their users (both companies and consumers) which, as we saw, are demanding for a greater sense of customized individuality.

This is not the time for just strategic planning. This is the time for addressing strategy by defining a tactical plan. Let us engage digitally with 2020, let us act!

References

[1] World Economic Forum, "Digital Transformation Initiative," World Economic Forum, Geneva, 2017.

[2] D. Plummer, M. Reynolds, C. Golvin, et al., "Top Strategic Predictions for 2017 and Beyond: Surviving the Storm Winds of Digital Disruption," Gartner, Stamford, 2016.

[3] Walker Information, "Customers 2020 - The future of B-to-B customer experience," Walker Information, Indianapolis, 2013.

[4] D. Sovie, J. Curran, M. Schoelwer and A. Björnsjö, "Dynamic Digital Consumers - Ever-changing expectations and technology intrigue," Accenture, Dublin, 2017.

[5] A. N. D. d. A. Aguiar, Interviewee, *The impacts of Telco, IT, Contents convergence and digital transformation in society by 2020 - Companies in 2020*. [Interview]. 16 February 2017.

[6] D. Newman, "Top 10 Trends For Digital Transformation In 2017," Forbes Media, Inc., 30 August 2016. [Online]. Available: https://www.forbes.com/sites/danielnewman/2016/08/30/top-10-trends-for-digital-transformation-in-2017.

[7] European Comission, "Digital Single Market: Coordination of European, national & regional initiatives," European Comission, 29 May 2017. [Online]. Available: https://ec.europa.eu/digital-single-market/en/cordination-european-national-regional-initiatives.

[8] B. Marr, "What Everyone Must Know About Industry 4.0," Forbes Media, Inc., 20 June 2016. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2016/06/20/what-everyone-must-know-about-industry-4-0.

[9] "Industry 4.0 - Challenges and solutions for the digital transformation and use of exponential technologies," Deloitte, Basel, 2015.

[10] OCDE/European Communities, Oslo Manual - Guidelines for Collecting and Interpreting Innovation Data, 3rd edition, Paris: OECD, Eurostat, 2005, p. 162.

[11] IBM Institute for Business Value, "Digital Transformation - Creating new business models where digital meets physical," IBM Global Services, NY, 2011.

[12] S. Hartley, "Digital Economy 2025: The World of 2025," Ovum, London, 2015.

[13] European Comission, "Reform of EU data protection rules" European Comission, 31 December 2016. [Online]. Available: http://ec.europa.eu/justice/data-protection/reform/index_en.htm.

[14] V. Tripathi and A. Gupta, "Hype Cycle for Sustainability, 2016," Gartner, Inc., Stamford, 2016.

[15] EMC Digital Universe; Research & Analysis by IDC, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," IDC, London, 2014.


Smart Hom
Lighting
Camera

02

Smart living powered by Digital Service Providers

Filipe Cabral Pinto, Altice Labs
filipe-c-pinto@alticelabs.com

Fernando Santiago, Altice Labs
fernando-m-santiago@alticelabs.com

Isabel Borges, Altice Labs
isabel@alticelabs.com

Bruno Parreira, Withus

Paulo Pereira, Altice Labs
paulo-s-pereira@alticelabs.com

Mário Rui Costa, Altice Labs
mcosta@alticelabs.com

In this digital age, the society evolves faster than ever! Technology creates new behaviours and customers reflect that evolution; they absorb the trends and demand answers to their needs. CSP have to be completely committed to new models of smart living and the best way to do so is by becoming a digital service provider. The goal is to maximize customer satisfaction by offering a new generation of digital services, affordable and convenient to customers, in the form of an addictive and simple journey.

Keywords

Smart Living; Customer Centric; Communication Service Provider; Digital Service Provider; Trends

Introduction

The digital revolution is having a tremendous impact on the creation of a new society. People now live surrounded by cutting-edge technology in a constantly changing world. The quick and widespread access to information is shaping people, making them more demanding. There is a new way of life on the way and there is no escape from it.

Customers are already reflecting these changes. They are now much more informed knowing quite well what they want. Their standards are higher than ever before, requiring first class treatment. What matters now is the experience as a whole during their journey with a supplier, and loyalty to service providers is a thing of the past. In addition to a relationship of excellence, customers also demand services tailored to their new lifestyle. They are expecting a myriad of integrated services supporting their daily personal and professional lives – I need to communicate, I want to have access to relevant content, I want to be permanently connected to my objects, I need to manage my living spaces, I need to control my own business, always in a private and secure way. And they will expect to have those services on the go, always accessible, and for a convenient price. Customers now require all the attention of the service providers and they want them to support their new way of life.

Communications Service Providers (CSP) have no alternative; they have to change their mindset to support new smart living models required by increasingly digital customers. CSP digital transformation initiatives must put customers at the centre of operations while providing them with the full support for their new way of life. To engage the new generation of customers, CSP have to reposition their business and reshape operations to offer an experience of excellence addressing a brand new emergent digital services industry.

This article presents a reflection on the demands of future customers in a digital society and their impact on the transformation of CSP strategy to answer the new smart living requirements.

A new customer journey

The new generation of customers relates much more dynamically to operators than past generations. For years, the relationship with the CSP has remained unchanged. The commercial offer was much more limited by technology than today and the concern with customer satisfaction wasn't the main priority at all.

Digital transformation is all about the requirements of this new society. Investment in technology is being made to streamline processes and to leverage new business opportunities. New experiences come about radically changing the way people interact, either personally or professionally. CSP customers started to demand much more than simple voice calls, they want ubiquitous connectivity at suitable prices to access their online services. The competition made customers even more demanding, making loyalty a scarce commodity. Moreover, OTT providers have come to market, offering quality services in exchange for personalized advertising, turning CSP into a simple bit pipe provider. It has never been so easy to change service provider and that's what unsatisfied customers will do if the following three dimensions are not met: price, service offer and experience – see **Figure 1**.

Reducing prices is perhaps the most immediate way to captivate new customers and keep current ones on board. Nevertheless, these benefits typically cling to loyalty contracts, forcing customers to something that they are running away from: maintaining the relationship for a preset period of time. Changing pricing does not, by itself, differentiate service or emotionally engage the customer with the operator's brand. Pricing alone can be seen as a means to achieve something, but not an end in itself. And it is well known that the customer runs away as soon as he receives a better overall proposal from another provider.

FIGURE 1 – Customer satisfaction main dimensions

The offer of new value-added proposals can effectively make the difference. Customers require new services and packing them together seems like a good bet. The “All IPzation” of networks and media services have been a generous lifeline for the core CSP’s business. The Multi-Play offers are on the ground making available seamless access to services at home, or on the go, having proved to be useful as well for the Business-to-business (B2B) market. Internet access, voice communications, Internet Protocol Television (IPTV) services and even home and office automation are now together under a single bill. These integrated offerings attract less cost-sensitive segments, allowing the operator to mitigate the effect of the decrease in revenues in its traditional services. In fact, the Internet OTT players have eaten a large chunk of the operator’s business. Skype is responsible for more than a third of the international voice traffic while WhatsApp, Viber, and Apple’s iMessage applications account for more than 80% of all messaging traffic [1]. Even without having a distribution network or even a private data centre, Netflix is reaching the impressive mark of 100 million streaming subscribers. All these figures force CSP to clearly redefine their portfolio strategy by thinking about future customers’ needs, new business dimensions, the way relationships will have to be maintained and nourished, and the way they have to evolve to make the difference.

Differentiation comes also through a relationship of excellence. Customer experience must be pleasant in the relationship with the CSP. Omnichannel access will have to be a reality enabling ubiquitous and personalised interactions. The quality of service must be unquestionable, tailored to that customer, as if he/she were unique. The processes should be simplified and adapted to the customer context, creating a pleasant customer journey, not neglecting any point of contact, from the contract start, until the contract cessation moments. **Figure 2** presents the key interactions taken during the relationship with the provider, taking into account the purchase decision, the service usage and support given. Only positive experiences, along with all the steps, engage clients, making them stay and not slam the door and walk away.

FIGURE 2 – Main customer journey steps

Towards smart living

The customer journey concept goes hand in hand with new trends. The customer needs are changing in alignment with societal and technological evolution. And CSP have to anticipate these new lifestyles.

Societal trends

Digital revolution has brought intense changes in the way people live and do businesses. The society has undergone deep modifications, having a direct impact on the creation of new consumption behaviours. These changes are also affecting CSP customers. They now have new standards and are much more demanding. **Figure 3** presents the societal evolution of CSP customers taking into account the smart living trends.

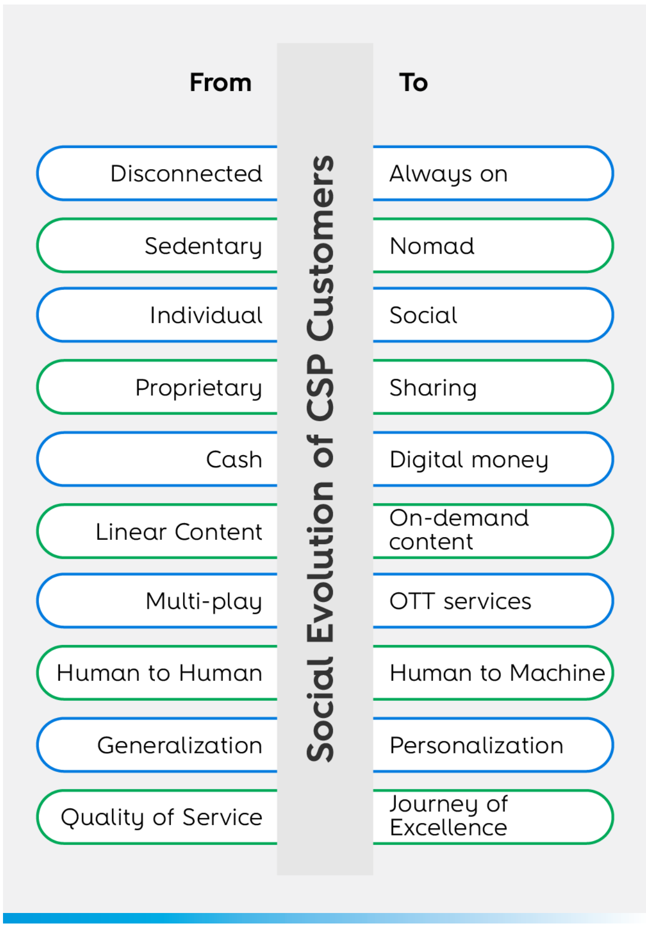


FIGURE 3 – CSP customers' societal evolution

- Connectivity is by now a daily life asset. Customers are already "always on", but they are asking for more bandwidth, full geographic coverage and suitable quality to take full advantage of the new digital trends;
- People are no longer in the same place most of the time. They now move easily, change frequently from city, and even from country, and no longer have a fixed residence. This is also true or even more relevant for businesses. Nomadism is back!
- People and businesses are much more social. They share everything and anything else on the network. They make declarations of love, post harsh comments, show what they are about to eat! They add an endless number of photos, exhibit in small videos their children's skills and introduce their pets to the community. They post everything but almost always hidden behind a computer! If you have a business, this is exactly what your customers are doing with their experience of using your services. You must be prepared as well;
- The digital sharing economy has come to stay. People pick up an Uber to move around, use Airbnb to rent a house for the holidays. They share objects with neighborhood in Streetbank and can still leave their pets in a sitter's home through the DogVacay site. Owning is no longer a goal. The key objective is to meet its own needs in a rational and efficient manner;
- Safety and convenience are two of the main factors that potentiate the Customer migration to cashless. Smartphone may be at the centre of this shift and mobile payments are one of the trends that promise to answer to these requirements;
- Watching TV is ever more a grandparent's behaviour. Being waiting for an episode is something that today does not make sense at all. Even worse is waiting in front of the television for the new season of "Game of Thrones". Increasingly, people will demand access to content and not programs and will

want to see it in their different devices, when they want, wherever they are;

- CSP are converging to deliver multi-play bundles with mobile and fixed telephony, broadband Internet, streaming TV and video-on-demand (VOD) services. In the future the differentiation will be mainly achieved through the content and with new additional services, including OTT services. Customer needs, preferences and behaviours will drive the composition of enhanced service packages;

- Human interaction with machines has been gaining ground. Chatting with bots is now common, and talking with intelligent personal assistants, like Alexa from Amazon, Apple Siri or Google Assistant, no longer impresses anyone. The control of spaces becomes also mandatory; the use of machines will guarantee the personalization of the different environments: the personal ones and those of the dependents such as children. The future brings a certainty: intelligent machines will transform the way society interacts;

- Personalization is a key requirement of future customers. They want to have the full power over their services, the content they access, the way they communicate. They negotiate what they are going to pay and want a personal relationship with their service provider;

- The perception and the feelings of customers, through the diverse digital and physical channels, about the services provided by companies are, in this digital age, a paramount factor for business success. Although they continue to ask for quality service, customers now require a relationship of excellence throughout their journey, taking into consideration the multiple touch points over the time.

Technological trends

There are new key technologies that directly impact on CSP future evolution which are described below:

- 5G is a paradigm shift integrating relevant technological changes and paving the way for a really flexible and convergent network with huge capacities, low latencies, high-speed mobility leveraging a vast potential for new commercial offers. Fuelled by consumer and business demand, governments, manufacturers and CSP are pushing the deployment of 5G technology that will revolutionize the way customers communicate, socialize, work and live. According to GSMA [2], 5G will enable 1.1 billion connections by 2025. The key objective is to facilitate the creation of a connected society where things, people and processes share data in a seamless way according to their momentary needs. The main scenarios are related to the evolution of mobile broadband connections, supporting massive machine type communications and ensuring critical systems. Long-term predictions point to a $12.3 trillion of global economic output by 2035 [3];

- Under the 5G umbrella, new business models taking advantage of Cloud, Software Defined Networks (SDN), Network Function Virtualization (NFV) and Self-Organized Networks will lead to network automation with global orchestration capabilities. From the operator perspective, the new paradigms enable better resource optimization, shorter time to market of new services and self-management customer portals promoting opex reductions. Moreover, Multi-access Edge Computing (MEC) will enable CSP to explore a more important role in content distribution by moving content closer to the customer potentiating new applications and new business opportunities. On the customer side, moving from a hardware-based capex model to a software-based opex service subscription model will benefit cost reductions and on-demand customization of services, with this aspect taking particularly relevance on the B2B market;

- Cyber & Physical Security is a key topic increasing its dimension. Companies and governments are becoming increasingly concerned about privacy and security issues with the growth of global scale cyber attacks. And the numbers reinforce that concern: cybercrime damage costs to hit $6 trillion annually by 2021 [4]; Hackers have available tools that may cause global panic due to the sophistication on the networks penetration level, personal data thefts, and spread of an unnumbered list of intelligent computer viruses. With the emergence of the Internet-of-Things (IoT), the risks ceased to be digital only and expanded to the physical side of the internet. One of the main characteristics of cyber and physical security is the continuously evolving nature of the risks. This requires new transversal approaches to react dynamically and in an adaptive way to the different attacks. To enable rapid response, the industry needs to embrace new disciplines to proactively detect abnormalities that can result from complex attacks, and use those assets to provide it as a service to their customers, particularly on the B2B domain;

- Internet of Everything (IoE) empowers the Internet with connected Things. The data gathered by sensors and the ability to act in the surrounding context opens doors to new logics that foster innovation and value creation. The materialization of these developments will have tremendous implications for the creation of new businesses, anticipating a multi-billion dollar market in the coming years. According to Gartner [5], 20.8 billion connected things will be in use by 2020. Moreover, International Data Corporation (IDC) predicts global Internet of Things spending to reach nearly $1.29 trillion in 2020 [6]. There will be direct impacts in the most diverse areas, such as health and wellbeing, home automation, agriculture and environment, industry 4.0, mobility or smart cities. CSP have here a strong possibility to offer new services answering future customer needs. Customers look for new life models where they control not only their communications, services and content but also their living spaces to ensure the most appropriate environment at all times;

- More than ever, data plays a central role in mankind. Big data is a major asset made of large quantities of data, coming from different sources, requiring management and processing to enable the extraction of useful insights for business. According to IBM, "*by 2016, 90% of the world 's data had been created in the last 12 months.*"[7], and forecasts point to even more growth; it is expected the production of 40 Zettabytes of data by 2020 [8]! The size of the market is also growing and is expected to surpass $88 billion by 2025 [9]. Big data must deal with "four Vs", refering to four relevant perspectives: volume (the amount of data); variety (the forms of data); velocity (the speed of incoming data) and veracity (the quality of data). Artificial Intelligence techniques can then be applied, with particular emphasis on machine learning methodologies, to understand the surroundings and help in decision making. Well known use cases include physical and cyber security, marketing and advertising, conversational systems, retail analysis, healthcare and financial trading;

- Augmented Reality and Virtual Reality are two hot topics. They both take the user to a different dimension. The extension of reality with related information or the creation of new virtual environments enables new forms of interaction and allows a new level of realism. A new generation of applications is emerging offering fantastic experiences to customers, making it difficult to distinguish between what is real and what is virtually added. Some sectors are already taking advantage of these evolutions: in the field of entertainment, the Pokemon Go game is a paradigmatic case of extreme success; a set of applications in the area of tourism have emerged that greatly enrich the traveller's experience by adding important information regarding its context; the education sector is also making use of these technologies to improve the learning process and make it more appealing; and, naturally, the military is developing new programs in order to expand its forms of action;

- Conversational systems are gaining relevance on several areas improving the experience level of people interacting with digital assets. Chatbots, voice-activated assistants, virtual personal assistants, promise to revolutionize our lives in multiple domains. For example, Gartner predicts that *"by 2020, 30% of web browsing sessions will be done without a screen"* [10], and according to Business Insider, *"80% of businesses want chatbots by 2020"* [11]. This new trend results from the combination of several technologies, such as artificial intelligence, machine learning, natural language processing and IoT, among others, that are hitting a maturity level that allows having a more intelligent, efficient and natural interaction between human and computer. The proliferation of these technologies promises to have a transformative impact in the way the customers interact with enterprises, enabling a more seamless experience and channel diversification on the customer journey.

The Digital Service Provider for new lifestyle

In this digital age, the society evolves faster than ever! Technology creates new behaviours and customers reflect the evolution of this new society; they absorb the trends and demand answers to their needs as if there was no tomorrow. CSP have to be completely committed to new models of smart living in which customers want much more than low-cost connectivity.

The CSP path is drawn and points in the direction of the digital service providers (DSP). The goal is to maximize customer satisfaction by offering a new generation of digital services, affordable and convenient to customers, in the form of an addictive and simple journey.

Customer demand is forcing the transformation of the technology-centric paradigm into a customer-focused approach to increase customer satisfaction. DSP must be focused on providing personalized interactions based on enhanced customer knowledge. Data and analytics will be central to intelligent process automation impacting operational and business efficiency. DSP will become a service broker that enables self and third-party OTT services reinforcing the relationship with customers through the competitive advantage of owning assets that can positively contribute to the service usage experience.

As **Figure 4** illustrates, the strategy to evolve from a CSP paradigm towards a DSP, demands key changes, namely:

- Enhance customer knowledge level by building a strong relationship with him in all their lifecycle, keeping the customer at the heart of the business by adopting intelligent customer relationship management systems and taking advantage of big data and artificial intelligence;
- Increase customer engagement by investing in omnichannel interactions using analytics, machine learning technologies and conversational systems;
- Evolve business and operational processes towards the full automation improving customer response times and decreasing incidents with customer services by enhancing online multichannel self-provisioning systems;
- Provide services supported by new technological trends driven by intelligence to enhance customer average revenue per user/unit (ARPU), diversifying the sources of revenue.

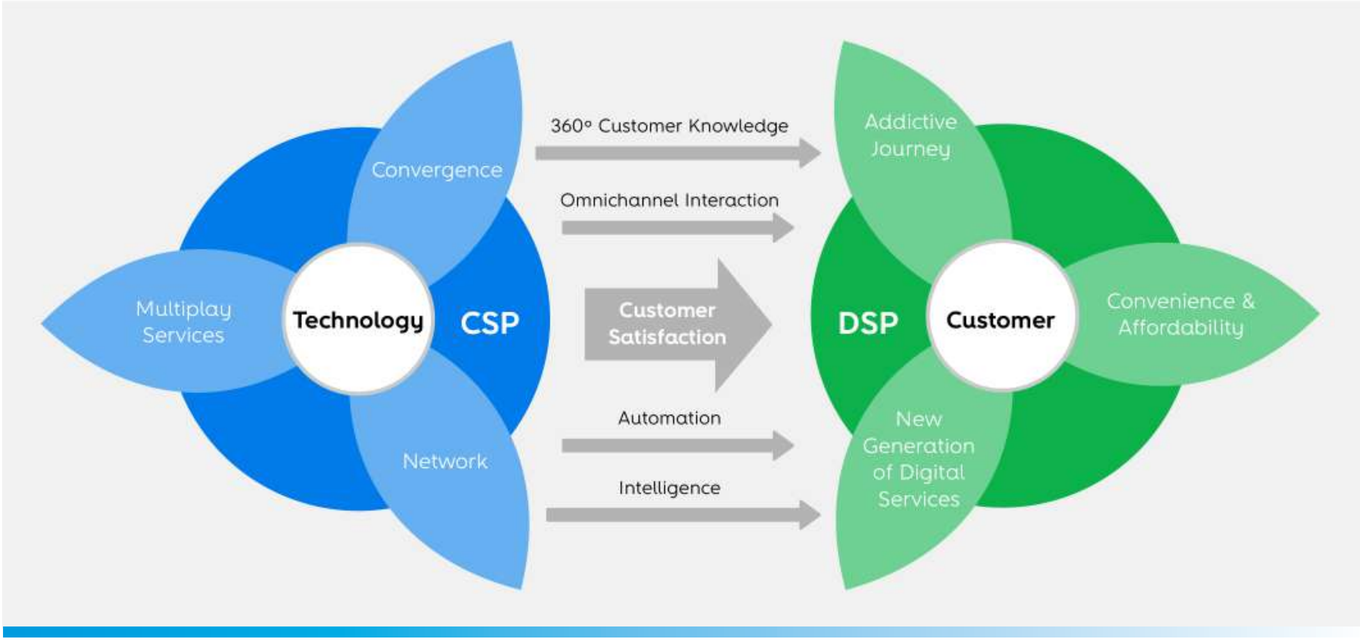


FIGURE 4 – Evolution towards a customer-centric DSP

Smart living and the smart home use case

Smart living is a lifestyle where people have full control of their physical and virtual interactions, either individually or socially, in an easy and personalized way. As seen in **Figure 5**, it is the daily trip crossing different smart domains in a new digital life.

Among the various domains, there is one of particular interest for future DSP: smart home is a bet that has to be made and won to keep DSP competitive in the smart living market.

A smart home is a living space where people are able to digitally interact with different home equipment and services, making life easier. The increasing use of IoT technology in household products is making possible to evolve the smart home proposition. Common offers include now: home security monitoring to keep homes safe; home automation to manage lighting, heating, ventilation, air conditioning or blinds; ambient assisted living comprising care services to ensure the safety and health of everyone in the household; monitoring energy, water and gas consumptions to save money and promote sustainable usage of resources; or even interaction with home appliances, such as dishwashers, clothes machines, toasters, microwaves, among other devices. But, for sure, the great success these days is the intelligent personal assistant that enable voice interactions, being able to completely interact either with people or connected systems towards a new way of living.

The smart home market is full of players eager to get a generous slice of the pie. In fact, pure digital and industrial-based companies such as Google, Amazon, Apple or Siemens are investing hard also in this domain [12]. In addition to the typical telco and media services that have been available, these players are now extending their presence at home, offering disruptive services to customers by letting them control a set of household products.

DSP are under (even more) pressure! They cannot lose the battle of smart homes, but the permanent and strong competition, with several fronts of attack, is forcing them to act

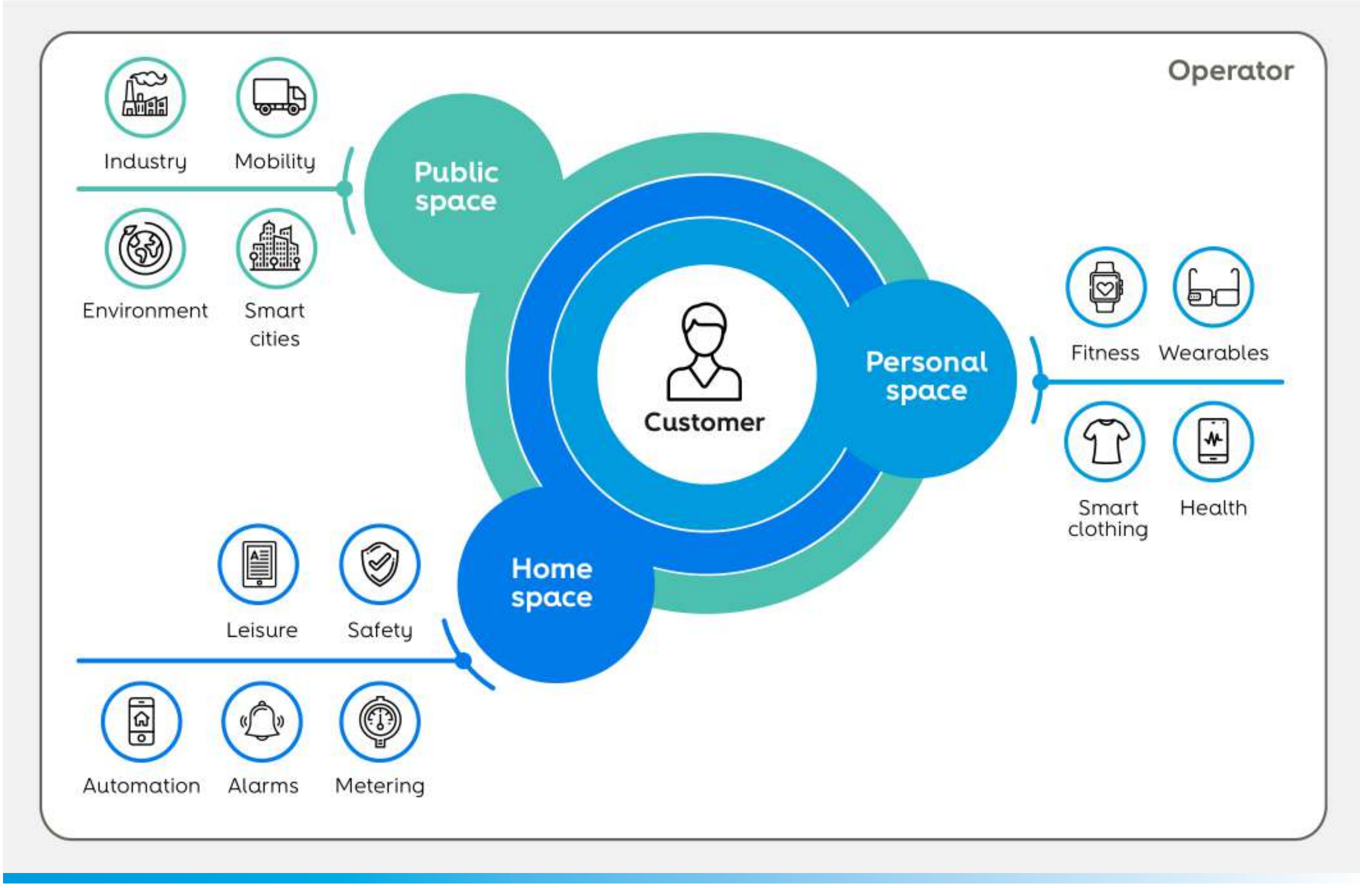


FIGURE 5 – Smart living interactions

fast. However, DSP have reasons to believe they can evolve to have a key role here since they already hold a set of goods essential to capture the value of this market. DSP control the entire customer relationship cycle, from sales to technical support, and own a big customer base in the home space that can be capitalized on. The existing relationship of trust allows them to be already in millions of houses, in which they installed home gateways providing voice, internet and TV services. But despite these competitive advantages, the DSP will have to adjust to the demands of its customers as a way to win this battle. They must provide a satisfactory journey regardless of the particular segment.

The wide variety of the market naturally makes this task difficult, bringing a set of associated problems derived from the great fragmentation. There is indeed a huge diversity of services in the home domain, provided by various entities, but typically they lack a common look and feel, forcing the customer to constantly settle into different environments. The large number of service providers requires different accounts and multiple invoices, and the contact points are dispersed and can only respond to part of the problems, not having a complete view of the customer. The processes, even when they are automatic, have their own logic, dependent on the company that manages them.

The smart home market requires a common interface able to put customers at the centre of operations, while offering a plethora of user-friendly digital services to interact with in-house environments, from media contents to home appliances. Future DSP are the right stakeholders to take this role due to its native presence in customer residences, but also for its ability to create partnerships with leading players in the creation of new commercial offers covering different home segments. Following the digital trends, DSP can provide open platforms to promote smart home offers, ensuring, besides connectivity, intelligent services with seamless interaction with physical and virtual assets in a full personalized environment. A truly interoperable platform can link a trusted ecosystem of partners leveraging the promotion of creativity and fostering innovation bringing added value to customers.

Wrapping up

Digitalization is a no-return path. Technological developments coupled with societal changes are creating a new lifestyle where information is vital to people's life. Changing behaviours is affecting the way people consume, making customers more demanding, in their daily dynamics. They now require seamless interactions with service providers and ask for personalized services bringing clear added value.

Smart living lifestyle is about people having full control of their physical and virtual interactions in an easy and personalized way, and CSP have a significant opportunity to have a decisive role on making it happen in a very important living space: the smart home, where areas such as IoT, security, monitoring, automation, ambient assisted living, interaction with appliances and personal assistants, open opportunities to explore and to bundle as attractive applications and services for the customers.

CSP need to evolve to answer customer requirements, placing them at the centre of their operations, while becoming agile in innovative digital services delivery. The engagement of customers must happen during the whole journey, at all touch points, to create a trusty relationship. CSP transformation towards DSP must happen now in order for them to become relevant players on the definition of our forthcoming digital lives in this connected society. This is the future, so be it!

References

[1] B. el-Darwiche, P. Péladeau, C. Rupp, and F. Groene, “2017 Telecommunications Trends,” 2017. [Online]. Available: https://www.strategyand.pwc.com/trend/2017-telecommunications-industry-trends.

[2] E. Obiodu and M. Giles, “5G era: age of boundless connectivity and intelligent automation,” 2017.

[3] IHS Markit, “The 5G Economy - How 5G technology will contribute to the global economy | IHS Markit.” [Online]. Available: https://www.ihs.com/Info/0117/5g-technology-global-economy.html.

[4] S. Morgan, “Cybercrime Damage Costs $6 Trillion in 2021, Cybersecurity Market DataCybersecurity Ventures,” 2017. [Online]. Available: https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/.

[5] Gartner, “Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015,” 2015. [Online]. Available: https://www.gartner.com/newsroom/id/3165317.

[6] M. Torchia and M. Shirer, “Internet of Things Spending Forecast to Grow 17.9% in 2016 Led by Manufacturing, Transportation, and Utilities Investments, According to New IDC Spending Guide,” 2017. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS42209117.

[7] M. Winans *et al.*, “10 Key Marketing Trends for 2017 Customer Expectations,” 2016.

[8] IBM, “Infographic: The Four V’s of Big Data | IBM Big Data & Analytics Hub.” [Online]. Available: http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

[9] Statista, “Global Big Data market size 2011-2026,” 2016. [Online]. Available: https://www.statista.com/statistics/254266/global-big-data-market-forecast/.

[10] H. P. Levy, “Gartner Predicts a Virtual World of Exponential Change - Smarter With Gartner,” 2016. [Online]. Available: https://www.gartner.com/smarterwithgartner/gartner-predicts-a-virtual-world-of-exponential-change/.

[11] Business Insider, “80% of businesses want chatbots by 2020 - Business Insider,” 2016. [Online]. Available: http://www.businessinsider.com/80-of-businesses-want-chatbots-by-2020-2016-12.

[12] M. Philpott, “2018 Trends to Watch : Smart Home Preparing for the mass market,” 2018.

03

Smart cities - the importance of sharing

João Bastos, Altice Labs
bastos@alticelabs.com

Telma Mota, Altice Labs
telma@alticelabs.com

Ricardo Ferreira, Altice Labs
ricardo-j-ferreira@alticelabs.com

While the demographic reorganization of the society continues, leading to the increase in the number of large urban concentrations around the world, new models of work organization and wealth creation are emerging, based on the concept of sharing and collaboration, which deserve to be investigated. In this context, national and European funded programmes are endorsing a number of collaborative projects, involving companies from several countries, aiming to research and innovate on new forms of collaboration in smart cities and sharing economy ecosystems. This is the case of H2020 Sharing Cities project, mentioned in this article.

Keywords

Smart City; Human-centric; Sharing Economy; Business Models

Introduction

The new digital reality into which we are being pushed every day is shaping and rapidly changing the rules that have been considered for centuries as the basis of societal, political, cultural and economic relationships in society. Although those relations were not likely to be immutable or built upon static rules per se, the broad and massive introduction of information and communication technologies (ICT), encircling so many areas of people's lives – from leisure to work, to healthcare or education – has brought with it the need for people to adapt to a new "way of doing" things, while at the same time it opened up new opportunities for society to move on towards the elimination of geographical barriers and the reduction of economic gaps between countries and communities in particular. Together, technology advances and data availability are reshaping long-time established social and economic relations among people but also between people and companies. Moreover, they are transforming the way cities are governed, both at the operational and strategic planning levels.

The strategy for smarter cities, as it is being commonly understood by municipalities worldwide, should promote a more open, inclusive and sustainable living in urban concentrations. In this sense, the search for more efficient solutions, for old and new city problems, aiming for better quality of life for their citizens should be centred in the creation of the necessary conditions for the emergence of innovative ideas, enabled by technology. The combination of high-speed communications, technological advances of all kinds and an innovation policy/culture, driven by the citizens' needs, is of paramount importance.

Yet, the smart city concept cannot be described or understood separately from other realities like social, demographic, economic or even political trends, that jointly are shaping the transition towards an economy based on resource sharing – the sharing economy concept. Both concepts aspire to improve peoples' lives, relying on technology for placing people and companies "face-to-face" while leaving to the user the opportunity to co-create, develop and share, in a wider and less regulated market. At the centre of this dynamic ecosystem, where technology seems to play the main role, are the citizens and the improvement of their quality of life must be the focal point of every initiative.

While the demographic reorganization of the society continues, leading to the increase in the number of large urban concentrations around the world, new models of work organization and wealth creation are emerging, based on the concept of sharing and collaboration, which deserve to be investigated. In this context, national and european funded programmes are endorsing a number of collaborative projects, involving companies from several countries, aiming to research and innovate on new forms of collaboration in smart cities and sharing economy ecosystems. This is the case of H2020 Sharing Cities project, mentioned in this article.

Societies are not the same anymore

World population

According to the United Nations habitat report [1], cities account for more than 50 percent of the world's population, consuming between 60 and 80 percent of energy and generating as much as 70 percent of the human-induced greenhouse gas emissions, primarily through the consumption of fossil fuels for energy supply and transportation.

The same report [1] indicates that, in the next ten to fifteen years, five thousand million people will be concentrated in cities while projections indicate that, by that time, urban population in developing countries will double, while the area covered by cites will triple. United Nations also reveals, in a report regarding the state of the cities [2], that in 2016 an estimated 54.5% of the world's population lived in urban settlements while one in each five –

	2016			2030		
	Number of settlements	Population (millions)	Percentage of world population	Number of settlements	Population (millions)	Percentage of world population
Urban	..	4034	54.5	..	5058	60.0
10 million or more	31	500	6.8	41	730	8.7
5 to 10 million	45	308	4.2	63	434	5.2
1 to 5 million	436	861	11.6	558	1128	13.4
500 000 to 1 million	551	380	5.1	731	509	6.0
Fewer than 500 000	..	1985	26.8	..	2257	26.8
Rural	..	3371	45.5	..	3367	40.0

TABLE 1 – World's population by size class of settlement, 2016 and 2030 [2]

approximately 1700 million people – lived in a city with at least 1 million inhabitants - see **Table 1**. By 2030, urban areas are projected to house 60% of people globally and one in every three persons will live in cities with at least half a million inhabitants.

By the mid of the century, the world's population should grow to about 9.7 thousand million people – see **Figure 1**. This means that between now and 2050 the world population will rise by as many people as those who lived on the planet in 1950 [3].

Although municipalities and urban regions' governments have long time realized the multiple impacts of the growing concentration of people in large metropolises, it is now more than ever

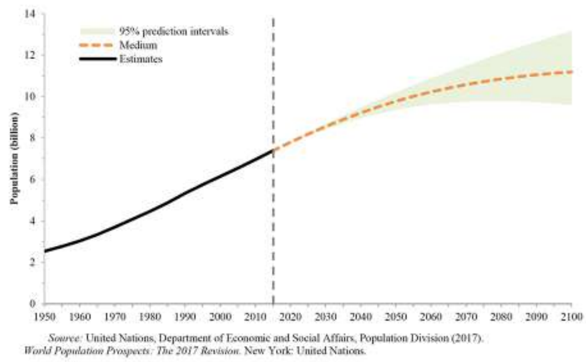


FIGURE 1 – Population of the world: estimates, 1950-2015 and medium-variant projection with 95% prediction intervals, 2015-2100 [3]

necessary that municipalities seek new and more efficient ways of managing the urban space and commit to organize effectively the lives of its citizens. New models of living and working are needed, built around integrated solutions that are open by design, affordable, scalable and offer a generic and replicable solution to the problems that are common to most cities, starting by connecting local government needs with citizen-centric value [4].

Digital inclusion

European Commission Europe's digital progress report 2017 indicates that 76% of European citizens now have access to fast broadband technologies (i.e. capable of delivering at least 30 Mbps download), although there is a significant unevenness between the most developed urban areas and rural areas, where fast broadband access is available only in 40 % of rural homes [5] – see **Figure 2**.

Europeans are using the Internet on a regular basis. In 2016, 79% of EU citizens went online at least weekly and 71% daily or almost (compared with, respectively, 76% and 67% a year earlier), whereas by contrast, only 17% of EU-28 SME and 38% of EU-28 large enterprises were actively selling online [5]. The same report also shows that the adoption of digital technologies varies strongly with company size, being that there

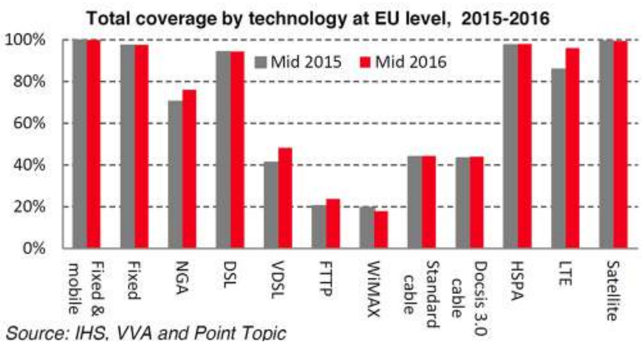


FIGURE 2 – EU coverage by technology (2015-2016) [5]

was an increase in Internet usage for business purposes by SME employees, from 20% in 2012 to 29% in 2016, while for large enterprises it remained stable at 64% over the same period. Still, there are a lot of opportunities to be explored, especially with respect to e-commerce, cloud computing or big data. As for EU-28 citizens, the report shows an overall progress on e-government use which allows us to conclude that users are starting to use more complex online services, which is a very good indicator of the success that smart cities' deployment and sharing economy models may achieve in the EU.

Smart cities

The smart city concept was initially built around the idea of spreading communication technology and sensors throughout the city and, later on, to create a "connected society" where new forms of interaction among citizens would emerge, fostering innovation and contributing to resources' usage optimization. Today, the smart city concept is commonly understood in a much more comprehensive way, seeking to facilitate the interaction between citizens and city administrations, to encourage the development of innovative solutions to urban life improvement – such as transportation, citizens' mobility, healthcare or energy efficiency improvement. A smart city must, therefore, be human-centric, open and non-discriminatory with regard to its citizens, an instrument used to promote openness, fairness, transparency and sustainability of the city while exploring the potential of technology to create new jobs and promote social inclusion.

As described by ISO "Working Definition", a smart city might be described as the one that "*dramatically increases the pace at which it improves its sustainability and resilience..., by fundamentally improving how it engages society, how it applies collaborative leadership methods, how it works across disciplines and city systems and how it uses data and integrated technologies,... in order to transform services and quality of life for those in and involved with the city (residents, businesses, visitors)*" [6].

The advent of smart cities opened the opportunity to introduce more fairness in the access to broadband communications to all segments of society, namely those deprived of it due to social, educational or economic reasons, reducing the digital divide and, not less important than that, has fostered innovation and quick deployment of ICT- enabled economic interactions among people and companies – the new sharing economy or, in a broader sense, collaborative economy. Both approaches consider exchanging of underused assets or services directly from the supplier to the user (without middlemen), although the latter assumes the existence of a distributed digital marketplace that matches the needs and the offers [7].

Yet, in order to enable a collaborative economy, it is necessary that an easy to access, multisystem capable, ubiquitous and trustable digital platform is available, allowing for anyone to access and share information and resources within a global marketplace of millions of connected people. In a sharing economy or collaborative environment, digital platforms and technology-based applications must be perceived as an enabler for the creation of a network of trust that makes the entire system run. Trust, ease of use and a solid sense of community are some of the factors to consider in pushing adoption of the sharing and collaborative economy forward.

Although there is still a long way to go before cities can be fully smart or be in the first row of the sharing and collaborative economy, city governments should take a deep attention to today's city needs and try to picture how to take the maximum benefit for their citizens and local companies within the framework of the new digital economy, in a connected society.

It's not about being smart – it's about being human-centric smart

No matter how you define the concept and in which context you define it, the idea of becoming a smart city comes together with the promise to make the city more efficient, more open, more technology aware and, in the end, the promise to improve the quality of life of its citizens.

The roadmap towards becoming a smart city may be seen as a three-stage process, matching the way how municipalities realize the power of technology and its contribution to a successful city development strategy, moving from technology driven, to city government driven and finally to a citizen-driven smart city [8].

- Stage one of this evolution roadmap – smart cities 1.0 – is characterized by the sometimes massive adoption of communication technology, without a proper analysis of the real needs and implications on citizen's quality of life. Such future-city visions, which arose in different geographies around the world in the beginning of this decade, have mainly been pushed by technology suppliers' argumentation, frequently missing out what should be its primordial objective – to simplify interactions between citizens and the city.

- Evolution to smart cities 2.0 becomes an objective for smart city's stakeholders when it becomes clear, especially for municipalities, that the so promised contribution for the quality of life of their citizens on the one hand, and the tremendous opportunity that the technology represented for shaping the future of the city on the other, were not being fully attained. The realization of the technology potential to shape city living, made city administrators take the lead in helping to determine the future of their city and what should be the role reserved for technology in the deployment of the smart city. Communication and information technologies should now be seen as enablers to improve quality of life of the citizens and visitors, and the municipalities would be responsible to identify the most promising opportunities for using ICT for that purpose. The European H2020 Sharing Cities project, described in chapter "Sharing models, interfaces and data" of this article and in which Altice Labs is involved, is a good example of this type of approach by potentiating technology and smartness to support the urban community on a sustainable basis.

- Next generation of smart cities – smart cities 3.0 – is the one that is seeking for the active engagement of their citizens in the shaping of the city and the urban community, by realizing not only the power of technology by itself but to foster new economic relations among people, resident or not, potentiating the capacity of their citizens to detect needs before the city administrators and use their innovative capacity to collaboratively work to find solutions and develop new technology-based services, based on individual or co-created models. Moreover, citizens' driven smart cities development models tends to be more socially inclusive by broadening the access to information and resources to a large number of people, allowing for local sharing activities to emerge.

Although technology emerges as the first thought when you talk about smart cities, to be smart a city should first and foremost to be a human-centric city, where technology should play the role of a valuable enabler to push the level of resources optimization and assets management, never losing the focus on its primary objective – to improve the lives of its citizens.

The sharing economy business ecosystem

The sharing economy concept and the sharing-based business models emerged in recent years as an ecological, sustainable, democratic, rational and management wise disruptive approach to the traditional perspective of doing business and owning assets. Much of this perception comes from the thought that in a sharing economy people have access to under-utilized resources from other people or companies at a small fraction of the cost of ownership. Thus sharing is perceived as constituting a win-win scenario for all actors involved, making sense for the consumer, the asset owners and the community alike, being it in either for-profit or non-profit business scenarios.

But sharing is not a new practice. Russel Belk [9] pointed out in 2014 that sharing is a phenomenon as old as humankind, while collaborative consumption and the sharing economy are phenomena born of the Internet age. He also stated that "*collaborative consumption is people coordinating the acquisition and distribution of a resource for a fee or other compensation*". Considering the definition of sharing economy proposed by Pablo Muñoz and Boyd Cohen yearly this year, "*sharing economy is a socio-economic system enabling an intermediated set of exchanges of goods and services between individuals and organizations which aim to increase efficiency and optimization of under-utilized resources in society*" [10].

These definitions direct one's thoughts to the sharing of assets, services and information among strangers, geographically distributed, supported on Internet access and digital service platforms, for the efficient temporary access and use of non-owned assets or related services, by opposition to the sharing and borrowing with one's trusted relatives, neighbours and friends, as in the old days. That is to say that "trust" is the paramount enabler of "sharing economy".

The old limited sharing and borrowing paradigm evolved into the sharing economy concept thanks to the Internet's digital service platforms that massively lowered the transaction costs for searching, contacting and commercially engaging with strangers. Without these digital service platforms, the transaction costs alone would render impossible any mass market ecosystem around collaborative consumption and sharing. In fact this evolution created new markets both for substitution as well as for complementation of the existing ones and altered the industry structure, namely by leading to increased possibilities for smaller business operations to function while, at the same time, leads to the creation of large business companies thriving on transaction costs [11].

New business models in a shared economy

In fact, in the past few years, several innovative business model experiments have emerged in the sharing economy business space, like crowd-funded business, cooperative-based business or alternative currency business, proving the vibrant and highly dynamic ecosystem around sharing, which will probably reshape the future digital business landscape and market regulation.

For the sharing economy business space to thrive some core elements are required to exist in the underlying business models, namely:

- A user friendly digital service platform, capable of delivering and monetizing shared base services with a high trust rating, readily accessible in a cost-effective way for both customers and resource owners to use;
- Under-utilized resources to be made available through the platform for potential customers to use and share in a pay-per-use model;
- The capacity for high engagement of actors leveraging the network effects, to generate positive externalities;
- Differentiated services and value propositions, leveraged on specific artefacts or specific data (big data, correlated data, geo-referenced data, spatiotemporal personal privacy-compliant data, and so forth).

If in the private sector and in a for-profit perspective, the above core elements can be the starting conditions for a successful business model design for the sharing economy business space. But at the public city level, the sharing economy concept implies additional dimensions to be considered in the process. City governments must consider non-profit, or at least cost-driven, measures in the business model design aiming to achieve a smarter city, thus diverse measures should be thought-out and orchestrated on the overall sharing economy city model, for areas such as smart e-governance, smart planning and forecast, smart economy, smart mobility, smart environment, smart building, smart tourism, smart people, smart living, smart citizen voluntarism, and so forth.

Furthermore, the sharing economy at the public city level implies that a considerable part – if not all - of the cost savings and of the additional value generated within this ecosystem should be passed along to citizens, in the form of lower taxes, lower fees and transaction costs of interacting with the government authorities, as well as promoting access to better and more effective public services. Therefore this public service perspective should also be considered as a core element in the business model design for the sharing economy at a public city level.

Sharing models, interfaces and data

A real smart city means, on one hand, a city with enhanced efficiency, sustainability, equitability, self-governability and, on the other hand, more automated, instrumented, connected, responsive and intelligent. In real smart cities their governments must make use of an integrated multidimensional digital services city sharing platform as the foundation on which to build their smart city upon, preferably based on integrated solutions that are open by design, affordable, able to evolve and scalable.

The H2020 European Project Sharing Cities [12] leverages this concept by bringing together six European cities. London, Lisbon and Milan are the so-called lighthouse cities, which have committed to cooperate with their fellow cities Bordeaux, Warsaw and Burgas. The project started in January 2016, with duration of 5 years. It is led by the Greater London Authority (GLA) and has the participation of 34 other partners. The Portuguese consortium is led by the Municipality of Lisbon (Câmara Municipal de Lisboa - CML) and includes a total of 8 partners: CML, EMEL, EDP, Lisboa E-NOVA, Instituto Superior Técnico, Reabilita, CEIIA and Altice Labs. This project has the following main motivations:

- **Scale-up:** by proving that properly designed smart city solutions, based on common needs, can be integrated into complex urban environments and increase in social, economic and environmental value;
- Make it **digital** and **data-driven:** applied to the improvement of existing solutions, to the design and running of a new city infrastructure, and to the development of new digital services whose main purpose is to help citizens make better choices around mobility and energy efficiency;
- **Open-up** and **accelerate the market:** to understand, develop and trial business, investment and government models, essential for the true aggregation and replication (through collaboration) of smart city solutions in cities of different sizes and maturities;
- **Share and collaborate** along with the **society:** to respond to increasing demand for participation; to enhance mechanisms for citizens' engagement; to improve local governments capacity for policy making and service delivery through collaboration and co-design; resulting in outcomes that are better for citizens, businesses and visitors.

The main goal of this project is to aggregate demand, standardise procurement and deliver common and replicable innovative models across the above mentioned cities. Focussed on piloting energy efficient districts by shift thinking irreversibly to the local renewable energy source and sustainable models of e-mobility, the scope also includes the exploitation of city data to maximum effect, expecting to successfully engage with citizens and foster local level innovation with the creation of new businesses and jobs.

To accomplish these goals the project identified **people, place** and **platform** as the main city dimensions which must come together to address properly the real city needs – see **Figure 3**.

People – Approaches, guidelines and tools are being selected to develop a deep understanding of society, and the means by which people actively participate in making their districts better places, through sharing services and delivering better outcomes. Currently, several tasks are being conducted close to the population to ease the engagement process and understand better what are the real population needs and aspirations. The initial set of use cases has been identified and represents mainly the municipality perspective and the data and services that will be offered to the population under the scope of public interest (e.g. public building energy management). A summarized list is provided in **Table 2**:

FIGURE 3 – Sharing cities vision [12]

Sustainable energy and smart mobility are core areas of interest to municipalities in many different perspectives. Each district has its particularities and monitoring their behaviour will help planning ahead more efficiently. Thus, public information hubs and end-user applications will be provided to better address the daily life of citizens in each city. It is important to deeply understand users' differences as well as their commonalities to promote sharing of experiences, leverage community sense and improve civic behaviour. Applications and services

Type	Services	Clients
Energy & Lighting	Sustainable building energy management and Demand-Side Response (DSR)	Municipalities/Users
	Public heating optimisation	Municipalities
	City energy planning and forecasting (EV/PV)	Municipalities
	Lamppost management	Municipalities
	Public hubs – WiFi, noise detection, crowd management, pollution control, traffic info, etc.	Users
Mobility & Parking	Fleet management (e-cars)	Municipalities /Companies
	Logistics (delivering - route optimization)	Companies
	Electrical vehicles sharing (e-cars and e-bikes)	Users
	Public mobility hubs – charging spots, parking, electrical vehicles	Users
	Optimize parking (cars and bikes)	Users
Retrofitting	Urban rehabilitation and building resource optimization	Users

TABLE 2 – Initial use case summary [12]

should work across cities from different countries and data should "follow" the users wherever they go to ensure a seamless experience and prove the value of the proposed solution.

Place – Also in progress is the selection of technologies to properly address city infrastructures and sensoring; to support low energy districts, smart lampposts, sustainable buildings and cities, electrification of mobility and fair managing of fleets, parks, etc.

In parallel with the work conducted close to the population, sensors and intelligent meters are currently being placed (or replaced) in buildings, lampposts, vehicles, parking places, etc. to gather the proper data from the city (**place**). **Table 3** summarizes the number of resources foreseen in each city (Greenwich – London, Downtown – Lisbon and Porta Romano – Milan):

Measure	Greenwich	Downtown	Porta Romano
Citizen Engagement	Collaborative activities to catalogue current practices across cities, enhance, and exploit. Co-creation of sharing services for relevant measures with active citizen participation. Alignament of this sharing layer with the urban platform. Development of a district bond scheme to incentivise positive behaviour change		
Building Retrofit	357 public homes, 5 buildings w/PV & RSHP	1 public housing block, 2 private homes, 1 public building, PV	300 private housing homes, 5 mixed owner blocks, PV
Sustainable Energy Management Serv.	Real-time demand response. Virtual energy plant. Smart heating management	Optimised building consumption, micro-generation, consumption/production mal, adaptive street lighting, traffic monitoring	Real-time demand response and energy optimisation, and micro-geograhical information and visualisation
Shared e-mobility			
e-vehicles Car Sharing	8 autonomous shared e-vehicles 20 e-vehicles	62 e-vehicles	62 e-vehicles
e-bike Sharing	10 e-bikes for council staff 30 e-bike scheme / 4 stations	30 e-bikes and 2 dock stations	150 e-bikes / 14 stations
e-vehicles Charging	20 charging points (4 rapid)	29 public charging points (1 rapid); 5 private	60 charging points (20 rapid)
Smart Parking	300 parking places	30 parking bays	125 parking bays
e-logistics	Trial of 4 shared e-vehicles	80 e-logistics vehicles	10 e-logistics vehicles
Smart Lamp Posts	400 lampposts	250 lampposts	300 lampposts
Urban Sharing Platform	Development of a common open standard reference architecture and urban sharing platform for exploitation beyond the core cities *"designed by three, mean for many"*		

TABLE 3 – List of resources (London, Lisbon and Milan) [12]

Data from all these resources will be collected, aggregated, stored, analysed, securely exposed and visualised by a common platform explained below.

Platform – The objectives are to develop an Urban Sharing Platform (USP) that aggregates and manages data from a wide range of sources, built using common principles, open technologies and standards. The USP, designed under the sharing concept, is the glue between the two other city dimensions: **people** and the **places**. Developed by Altice Labs, it is responsible for gathering, aggregating and analysing all the data coming from sensors, gateways, local management systems and other smart devices distributed by different locations in each city.

A layered reference architecture (as represented in **Figure 4**) is the foundation of the USP and includes a set of blocks which define the main functionality capable of providing generic services to a smart city. This architecture is:

- Modular which enables, on one side, the reuse and, on the other side, the use of different technologies in different layers and different blocks, therefore, in different cities;
- Transversal which enables interoperability through very well defined APIs and data models. On one hand, the platform will offer standardised procurement in order to integrate new data sources as the city grows and digitalises; on the other hand it will expose secure interfaces to be used by multiple third-party services and end-user applications; public data will be available for both public and private/enterprise use (private data will be subject to the new General Data Protection Regulation (GDPR)).

Moreover, the architecture includes 4 layers:

- Sharing/Service layer: it includes marketplaces, presentation and reporting tools, KPI management and business intelligence. Data will be available in a secure

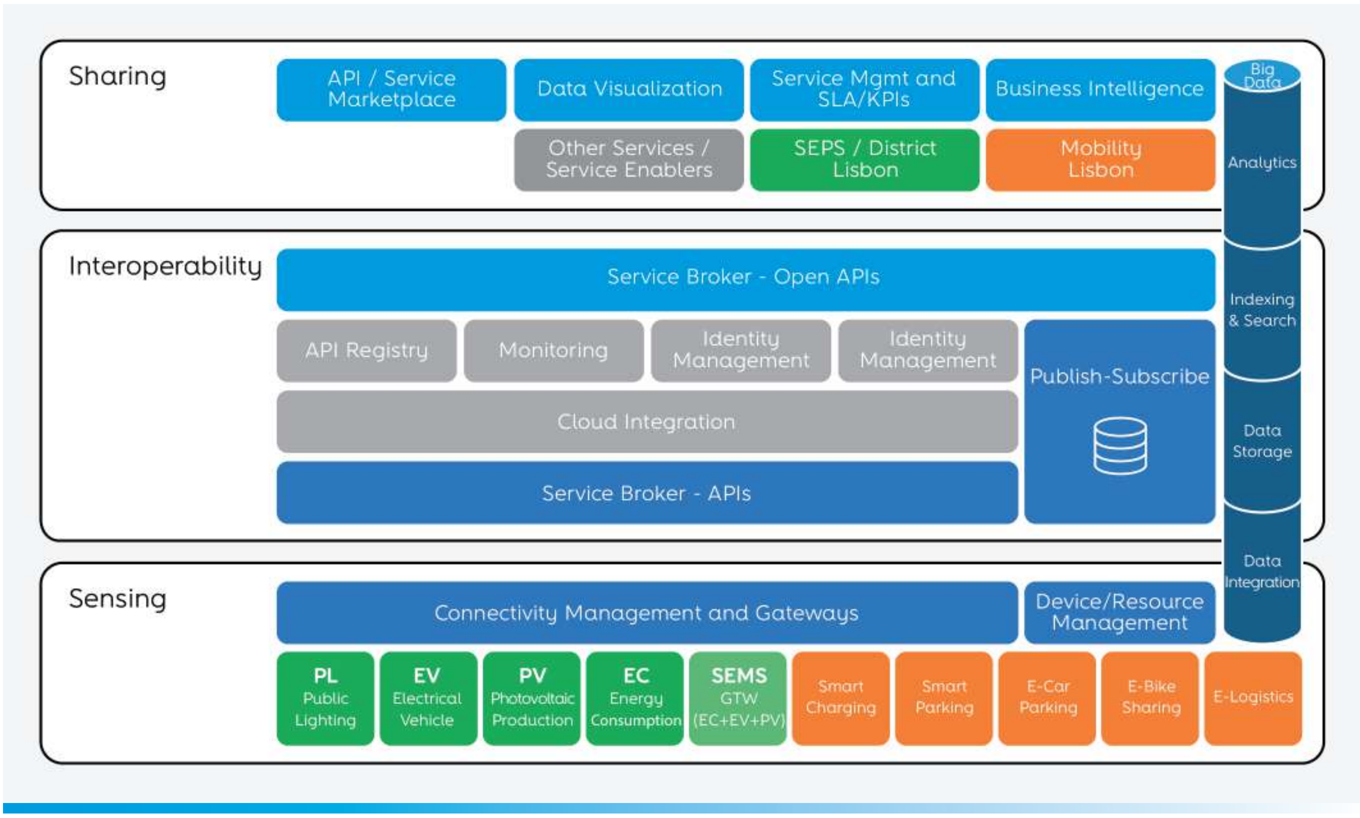


FIGURE 4 – USP reference architecture, resources and services [12]

(or open) manner at the urban platform according to the use cases requirements and the different ecosystem stakeholders and clients' needs (e.g. citizens, municipality, city managers, merchants, start-ups) through proper tools (e.g. marketplace, API, web portals, applications, dashboards);

- Interoperability layer: it provides processing capabilities, data aggregation, enrichment, analytics and anonymization if needed. At this layer, data is converted into useful knowledge (by applying machine learning techniques when necessary) to be consumed by the different third-party services and applications. Internally it includes functionalities as identity management, event processing, publication and subscription mechanisms, registration, monitoring and service management. APIs will be provided to standardize the access to the platform;

- Sensing layer: it includes a layer of adaptation and parsing of data (if needed) to convert legacy formats into standard data models. Also sensing resource (e.g. sensors and gateways) registration and management are provided;

- Big data layer: this transversal set of components cross all the previous layers. It integrates typical database management functionalities as storage, searching and filtering, caching and basic data analytics.

The areas addressed by the project include urban rehabilitation, mobility and energy sustainability. The results will determine strategic interventions in critical and specific locations in each city. According to the use cases previously identified, resources/data that will be considered by the project (represented in **Figure 4**) include:

- Public Lighting (PL);

- Electrical Vehicles (EV);

- Photovoltaic energy Production (PV);

- Energy Consumption (areas of intervention) - two aggregated energy systems:

    - a Sustainable Energy Management System (SEMS) for Buildings/Districts;

    - a Sustainable Energy Planning System (SEPS) for the City.

- Charging stations;

- Public parking;

- Fleet management.

On top of the USP (integrated with the sharing layer) a set of services will be provided to fulfil the needs of the energy and mobility use cases which will serve both the municipality and the citizens.

The advantage of the USP is to provide independence between the sensing layer and the sharing/service layer while, at the same time, offer a standard and secure access to data. Resources may belong to different providers and be located in different places in the city; data may be provided in different formats (in spite of serious encouraging of using directly the standards APIs and data models provided by the platform in order to standardize procurement); different levels of privacy may be required; real-time and historical data may be needed; raw and processed data may be consumed by the different applications with heterogeneous levels of access rights - the USP will be able to cope with all these requirements.

Wrapping up

The world we live in is changing more and more rapidly and the vast amount of data that is being generated daily and exchanged among appliances, applications and people, is fuelling that transformation either through technology development and innovation, social relations or even policy making. The consistent trend, for people to concentrate in large urban settlements, calls for innovative solutions in the areas of housing, transportation and mobility, education, healthcare or energy management. New jobs and new business contexts will also necessarily appear and as a direct consequence of all that, the way cities are governed will have to change.

Data and information flow is the key that should be used to make cities a better place for people to live and work. Municipalities should learn to manage the urban space focussing on integrated, affordable and scalable solutions, looking for solutions that provide answers to their operational requirements and create a perceivable value for citizens at the same time.

The smart city concept, despite being traditionally associated to the reign of technology, should instead be seen as the guidebook about the way to use information and communication technology to improve the quality of life of the citizens. The spreading of broadband access to Internet and digitalization are undoubtedly the principal enabler of the entire process. To be called smart, the city should seek to create an encouraging environment to innovation and co-creation, offering their citizens opportunities for collaboration and sharing powered by the synergies arising from the combination of urban and digital space. This is a radically new vision of city's development, tying together technology, social inclusion and sustainability.

Economy is changing too. Digital economy is growing worldwide faster than the global economy as a whole [13], partially driven by the substantial lowering of transaction costs inherent to the online world, traditional business models are also consistently evolving to adapt to this digital transformation. Sharing economy has moved its base from the traditional "asset own" paradigm to a new and disruptive paradigm, based on the temporary access and sharing of assets, services and information. However, to be fully successful, two paramount factors should be taken care of – trust and reputation. Under the sharing or collaborative paradigm, digital platforms and technology-based applications are entailed to cooperate towards the creation a network of trustworthiness, which is the foundation that allows the system to work.

In a global marketplace of millions of people, data and analytics are the new feedstock that make possible to obtain comparative advantages over competitors by speeding up decision making processes. Data collection and analytics are the melting pot where a multitude of sources of data, either produced by citizen´s devices or machine-generated data, residing outside of the control of a single public or private entity, to extract new individual or business oriented information.

The H2020 European Project – Sharing Cities in which Altice Labs is a full member, leverages this concept by proving a real ground to make smart cities and the sharing economy concept a reality. Sharing Cities project aims to investigate new forms of collaboration on sharing economy in the context of smart cities development by enhancing mechanisms for citizens' engagement and improving Municipalities capacity to define policies and deliver innovative services through collaboration and co-design.

In order to achieve this, the project promotes collaborating between cities and communities to drive fundamental behaviour changes by sharing. By fostering international collaboration between industry and cities, the project intends to lead by example, while it seeks to develop affordable smart city solutions by integrating commercial-scale products with high market potential. Also it will offer a framework for citizen engagement and collaboration at local level which will strengthen trust between cities and citizens. The demonstrations districts, located in the cities of Lisbon, London and Milan, will implement replicable urban digital solutions and collaboration models. Each city will retrofit buildings, introduce shared electric mobility services and install energy management systems, smart lampposts and an urban sharing platform by following the same best practices and guidelines. To validate the overall approach the "fellow" cities of Bordeaux, Burgas and Warsaw will replicate the solution.

Finally, there is no technology available to make a city smart or force the economy to create new jobs and promote social inclusion. Technology will always be a (powerful) enabler of transformation by allowing for openness, fairness and transparency of data and information.

References

[1] L. Narayan, World Cities Report2016 - *Urbanization and Development: Emerging Futures*, no. 8. 2014.

[2] United Nations, "The World's Cities in 2016: Data Booklet.," *Econ. Soc. Aff.*, p. 29, 2016.

[3] P. D. United Nations, Department of Economic and Social Affairs, "World Population Prospects The 2017 Revision Key Findings and Advance Tables," *World Popul. Prospect. 2017*, pp. 1–46, 2017.

[4] ATIS, "Smart Cities: Technology Roadmap," 2017.

[5] European Commission, "Connectivity - Broadband market developments in the EU," pp. 1–37, 2017.

[6] F. Dadaglio and D. Welsh, "ISO Smart Cities - Key Performance Indicators and Monitoring Mechanisms," no. May, pp. 3–4, 2015.

[7] FastCompany, "Defining The Sharing Economy: What Is Collaborative Consumption–And What Isn't?," 2015. [Online]. Available: https://www.fastcompany.com/3046119/defining-the-sharing-economy-what-is-collaborative-consumption-and-what-isnt.

[8] FastCompany, "The 3 Generations Of Smart Cities," 2015. [Online]. Available: https://www.fastcompany.com/3047795/the-3-generations-of-smart-cities.

[9] R. Belk, "You are what you can access: Sharing and collaborative consumption online," *J. Bus. Res.*, vol. 67, no. 8, pp. 1595–1600, 2014.

[10] P. Muñoz and B. Cohen, "Mapping out the sharing economy: A configurational approach to sharing business modeling," *Technol. Forecast. Soc. Change*, vol. 125, no. Supplement C, pp. 21–37, 2017.

[11] A. H. Henten and I. M. Windekilde, "Transaction costs and the sharing economy," *Info*, vol. 18, no. 1, pp. 1–15, 2016.

[12] Eurocities, "Sharing Cities." [Online]. Available: http://www.sharingcities.eu/?_sm_au_=iTrV6VTvssvtVtps.

[13] P. Assembly, "Euro – Latin American Parliamentary Assembly Resolution: on a digital agenda in the EU and the LAC countries," no. September, 2016.

04

Digital innovation labs and the digital CSP

Pedro Carvalho, Altice Labs
pcarv@alticelabs.com

In order to collaborate or compete with the digital giants, CSP need to adapt their networks and systems, to train/hire/convert their staff with the relevant skills, to change their internal culture and market image and to establish the right partnerships with customers and complementary companies. Digital innovation laboratories appear in this context as a very relevant engine to speed-up the transformation process for CSP. More than physical laboratories, these concepts may be the expression of a digital strategy composed of several different tools and initiatives.

Keywords

Innovation; Digital Disruption; Startups; CSP; Labs

Introduction

As all the world physical entities get interconnected (individual wearables, homes, buildings, cities, cars, factories) the digital giants (e.g. Google, Amazon, Facebook, Apple, Microsoft) are focusing most of their efforts on integrating these new connected domains into their digital ecosystems, acting either as front-ends or as cloud computing intelligence providers for other previously established ecosystems, effectively merging with those and capturing a part of their traditional value chains. So far this effort has focused mostly in the B2C arena (individuals, citizens), therefore leaving the B2B area more open for other players like the Communications Service Providers (CSP) and media providers, at least for now.

In order to collaborate or compete with the digital giants, CSP need to adapt their networks and systems, to train/hire/convert their staff with the relevant skills, to change their internal culture and market image and to establish the right partnerships with customers and complementary companies.

Digital innovation laboratories appear in this context as a very relevant engine to speed-up the transformation process for CSP. More than physical laboratories, these concepts may be the expression of a digital strategy composed of several different tools and initiatives.

This article explains what digital innovation labs are and what they aim by briefly analysing relevant experiences, pitfalls as well as suggesting possible approaches to take the most out of them.

What are digital innovation labs?

Around 2015, a new innovation trend emerged and started to show up into all kinds of news items and blogs all over the world. It went like this: "Company XYZ launched a new innovation laboratory" (for illustrative examples, see links provided on references [1] [2] [3] and [4]).

This trend hasn't stopped since then, even though innovation laboratories (Labs) can now present some variants in name and form, such as "digital innovation labs", "open innovation labs" or even "open innovation platforms" – for the sake of simplicity and up-to-date terminology, in this article we will use the name "digital innovation labs" (DI-Labs).

In fact, the relevance of DI-Labs only ramped up as a result of the increased levels of digital disruption showing up in many industries. The creation of DI-Labs is seen by the experts as one way to accelerate the digital transformation efforts of most businesses, particularly for bigger, historical and incumbent companies, where transformation efforts are usually harder and take longer.

But what are DI-Labs and what is their purpose?

There is not a single correct answer as for some companies the term can simply mean a physical laboratory while for others it is the lighthouse for a whole new digital strategy composed of many different tools. But the main objectives of DI-Labs could be summarized as:

- For many companies, one of the first objectives is to raise the level of digital awareness and competencies (the so called "digital IQ") of their staff, through education, experimentation and dissemination of the "digital era" concepts;
- Experimenting with new technologies, prototyping potentially new and disruptive products and services and designing new business models with minimal risk is another objective at the top of the list;
- Not less important, DI-Labs might aim at engaging with partners and customers in experimental projects, possibly running co-creation ventures. Those could eventually

be performed as spin-offs from the main company;

- Finally, DI-Labs can also act as interfacing incubators for startups working closely with the owning company, possibly running incubation programs in partnership with established startup incubators and accelerators. This is usually associated with some form of corporate venture capital strategy, which can later take stakes in the most promising companies.

Considering one popular innovation types matrix (**Figure 1**), one could say that DI-Labs would aim to address "breakthrough" and "disruptive" innovation within large corporations.

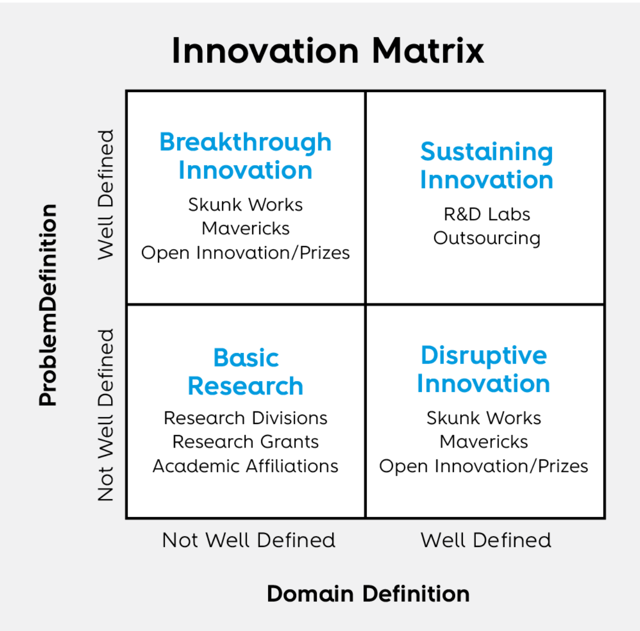


FIGURE 1 – Popular innovation types matrix [5]

CSP's digital disruption

CSP are among the most impacted companies by the digital transformation: they are at the same time a fundamental instrument for this transformation and deeply impacted by it, as it opens the door to all kinds of new competitors, from the digital giants to relatively small companies riding the digital wave.

According to Gartner [6] during the next few years the combination of key technologies like Data Analytics, Artificial Intelligence, Automation and Software Defined Networks & Services will have a highly disruptive and transformative impact on CSP. Their fiercest competition will increasingly come from the likes of digital giants like Google, Amazon, Microsoft, Apple, Facebook, Netflix, which are already outspending CSPs in R&D and therefore setting the pace and scope for disruptive changes, innovation, new business models and customer's expectations on their service providers of all kinds.

CSP will, in the near future, be at the convergence point of the above mentioned disruptive technologies, with the crushing platform-type business pressure from the digital giants and, not least important, the increasing pressure of a generation of customers who grew up with the internet as their personal, professional and social centre. One could certainly call this convergence the perfect digital business disruption storm – See **Figure 2**.

CSP will need to adapt their "personality" to this new reality, by:

- Investing in internal transformation of both infrastructures and human resources, in order to gain key technological know-how, flexibility, innovation capabilities and adequate response time to challenges – the so called "digital transformation";

- Investing in innovation (products, services, business models), both internally and externally, in close alignment with the market trends and needs;

- Establishing a network of partners with the right experience in the Internet inspired, platform-type business models;

- Focusing on providing platform-like services,

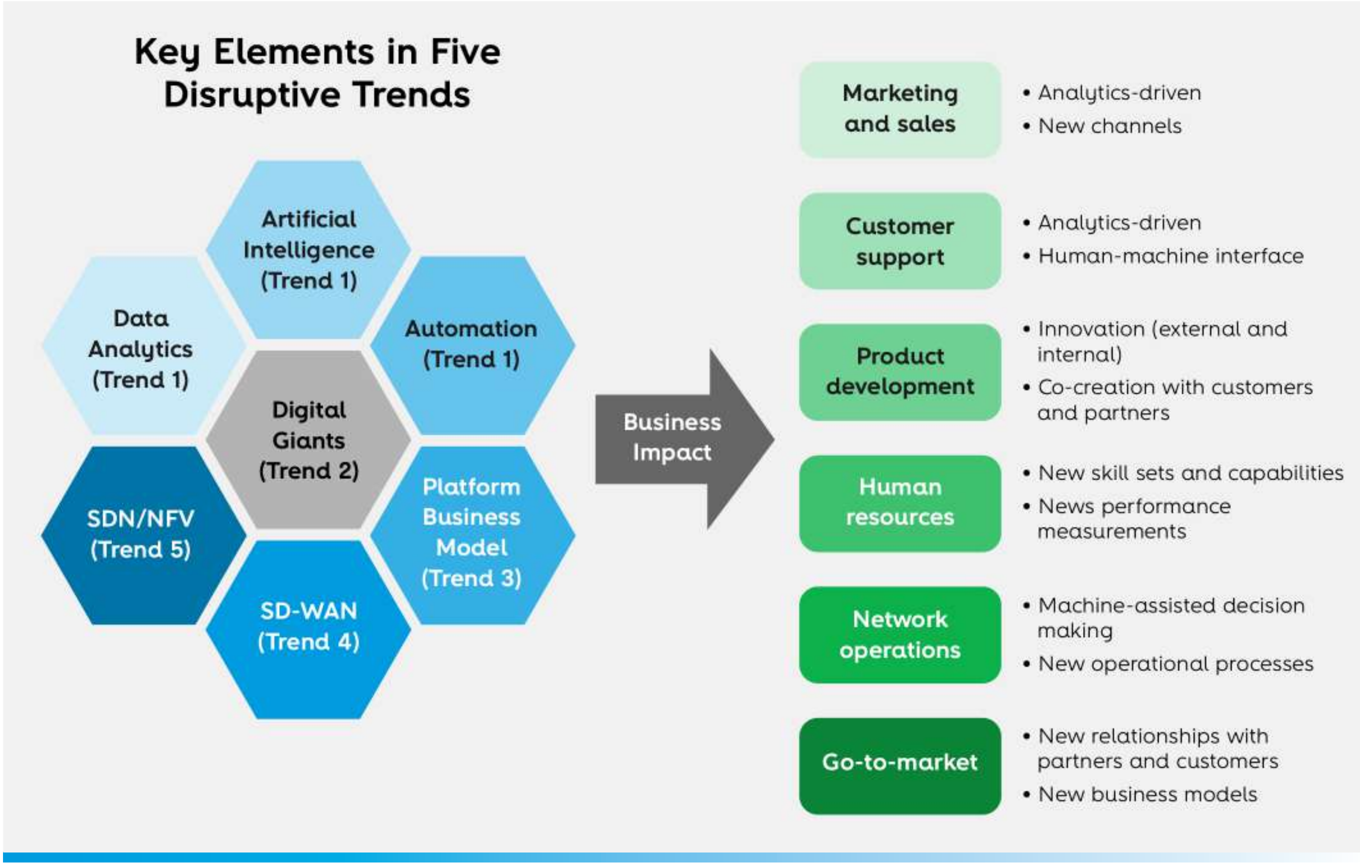


FIGURE 2 – Near term disruption for CSP [6]

establishing network effects by providing key enablers for their partners businesses, in order to attract them;

- Defining and controlling the right key performance indicators (KPI) for the measurement of the effectiveness of the investments in new technologies.

Setting-up DI-Labs, as the visible core of a digital transformation strategy, could certainly be one of the strategic tools to be adopted by CSP in order to accelerate their internal digital transformation and their capacity to connect with partners and end-customers, therefore raising their “innovation in collaboration” capabilities. Some of the bigger CSP are already following this track, as we will see in the next section.

CSP’s digital transformation initiatives

Most of the well known CSP have realized this urgent need for change and are speeding up their digital transformation with several initiatives publicized and even more initiatives kept secret internally. In this section we have a quick overview of the most representative initiatives with CSP, through an analysis of publicly available information.

The Verizon case

Verizon is recognized by most analysts as one of the most effective CSP in the world. Their advanced Long Term Evolution (LTE) network covers 95% of the USA population, their wireline networks are one of the first choices for both consumers and enterprises (Fios – fibre optics Pay TV service and global IP backbone). Verizon is diversifying its activity into digital media and advertising (through acquisitions), Machine-to-Machine (M2M)/Internet of Things (IoT) (leveraging its LTE network) and OTT services which could be consumed outside its network boundaries. However, Verizons's strategy focus is clearly set on connectivity: delivering the best services and attracting more and more customers to its networks and platforms.

Verizon's innovation strategy [7] is mostly based and managed from two innovation labs (San Francisco - California and Boston - Massachusetts), both locations are well known technology hubs for startups and investors. Additionally to these Labs, Verizon also created a corporate venture capital instrument instrument, the Verizon Ventures which invests on early-stage funding of startups presenting products closer to the strategic plans and which could therefore be capable of serving the company's short to mid-term business objectives. According to the references, Verizon Ventures has over 50 startups on its portfolio and a total of 9 exits (initial public offerings and acquisitions), with an average of six startup deals per year.

Verizon runs a dedicated innovation program ("Verizon innovation program" [8]) which serves as a framework for Verizon and its external partners to work together in areas like Advanced Communications, Networking, Energy management, Healthcare, M2M/IoT, Media & Entertainment, Retail and Smart Accessories. Entrepreneurs can join Verizon at the innovation Labs to further develop their products, using the company's networks and platforms. Within the scope of this program, Verizon seems to be open to work in innovation areas not totally aligned with its core business, rather willing to experiment with partners with total open-mindedness.

The AT&T case

AT&T has always been the reference operator in terms of long term strategy and innovation capabilities. The digital innovation strategy of AT&T [9] comprises several complementary tools, namely:

- The AT&T Labs ("Discover the science behind the service"), where fundamental research and innovations and digital transformation are pursued. The most relevant initiatives of the Labs revolve around Software Defined Networks (SDN) and Network Function Virtualization (NFV) and the Enhanced Control, Orchestration, Management and Policy (ECOMP) software platform and Open Network Automation Platform (ONAP) which are expected to revolutionize the future of CSP's networks;
- The AT&T Foundry, a set of 6 specialized locations "where ideas are made", together with partners. Those "foundries" can be seen as DI-Labs, specializing in network transformation and consumer services (Palo Alto), home and connected cars (Atlanta), connected healthcare (Houston), IoT and enterprise services (Plano 1, Plano 2, both in Texas), operational efficiency and cybersecurity (Israel). Altogether, the foundries collaborate with an average of 500 startups every year;
- The Innovation Pipeline (TIP) – internal crowd-sourcing platform, with associated seed funding for the best ideas to be developed, either internally or through partners. It also acts as an engagement and team building tool for employees;
- AT&T Venture Capital, funding startups and spin-off ideas.

The Deutsche Telekom case

Deutsche Telekom's (DT) digital "Innovation+" strategy [10], coordinated by the "Technology and innovation business unit", focuses on three different lines: in-house developments, partnerships and startup funding (venture capital), in this way trying to reach an healthy balance of internal and external product and service development while covering the entire digital ecosystem. The agents for this strategy are ”Telekom Innovation Laboratories” (T-Labs) for internal innovation, “Deutsche Telekom Strategic Investments” (DSTI) for partnering with expert companies and hub:raum as a startup incubator (with associated corporate venture capital mechanisms), the external innovation engine of DT and main promoter of collaborations in several areas, namely on OTT services. Apart from these three agents, T-Systems, the fully owned ICT services company of DT, plays a major role in digital services and products delivery, due to its large expertise in information systems, with major impact in the areas of Information and Communications Technology (ICT) and cloud projects, IoT and security. DT also promotes an internal entrepreneurship and innovation line of work, which supports business ideas generated inside the company (the Telecom Innovation Pool).

The Orange case

Orange is considered one of the most advanced digital CSP by the market analysts, totally focusing its core strategic vision on digital innovation [11]. This focus can be found in the company’s five year strategic plan published by Orange in 2015 and named “Essentials 2020” [12] which shows a clear direction to digital-services launch, but also to an internal and external digital transformation.

Orange has a clear organization for digital innovation, consisting of several different instruments, namely Orange Labs (research and network development), Orange Fab (startup acceleration program, covering more than 14 locations in the world, now linked to “Go Ignite”, a multi-operator startup incubator), Orange Gardens (innovation ecosystem center for co-innovation with partners), Orange Valley (web development and innovation) and Orange Digital Ventures (Corporate Venture fund, for external investments). In this mesh of Labs, partnerships and financing instruments, Orange develops projects in all the relevant digital disruption areas.

The Vodafone case

Vodafone is one of the largest mobile operators in the world, featuring more than 460 million mobile customers, 43% of which are heavy data service consumers.

Vodafone’s official strategy [13] focuses on three pillars: Data, Convergence and Enterprise. This strategy translates into the modernization of its broadband and LTE networks and preparation for 5G networks, the provision of triple play and quadruple/quintuple play services, with TV and fixed broadband over fibre and the focus on enterprise services like IoT and Cloud. Unlike many of its competitors, Vodafone has never decided to create a digital service unit, partly due to not seeing clear benefit derived from this move and partly due to some bad results in the past (e.g. Vodafone Xone and Vodafone Ventures).

In terms of innovation and digitalization Vodafone uses a very pragmatic approach: if it wants to enter into a new business area, the company buys the relevant player or players in the area. If it perceives that a new “innovative” product is not producing the expected results the company immediately discontinues it and eventually sells the associated assets. Most consumer services on top of the core connectivity business are provided through App and content partners (e.g. video content, music). The most visible digitalization effort from Vodafone in the last years was its rather successful “My Vodafone App” which has been widely adopted by customers and is said to reduce churn and raise the Net Promoter Score (NPS) of the company above its direct competitors.

Traps and pitfalls

We strongly believe that DI-Labs are fundamental instruments for CSP in the continuous process of trying to stay close to your partners and customers and at the same time ahead of your competition. However, if not managed with caution, the Labs might become a simple waste of time and money for all of its stakeholders. Below we try to summarize the main known traps in this area.

Ideation and creativity vs innovation

People tend to think that innovation is simply the generation of new ideas or creativity, while innovation is actually the process which goes from generating ideas, to selecting the most promising, develop viable and sustainable business models and actually launch the product or service to the market where it generates enough revenue during it life span. In other words, an innovation is an idea which found a market!

If companies let their Labs turn into creativity and ideation Labs only, it's easy to fall into the trap of the "Innovation Theater" where you do all the visible things that the really innovative companies do (e.g. Apple, Google) but without a purpose behind it and the results to show in the end. They're just putting up a show for the media!

R&D vs innovation

Also very common especially for Labs run by academic teams, is to believe that research and development results are innovation results. Again, if the R&D results are not incorporated into a product or service with a sustainable business model what you will have are inventions, eventually patents, which can be a valuable asset in itself, but still not innovations.

Misalignment with corporate strategy

Even if the innovation lab is well managed and actually produces innovative and sustainable ideas, if it is not fully aligned with its main company's corporate strategy it will tend to lack focus in its activities, producing all kinds of promising innovations which die as orphans, because no business unit in the company adopted them as part of their product portfolio nor will the venture capital units want to bet their money on funding those ideas.

To put it simply, DI-Labs need to have a strategic focus, helping them to build the right culture, attract the right people and develop the right core competencies.

Lack of impact

Last in this list but still extremely important, comes the need for impact on company revenue. Companies creating a DI-Lab must keep this in mind when preparing the management team.

In fact, after two or three years of activity, everyone expects to have new products and services, created at or with the DI-Labs in partnership with a business unit, launched to the market and generating new revenues, capable to compensate the lost revenue with more mature products. The Labs must therefore have other metrics, besides simple activity metrics, capable of demonstrating actual impact on company new product launches and revenue. That is, unless the company sets up an innovation laboratory only for internal cultural change purposes or for media show-off, but those tend to get useless with time and get shut down.

Altice and digital innovation labs

The Altice Group vision positions the group with the ambition to become a converged telco, content and advertising service provider, with data as enabler. The group is composed of several CSP, media and advertising companies, in different geographies, ranging from smaller scale traditional operators to mid and large-scale cable and triple-play operators, providing both generic services and exclusive content to their B2C customers and a set of traditional B2B convergent services to SME and enterprise customers.

Based upon the DI-Lab concept, the group may position itself ahead by defining a digital innovation strategy to be implemented across all Altice DI-Labs.

Having in mind the group's organization, the best approach for Altice might be an orchestration from the CTO-level coordination considering the following (also illustrated in **Figure 3**):

- DI-Labs coordination:
  - Defining the focus and the key areas of work;
  - Defining the specialization of each one of the physical locations;
  - Establishing key working relationships (partnerships);
  - Defining the right KPIs.
- Altice Labs, the structure for innovation and product development, which exists physically in different locations and featuring:
  - Product development teams;
  - Research and innovation projects in local collaboration;
  - Research with local universities (AlticeLabs@University);
  - Exploratory labs.
- Altice@University:
  - Partnership agreement with key universities, in order to promote the digital culture change and needed skills among the staff of Altice, through a set of carefully organized courses and seminars.
- Altice Technology Partnerships:
  - Partnerships with suppliers and technological SME;
  - Partnerships with integrators and partners using the platforms of Altice.
- Altice incubation and acceleration:
  - Identification of relevant startups and participation in their incubation and acceleration, possibly resorting to established incubators in the relevant geographies;
  - Promoting and supporting spin-offs from Altice companies;
  - Supported by Altice venture capital.

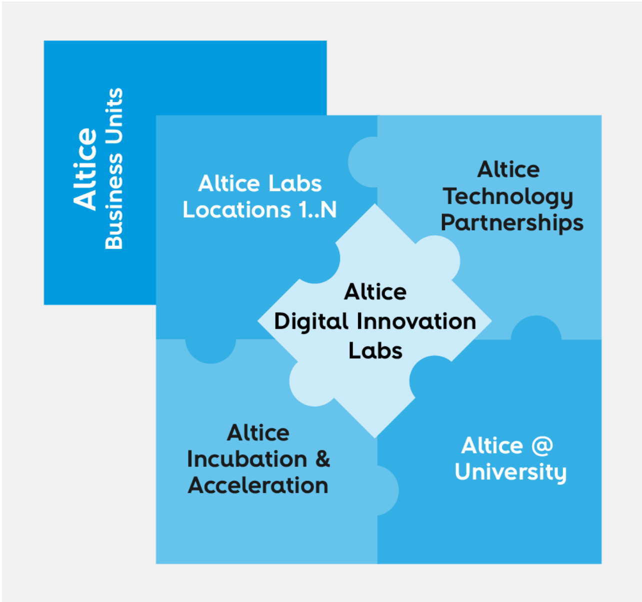


FIGURE 3 – Altice DI-Labs strategy

To be effective, the focus for Altice DI-Labs could be both internal (e.g. changing the cultures, defining and owning the future networks and platforms architectures, developing and deploying key solutions internally) and external – that is, focusing on collaboration with external partners and clients and building true ecosystems.

Final thoughts

The cycles of innovation are accelerating like never before, as digital disruption becomes more and more a daily reality in every aspect of people and businesses life, while at the same time the expectations of people on the outcomes of the usage of digital technology raises every day, becoming virtually unlimited.

These factors combined place a lot of pressure on all types of business but exert a crushing force on large, traditional and incumbent businesses which are challenged everyday by new and lean digital entrants, capable to steal their customers in a very short time. This is clearly the case for CSP and traditional media companies.

We believe that setting-up and promoting DI-Labs as the ones discussed in this article are a necessary step for CSP, particularly for those which were monopoly, incumbent, significantly big or old enough to be from a pre-digital age. Key aspects to consider when setting-up the CSP DI-Labs are:

- Clearly defining the objectives for the Labs – what do we expect to get from this laboratory, both internally (raising the digital IQ, performing the digital transformation, etc.) and externally (new partnerships, products and services, identifying and funding new startups, etc.)?
- Setting-up a multidisciplinary team – you will need the creatives, the designers, the engineers, the scientists, the product owners and the business experts to make it all work and actually produce innovation;
- Clearly defining the reporting line/lines inside the company in order to maximize the impact of the DI-Labs, both internally and externally;
- Deciding on location or locations – if you need lots of software engineers and telecom experts you should choose a location close to good universities in these areas. One location will probably not be enough for a large CSP as all the required skills will probably not exist in a single location;
- Deciding on the level of "openness" of your Labs – will they be internal? Will they involve business partners and end users? Should they be able to act as incubators for ideas and startup/spinoff companies? To what extent should the more fundamental research be performed by associated universities? How do you leverage and incorporate the huge ideation potential coming from your employees?
- If considering incubation, what should be the associated corporate venture capital strategy, in order to gain some level of control over products developed by incubated companies?
- How to link the activities and results of the Labs with the regular enterprise operations, product development and engineering as part of an integrated process of digital transformation? In other words, how do you guarantee that the Labs contribute to change the culture and skills of the organization and that the innovative products reach the production and distribution channels of the company;
- Setting-up the right KPI, in alignment with the corporate strategic objectives and metrics, possibly with different timeframes and more room for failure, but still with an eye on the revenue generated by new products created at or with the Labs. KPI examples are:
  - Size of innovation ecosystem (e.g. number of partners/developers using the ecosystem);
  - Cost reducing effects of innovation (internal);
  - Number of Minimum Viable Product (MVP) with a sustainable and validated business model;
  - Number of innovative products entering the product portfolio of the company;
  - Revenue coming from innovative products or services every year;
  - Number of venture capital deals/year + capital invested, etc.

With this article we expect to have contributed to the reader's understanding of the digital disruption for CSP and the possible lines of action to fight it (or align with it), namely the creation and running of DI-Labs by the CSP.

References

[1] The Guardian, "Guardian US to launch news innovation lab focused on using mobile technology to create deeper journalism with $2.6 million from Knight Foundation," 15 June 2015. [Online]. Available: https://www.theguardian.com/gnm-press-office/2015/jun/15/guardian-us-to-launch-news-innovation-lab-focused-on-using-mobile-technology-to-create-deeper-journalism-with-26-million-from-knight-foundation?_sm_au_=iTrRWHfQ26j610wF.

[2] Milton Exhibits, "PayPal Innovation Lab Launch," Milton Exhibits, 17 August 2016. [Online]. Available: http://milton-exhibits.com/en/paypal-innovation-lab-launch.html?_sm_au_=iTrRWHfQ26j610wF.

[3] Essential Retail, "NCR launches Innovation Lab in US tech hub Georgia," Essential Retail, 23 August 2016. [Online]. Available: http://www.essentialretail.com/news/article/57bc6a6fe144e-ncr-launches-innovation-lab-in-us-tech-hub-georgia?_sm_au_=iTrRWHfQ26j610wF.

[4] Database - Trends and Applications, "Dell Services Launches Innovation Lab for SAP HANA," Database - Trends and Applications, 8 August 2016. [Online]. Available: http://www.dbta.com/Editorial/News-Flashes/Dell-Services-Launches-Innovation-Lab-for-SAP-HANA-112854.aspx?_sm_au_=iTrRWHfQ26j610wF.

[5] G. Satell, "How to Manage Innovation," Forbes, 7 March 2013. [Online]. Available: https://www.forbes.com/sites/gregsatell/2013/03/07/how-to-manage-innovation-2/#c248104785a3.

[6] K.-Y. Foong, A. Nandan and J. Skorupa, "Market Trends: Top Five Disruptive Trends for CSPs, Worldwide, 2017-2022," Gartner, Stamford, 2017.

[7] F. Radicati, "How Telcos Innovate: Verizon," Ovum, London, 2016.

[8] Verizon Wireless, "Verizon Innovation Program," Verizon Wireless, 02 January 2017. [Online]. Available: http://innovation.verizon.com/content/vic/en.html?_sm_au_=iTrRWHfQ26j610wF.

[9] AT&T, "AT&T Innovation," AT&T, 02 January 2017. [Online]. Available: http://about.att.com/innovation?_sm_au_=iTrRWHfQ26j610wF.

[10] A. Satari, "Digital Profile: Deutsche Telekom Group," Ovum, London, 2017.

[11] P. Lambert, "Digital Profile: Orange," Ovum, London, 2017.

[12] Orange, *Essentials2020, Orange's new strategic plan*, Paris: Orange, 2015.

[13] D. Talmesio, "Digital Profile: Vodafone Group," Ovum, London, 2017.

05

Merged reality for everyone

Fausto de Carvalho, Altice Labs
cfausto@alticelabs.com

Leonel Morgado, Universidade Aberta & INESC TEC
leonel.morgado@uab.pt

This article addresses some interesting challenges and business opportunities within the promising merged reality ecosystem, which offers the vision of bringing together virtual, augmented and physical realities, seamlessly. The article also links the current status of this field with exploratory research and development work carried out by Altice Labs.

Keywords

Mixed Reality; Augmented Reality; Virtual Reality; Natural Interaction

Introduction

Virtual reality (VR) is currently under the spotlight, with multiple big industry players truly hyperactive introducing affordable technology, sophisticated services and exciting content.

There are interesting variations and combinations of reality and virtuality along Milgram et al.'s "Reality – Virtuality Continuum" [1], spanning from the physical world to its recreation as a completely digital environment using virtual components. The middle ground is the so-called mixed reality (MR), which includes augmented reality (AR) – where virtual elements are overlaid on the physical world – and augmented virtuality – where data and video from the physical world impact the virtual environment.

Most of the applications and services in this domain are usually associated with a paraphernalia of technology, such as controllers, sensors and cameras, making it rather challenging to create a simple and pleasant user experience. Nevertheless, 360° video, low-cost display devices (including smartphone adapters), gestural and somatic interaction approaches, integrated tracking and increased mobility have been improving significantly and truly immersive experiences are now becoming possible and quite affordable.

In this article we connect the current status of this field with exploratory research and development work carried out by Altice Labs, and we address some interesting challenges and business opportunities within the promising merged reality ecosystem, which offers the vision of bringing together virtual, augmented and physical realities, seamlessly.

From virtual reality to mixed reality

Three different technologies have evolved in parallel, and are now witnessing a convergence, raising the tantalizing prospect of widespread availability as MR for everyone: VR, somatic controllers and AR.

VR was the first to emerge. After decades of trials and announcements, the technology has finally caught up with the mass market needs. Three-dimensional (3D) graphics are now commonplace, not only in desktop computing but also for handheld devices such as smartphones. For years, simply having access to 3D virtual spaces was deemed "VR". Since it is now so common, however, the term evolved. By mentioning "VR", one now expects to have the visual experience of being inside the virtual space, using immersive devices much better, convenient and affordable than the heavy and high-latency devices of the 1990s. High-end devices are now priced at the €400-€600 level of video game consoles (PlayStation VR, Oculus Rift), except some more refined alternatives (HTC Vive: €900-€1000). Low-cost devices are almost free: using cardboard or plastic assemblages, people can use their own smartphones to enjoy immersive VR. Branded solutions, such as Samsung Gear VR, cost a few dozen euros, while Google Cardboard and similar devices may cost only €5.00 or even less.

Somatic controllers (i.e., motion-based or gesture-based, for natural interaction) have matched the widespread availability of VR, both in availability and diversity of low-cost devices. Arguably the first major market success was Nintendo's Wii console controller, whose popularity caught by surprise most technology and video game media pundits, leading to a variety of competing offers, such as PlayStation Move, all based on simple acceleration tracking and visual tracking of the controller's position. In its wake, Microsoft Kinect offered full-skeleton and depth sensing for €100 and originated a cascade of application scenarios, from end-user entertainment to medical applications. The growth in media interest and business investment means that dozens of somatic controllers are now available, from Microsoft Kinect alternatives to full-body tracking suits such as PrioVR, alongside specialized devices such as Leap Motion controller for hand and finger tracking, the Myo gesture control armband for detection of muscular activity in the arm, or spatial handheld controllers specifically designed to interact with VR spaces seen through Oculus Rift or HTC Vive. Wide adoption was the main doubt in this equation, but the surprising success of a high-range offering (PlayStation VR), which surpassed Sony's expectations [2], allows promising expectations in this regard.

There were no such doubts regarding AR popular appeal. It grew in success, step by step. Cameras on smartphones have been used from very early on to extract information from signs in the physical world, such as bar codes or their bi-dimensional variety, QR codes. For instance, QR codes have recently been used in wine bottles to strengthen the relationship with the consumers: using their smartphone, they can find out more about the wine and get wine tasting reports or access social media [3].

A classical example of AR is the Layar Reality Browser, introduced in 2009. Its development platform allowed developers worldwide to provide a multitude of layers of georeferenced data, that could be explored live with the mobile camera application (App) as points of interest in the vicinity of the user [4].

Full-fledged, animated, 3D AR has long shown its appeal to mass audiences. A major breakthrough in popularity came in 2009 when Sony launched a camera attachment to the PlayStation portable (PSP), its portable video gaming console. This attachment was bundled with the Invizimals game [5], which not only got critics' awards but also achieved large public success, warranting the release of numerous sequels and expansions. Children would seek out "invisible to the naked eye" creatures called Invizimals everywhere: on their homes, on their pets, on their clothes, using the PSP camera to track appropriate colours.

Invizimals were not, however, only creatures visible on a device's screen. Using collectable cards as markers, children would place the creature in battle positions in the physical world and then watch them battle it out in their full colour, animated glory, superimposed with the actual physical settings (**Figure 1** is a screen capture of a battle).

FIGURE 1 – Invizimals ready to battle it out in MR

Exploration and technological experimentation

Cutting edge areas often lead to the emergence of innovative ideas with the potential to satisfy new or existing market needs, to improve existing products, services, processes or the organization itself. Thus, Altice Labs has a tradition of continuously engaging in collaborative Research, Development and Innovation projects as part of a sustained strategy for technological leadership. Throughout the years, internal capabilities, creativity, experience and knowledge have been leveraged by projects and partnerships with world class universities, R&D institutions, suppliers and customers in several projects to monitor, study and explore the advances in multiple relevant areas including the above mentioned VR, AR and natural interaction.

More than just getting acquainted with technology, we have been pursuing scenarios and use cases related to promising application areas, using several internally funded exploratory research projects to anticipate challenges and foresee opportunities.

VR was a fast pace emerging area when we started experimenting with Second Life environment, a decade ago. We explored the creation of 3D content with the University of Aveiro and the social networking and interaction aspects with the University of Trás-os-Montes and Alto Douro (UTAD). Furthermore, we worked concepts around chat bots, multimodal customer support teams' management [6] and virtual presence.

Synchronized online gymnastics, for example, may provide new possibilities for enhancing the physical and social well-being of people with restricted mobility. We went back to Second Life and OpenSimulator technology to prototype **Online Gym**, a virtual 3D platform where different users physically apart at multiple locations may attend together a workout session coached by a monitor, all of them connected over the Internet and represented by their avatars directly animated by the movement captured on the Kinect device plugged into each personal computer [7].

FIGURE 2 – Online Gym prototype used on Bang Awards installation

Furthermore, we reused the Online Gym prototype to create an installation for public interaction, in the context of an animation film festival - Bang Awards 2014, held at the Torres Vedras castle (**Figure 2**). In fact, exergame approaches created around VR avatars and bodily interaction, proved to be a great fit for well-being and eHealth scenarios: with the Instituto de Telecomunicações - Porto, we were able to further experiment with real time markerless MOCAP interaction in 3D virtual environments, creating the **Move4Health** serious/exergame for validation of the use of Kinect as a reliable device for physical rehabilitation involving large motor skills.

InMERSE is another important exploratory project carried on in partnership with the Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC), with researchers from UTAD and Universidade Aberta. A set of devices was analyzed in depth, both at a functional level and under a programming perspective: Oculus Rift, Leap Motion Controller, Google Glass, Microsoft Kinect and Myo Gesture Control Armband.

The main outcome was an open source framework - the InMERSE framework [8] - for multimodal gestural input. It enables the development of gesture-controlled applications that are independent of gesture-recognition devices and gesture-recognition methods [9]. As demonstrations, two main prototypes were created: a digital signage platform and demo, based on Leap Motion interaction, and the "First Armada of India" installation [10] integrating all the aforementioned devices, around a game concept depicted in **Figure 3**, recreating a notable moment of Portuguese history and literature.

FIGURE 3 – InMERSE installation concept: "First Armada of India" game scenario

The sitting player wears an Oculus Rift VR display and takes the role of the helmsman, steering the XVth century ship with his hands, detected by Leap Motion attached to Oculus, to avert opponent attacks. The standing player is Adamastor, the mythical giant trying to sink the ship. He gets its position as a compass heading value on Google Glass and throws rocks with movements detected by the Kinect and the Myo armband [11]. **Figure 3** shows these players in the same physical space, although they could be located remotely. **Figure 4** shows the perspective of the Oculus Rift player.

FIGURE 4 – "First Armada of India" in-game prototype view

Reaching the market

Beyond the exploratory initiatives presented in the previous chapter, some of our VR and AR projects actually reached real users, making stronger evidence of the exploitation potential of these technologies.

3D Virtual Worlds has been an area with big popularity among the education community, not only by virtue of the collaborative and the social networking features, but also due to the powerful role-playing capabilities, which are particularly adequate to many professional training situations. We explored those aspects together with UTAD in projects **MULTIS** and MULTIS II, establishing an extensive set of functional, integration and organizational requirements [12] to enhance our Learning Management System – Formare – with 3D Learning functionality that was afterwards made available to some of our customers (**Figure 5**). The approach was developed as a software architecture with separation of concerns (SoC) as its core focus. This MULTIS architecture may be deployed on other platforms, laying the grounds for obsolescence-resistant deployment of virtual and gaming technology [13].

FIGURE 5 – MULTIS instance in Second Life

Rama stands for Aveiro Mobile Augmented Reality and it's an Android/iOS App for browsing points of interest in AR, map view and routes. It was a development with collaboration from Load Interactive, created for Inova-Ria in the context of National Strategic Reference Framework (QREN) project "Parque da Sustentabilidade".

MEO Go VR is essentially an extension of MEO Go to VR through Samsung Gear VR powered glasses [14]. Developed for MEO by Altice PT in collaboration with Altice Labs and Gema Digital, the solution provides a unique and immersive experience with 360° content and live television. The App features 360° videos from multiple MEO partners and its environment transports users to a virtual living room where MEO customers with the MEO Go service can watch their favourite live channels on a giant screen (**Figure 6**).

FIGURE 6 – MEO Go VR

The future is arriving

The joint success of both VR and AR technologies is causing the emergence of new services beyond entertainment. IKEA's digital catalogue App in 2014 allowed customers to see how the furniture would look like in their homes by simply placing the paper catalogue on the floor; Holition Nails App enables trials of varnish on our own fingernails; Boucheron's allows us to try out jewellery (**Figure 7**)... the list is ever-growing.

Hands-free solutions are starting to emerge, albeit still exploratory: wireless headsets such as Google Glass, Microsoft HoloLens and Meta 2 (see **Figure 8**) are trying out new ways to enable AR to be used without locking users' hands to a specific device or position in front of a camera (as in Boucheron's example in **Figure 7**). Their applications are still being envisioned and

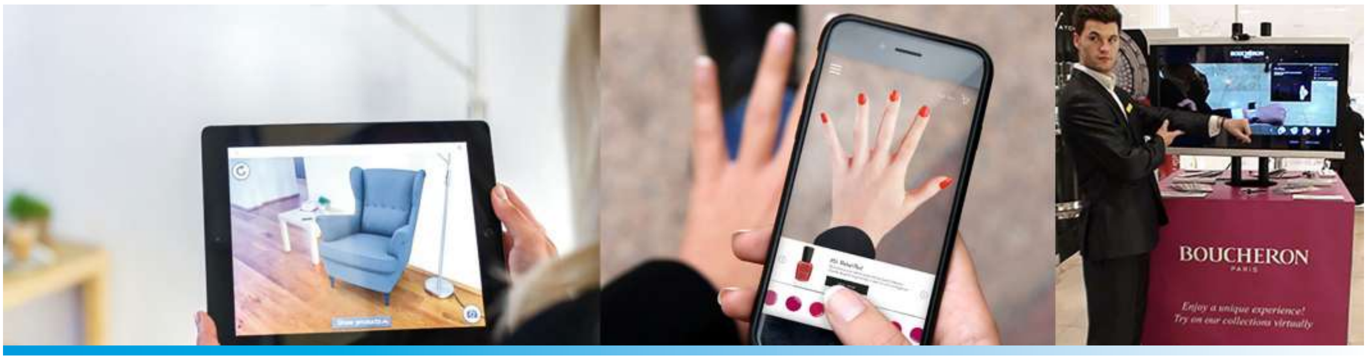


FIGURE 7 – Customer-oriented AR applications: IKEA, Holition, Boucheron

FIGURE 8 – Google Glass, Microsoft HoloLens and Meta 2 headsets

proposed, with business and industrial applications in the forefront (e.g. mechanical maintenance for manufacturing and aerospace industries), but treading the way for the future. Meanwhile, the availability of technology for indoor localization and the huge advances in mobile computer vision already made possible the debut of AR solutions to explore and navigate the complexity of underlying data and processes for management of sites, such as data centres and factories.

Industry experts project combined VR/AR revenue will reach over €100 billion by 2021, 75% being AR [15] [16]. Massification is still facing some hurdles, the first being the need for devices with the right specifications and with an affordable price target, which means high-tech headsets will be used mostly in professional scenarios for the time being. Nevertheless, recent developments show that smartphones may soon be paving the way to mobile AR widespread, with a forecast of one billion users and a market value of approximately €55 billion for the next 4 years. Terminals soon will incorporate the right features, e.g. Google ARCore, with manufacturers eager to have arguments to improve stalled sales. Developer ecosystem is spinning up, catalyzed by software platforms from major contenders, e.g. Apple and Facebook. Developers explore new capabilities and leverage the cloud, hence mobile data usage will multiply and telco will have an opportunity to add another profit stream to their 5G operations.

As interface devices become less intrusive, untethered and more intuitive, it's easy to realize scenarios being anticipated by researchers [17], with Internet of Things (IoT) and Artificial Intelligence (AI) evolving and combining with MR, profoundly disrupting and enhancing the way we perceive and explore information, manage knowledge and interact with equipment, buildings, people and the world at large.

The amazing and even upsetting concepts presented by Matsuda in his Hyper-Reality provocative and visionary movie [18] contain some glimpses of what possibly would be an ultimate Merged Reality (**Figure 9** and **Figure 10**): the physical world saturated and intertwined with personalized and context-aware streams of digital media, scenarios of urban augmented communication and interaction taken to a mind-boggling extreme.

FIGURE 9 AND FIGURE 10 – Hyper-Reality, speculating around the ultimate merged reality

We are not going that far in making science fiction real (or are we?), but clearly a cyberspace type of near future encompasses a multiplicity of business opportunities for operators, service and content providers, in rising key areas such as AI, IoT, advertisement, micropayments, cloud and virtualization, mobility, brokerage, identity, authentication, access and security and privacy management, etc., and many of the now futuristic 5G use cases become even more relevant. Working around these scenarios is certainly part of the homework for whoever is going to stay and thrive in the digital market for the upcoming years.

References

[1] P. Milgram, H. Takemura, A. Utsumi and F. Kishino, "Augmented reality: A Class of displays on the reality-virtuality continuum," in *Proceedings Volume 2351, Telemanipulator and Telepresence Technologies*, Boston, 1995.

[2] N. Wingfield, "Popularity of Sony's PlayStation VR Surprises Even the Company," The New York Times, 26 February 2017. [Online]. Available: https://www.nytimes.com/2017/02/26/business/sony-playstation-vr-sales.html.

[3] R. M. P. M. d. Santos, *A utilização de QRCode na indústria vinícola*, Vila Real: Universidade de Trás-os- Montes e Alto Douro, 2017.

[4] Blippar Group, "Layar," Blippar Group, 02 January 2009. [Online]. Available: https://www.layar.com/news/blog/2009/08/29/layar-reality-browser-21-launched/.

[5] Wikipedia, "Invizimals," Wikipedia, 16 July 2017. [Online]. Available: https://en.wikipedia.org/wiki/Invizimals.

[6] L. Morgado, P. Reis, F. d. Carvalho, T. Ribeiro, P. Mestre and C. Serôdio, "Architecture for transparent helpdesk system with multimodal presence," in *IEEE International Conference in Service-Oriented Computing and Applications (SOCA)*, Perth, 2010.

[7] F. Cassola, L. Morgado, F. d. Carvalho, H. Paredes, B. Fonseca and P. Martins, "Online-Gym: A 3D Virtual Gymnasium Using Kinect Interaction," *Procedia Technology*, vol. 13, no. SLACTIONS 2013: Research conference on virtual worlds – Learning with simulations, pp. 130-138, 11 April 2014.

[8] InMERSE, "InMERSE Framework," ND, 12 April 2017. [Online]. Available: https://bitbucket.org/Apidcloud/inmerse-framework.

[9] L. Morgado, L. Fernandes, H. Paredes and F. d. Carvalho, "Separating Gesture Detection and Application Control Concerns with a Multimodal Architecture," in *14th IEEE International Conference on Ubiquitous Computing and Communications (IUCC-2015)*, Liverpool, 2015.

[10] L. Morgado, P. Cristovão, L. Fernandes, R. R. Nunes, P. Martins, H. Paredes, L. Barbosa, B. Cardoso e F. d. Carvalho, " Primeira armada da Índia: novo conceito de jogo misturando realidades aumentada e virtual, gestos finos e amplos," em *SciTecIN'15 - Science and Technologies of Interaction*, Coimbra, 2015.

[11] L. Fernandes, G. C. Matos, D. Azevedo, R. R. Nunes, H. Paredes, L. Morgado, L. F. Barbosa, P. Martins, B. Fonseca, P. Cristovão, F. d. Carvalho and B. Cardoso, "Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype," *Journal Behaviour & Information Technology*, vol. 35, no. 11, pp. 907-918, November 2016.

[12] L. Morgado, H. Paredes, B. Fonseca, P. Martins, R. Antunes, L. Moreira, F. d. Carvalho, F. Peixinho and A. Santos, "Requirements for the use of virtual worlds in corporate training - Perspectives from the post-mortem of a corporate e-learning provider approach of Second Life and OpenSimulator," in *iLRN 2016 - Workshop, Short Paper and Poster Proceedings from the Second Immersive Learning Research Network Conference*, California, 2016.

[13] L. Morgado, H. Paredes, B. Fonseca, P. Martins, Á. Almeida, A. Vilela, B. Pires, M. Cardoso, F. Peixinho and A. Santos, “Integration scenarios of virtual worlds in learning management systems using the MULTIS approach,” *Personal and Ubiquitous Computing*, vol. 779, no. ND, pp. 1-11, 2017.

[14] MEO, “MEO Go VR,” 02 January 2017. [Online]. Available: https://meogo.meo.pt/conhecer/realidade-virtual.

[15] Digi-Capital, “After mixed year, mobile AR to drive $108 billion VR/AR market by 2021,” Digi-Capital, 02 January 2017. [Online]. Available: https://www.digi-capital.com/news/2017/01/after-mixed-year-mobile-ar-to-drive-108-billion-vrar-market-by-2021/#.Wag7hj75jDc.

[16] TelecomTV Tracker, “Worldwide spending on Augmented and Virtual Reality expected to double or more every year through 2021, according to IDC,” TelecomTV Tracker, 03 August 2017. [Online]. Available: https://www.telecomtvtracker.com/insights/worldwide-spending-on-augmented-and-virtual-reality-expected-to-double-or-more-every-year-through-2021-according-to-idc-11230/.

[17] MIT Media Lab, “Responsive Environments Group,” MIT Media Lab, 02 January 2017. [Online]. Available: https://www.media.mit.edu/groups/responsive-environments/overview/.

[18] K. Matsuda, “Hyper-reality,” 02 January 2016. [Online]. Available: http://hyper-reality.co/.


15,56 C
33,94 C
32,13 C

06

Industry 4.0 – the telco perspective

Alberto Rocha, Altice Labs
alberto-m-rocha@alticelabs.com

Pedro Fonseca, University of Aveiro
pf@ua.pt

Smart products manufacturing supported by intelligent processes leads to a global digital business. For the telco service provider new opportunities arise not only on the natural provision of resilient and secure communications but also on the creation of vertical and horizontal solutions, including data science (real-time analytics, machine learning and artificial intelligence), collaborative suites and specialized cloud services for ICT.

Keywords

Fourth Industrial Revolution;
Digital Transformation; Manufacturing Chain;
FIWARE; FITMAN

Introduction

"Industry 4.0" is certainly one of the buzz words of today. Germany has started, in 2011, a government strategy to revive the industrial sector by introducing computer systems towards the concept of a smart factory. In 2015, the European Commission launched the framework "Digitising European Industry" [1], describing it as *"an open initiative aiming to create a sustainable ecosystem of Future Internet Enablers and Digital Platforms to grasp the opportunities emerging with the new wave of digitalization of EU industry and society"*. Since then, similar initiatives have been taking place in several European countries like *"Industria Conectada"* in Spain, *"Fabrique du Future"* in France, *"High Value Manufacturing"* in the United Kingdom, *"Fabbrica Intelligente"* in Italy and *"Indústria 4.0"* or *"ProduTech"* in Portugal.

The evolution

Usually called the "fourth industrial revolution", Industry 4.0 was defined as the fourth stage of the Industrial Revolution (**Figure 1**). The Industrial Revolution started around the end of the 18th century, with the advent of workplace mechanization. The invention of the first mechanical loom, in 1784, by Edmund Cartwright, marks the beginning of a new era in goods fabrication and the uprising of the industry as the privileged means to produce goods. The mechanical loom was soon to be adopted in large scale: the number of mechanical looms in England grew from 2400 in 1803 to around 250 000 in 1857 [2]. Along with the mechanical loom, two other technologies were fundamental for the booming of industrial production: the invention of the steam machine, in 1781, by James Watt, and the Jacquard machine, in 1804, the first programmable device that used punched cards to automate the production of textile patterns. Together, mechanical loom (productive device), steam power (energy) and programmable cards (control) allowed, for the first time, for machines to take the place of people in the production of goods.

The second stage of the Industrial Revolution occurs in the 19th to the 20th century transition, with the adoption of electricity as the main power source, replacing steam. At the same time, the discipline of Industrial Engineering emerges with the works of Frederick Taylor, seeking to improve the efficiency of industrial work, which Henry Ford will apply in the concept of the production line. The usage of electricity allowed the introduction of relay based automation.

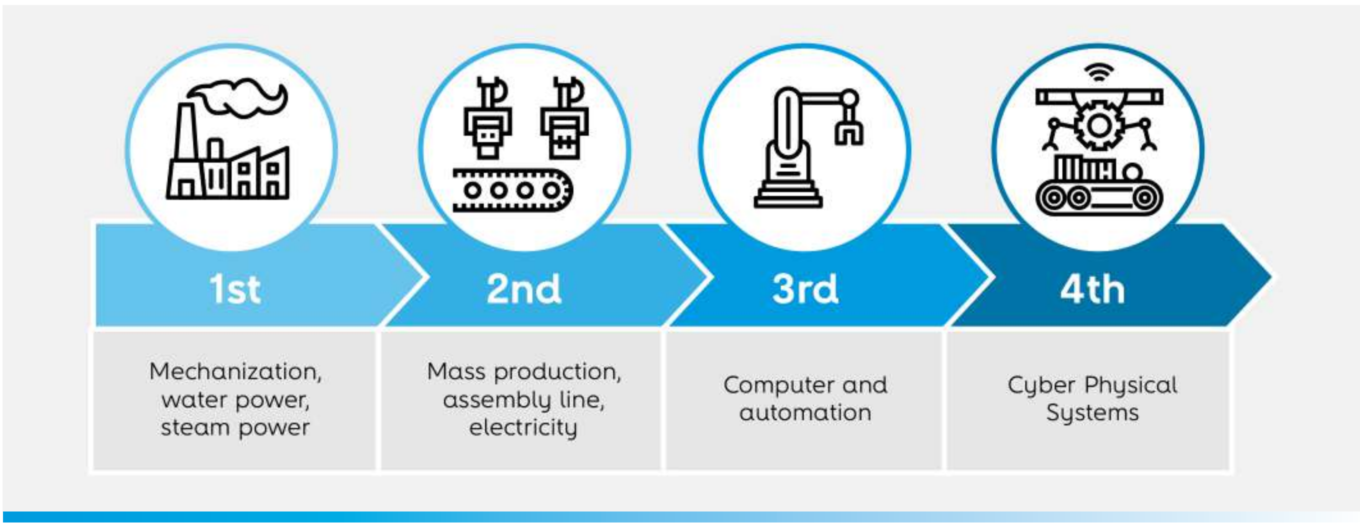


FIGURE 1 – Industrial Revolution evolution

By the 1970s, the developments in electronics, in particular, the possibility of producing cheap and reliable transistor-based circuits, open the way for the introduction of the Programmable Logic Controller (PLC), replacing hard-wired relays and cam-based systems with an electronic programmable device, creating the third Industrial Revolution. IT systems start to be applied to the factory floor and the first networks (or fieldbuses) to connect the controlling devices appeared.

The fourth Industrial Revolution, that we witness today, can be described as the intensive usage of Information and Communication Technologies (ICT) in the factory floor. It is fuelled by a combination of technologies that have become ubiquitous in the last few years: microprocessors and microcontrollers with ever-growing computational power, a global telecommunications network, providing wired and wireless connectivity nearly any time and anywhere, data storage of unprecedented size and scale, and algorithms with the ability to process this ever-growing amount of data. All at a monetary cost that does not cease to decrease, meaning that it is increasingly easier to have objects controlled by electronic devices, sending and receiving data wherever they are. This data will leave the record of its history in a data server somewhere in the world, which will be processed by some big-data algorithm, extracting features that, otherwise, would have stayed unnoticed. In fact, Industry 4.0, or any other of its names, is no more than the migration of the Internet of Things (IoT) concept to the factory floor, to become the Industrial Internet of Things (IIoT).

Industry 4.0 is based on the concept of Cyber-Physical Systems, or CPS, which are based on an electronic computational device (*cyber*) that has some means of interacting with its environment (*physical*). Cyber-Physical Production Systems consist of the application of CPS to manufacturing processes, creating *smart-objects* that transform the factory into a *smart environment*. This allows production to be configured in a much more flexible and agile way and creates new opportunities for new management and control processes.

Imagine a factory that produces different versions of the same product. Each version will have a different set of components, materials or colours, differences that must be considered by the production process. Imagine also a production line that contains feeders for all required components and assembly tools for all the production steps. Each piece to be produced may include a Radio Frequency Identification (RFID) chip that contains its unique identifier (serial number) and the recipe for its production: what parts, what colours, what operations, etc. That information allows each workstation in the production line to automatically reconfigure itself, in order to produce that exact piece, in all of its details. This way, the maximum flexibility is achieved and no extra costs are incurred to produce lots of one-pieces. At the end, the unique ID is also associated with all the production data for that piece, like temperature, insertion forces or the torque applied to each screw, etc, allowing for unprecedented levels of quality control.

The strategy for Industry 4.0 requires the implementation of three features [3]: horizontal integration, through the development of inter-company value chains; end-to-end engineering, covering the entire value chain; and vertical integration, linking business processes to the smart manufacturing shop.

Horizontal integration, depicted in **Figure 2**, allows the integration of different production plants in the world, where smart-objects are carried through the production chain, creating a coordinated global production facility. Also, it will allow correlating different parameters enabling to gather new knowledge that will enrich the eco-system (improving processes and productivity, reducing costs) [4].

End-to-end engineering relates to digital integration of all engineering process, creating a digital twin for all physical production lines while incorporating the customer requirements. In this way, companies can use virtual and augmented reality to get feedback from their customers, as John Deere is doing by testing new designs with their customers using virtual reality, decreasing cycle time from 27 to 9 months with cost savings of over $100k [5].

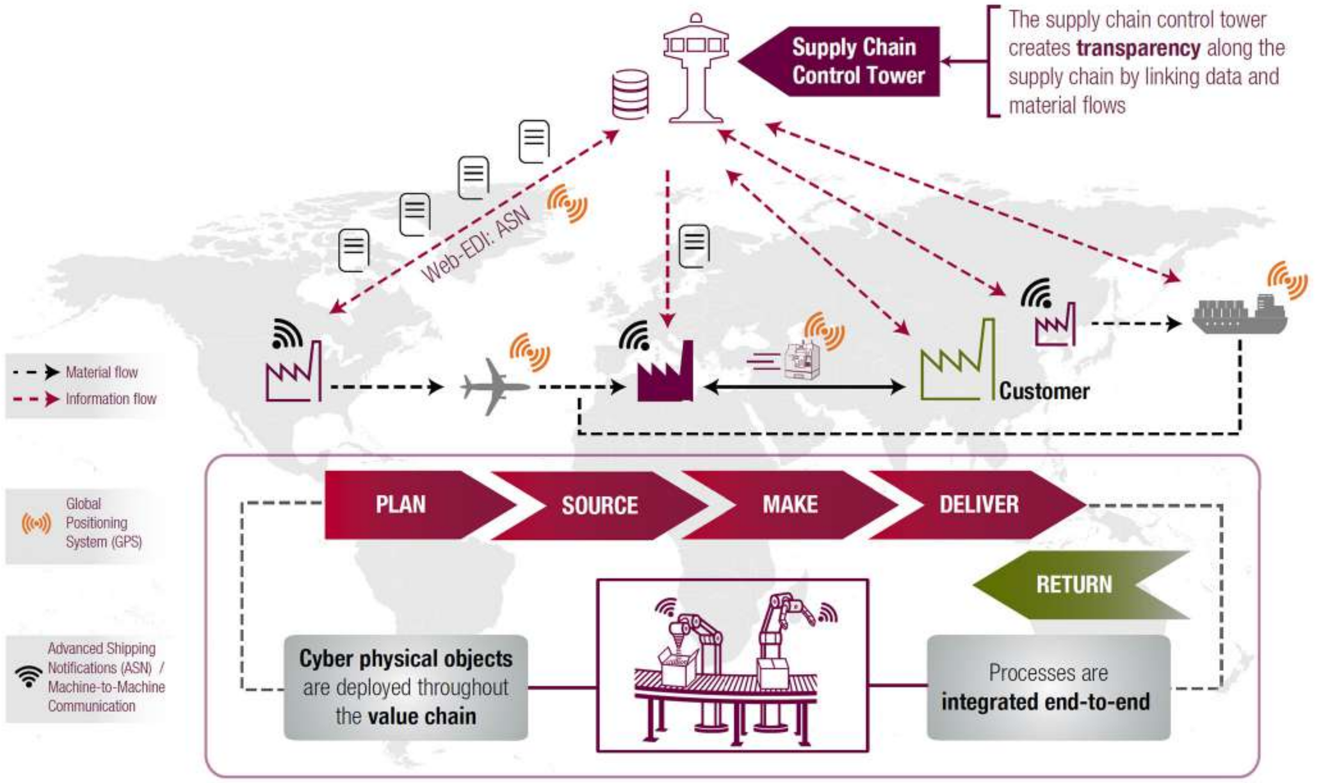


FIGURE 2 – Horizontal integration of global manufacturing chain [4]

Finally, **vertical integration** implies the total interaction of the management and business processes and the controlling devices at the factory floor, meaning that the production shop knows the customer orders and organizes accordingly, and the business and management layer is permanently aware of the conditions and state of the production line.

The new requirements for flexibility push for new forms of production flexibility. In the smart factory, the smart-objects will bid their production: they will send their production requirements to a set of production centres that may be in the same plant or in other plants. Each production centre will reply with its cost, delivery delay and other relevant production data that the bidder will use to decide the production site. The logic of manufacturing is radically changed and the concept of a self-organizing production system emerges. Smart factories could also take into account the outsourcing of a set of departments while keeping the focus on their core business: IT, marketing, purchase, distribution and other non-critical departments could change, depending on the 3rd party offered advantages.

Industry 4.0 is about interconnecting devices. Such a degree of interconnectivity is only possible if there is a common ground for all devices to communicate. Open standards will play an irreplaceable role in the advent of these new forms of production, as they are the only way to guarantee healthy competition among all players, avoiding artificial barriers. Initiatives such as FIWARE [6], which started with the European Union (EU) funded project “Future Internet Core Platform” during the 7th Framework Program, and the Open Platform Communications Foundation Unified Architecture (OPC-UA) [7], are paving the way to a set of open standards upon which the Factory of the Future can be built.

Currently, the manufacturing process already incorporates traditional IT tools to assist solving a particular problem:

- Product development and engineering, like Computer Aided Design/Computer Aided Manufacturing (CAD/CAM), Computer Aided Engineering (CAE), Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), Product Data Management (PDM), Product Lifecycle Management (PLM);

- Product, planning and governance, like Warehouse Management Systems, Manufacturing Execution Systems, Advanced Planning Systems, Computerized Maintenance Management System, Governance Risk and Compliance.

However, the main feature is that they are standalone platforms with proprietary or limited API: as such, all the potential that could be obtained by the integration of available data is lost. The challenge is now the interconnection of all these silos to exchange, collect and relate data, for instance using a cloud-based big data platform acting like an integration bus.

The Industry 4.0 is changing the paradigm of manufacturing strategy. But the expected benefits are **increased productivity, improved quality, higher flexibility, higher speed and higher competitiveness**.

The technology

Technology is transforming the way industry works. The most advanced software and communication technologies are now being applied to transform the manufacturing ecosystem and enabling the Industrial digitisation. The main technology drivers are described next.

IoT

"IoT is constituted by the real-time capable, intelligent, horizontal, and vertical connection of people, machines, objects, and ICT systems to dynamically manage complex systems" [8].

The deployment of a network of **Sensors** and **Actuators** at the shop-floor level enables data collection from multiple sources and automated execution of work orders.

The deployment of a **Tag** and **Tracking** or other based sensors system enables life cycle management of a product. From QR codes to RFID thin chips, it is now possible not only to know the exact location of an asset in the logistics department, in the distribution path or at client's premise, but also collect Key Performance Indicators (KPI) metrics of that asset, enabling to proactively fulfil the Service Level Agreement (SLA) contracts in the support services.

The interface to networked devices at the southbound layer is supported by multiple communication **Protocols**, like RS-232, RS-485, 802.1x, ZigBee, Z-wave, Thread, 6LoWPAN, LTE, LTE-cat0, LTE M, NB-IoT or 5G, enabling the control of heterogeneous devices.

Cloud computing

Private and public cloud services provide many of the managed services that applications needed – ex: scaling and orchestration templates, application catalogues for rapid deployment, monitoring and logging, multi-tenant isolation, rich API for developers and a self-service portal for operations.

IT resources can now be offered by a specialized service provider as a commodity, on a subscription basis (opex), enabling the pay-as-you-grow model, reducing costs and increasing the efficiency.

Big data and analytics

Huge amounts of data are being produced not only at the shop-floor but also in internal business processes, in partner networks, as well as in social networks. Data Science applied to the processing of multiple data sources (originally stored in a structured, semi-structured or non-structured format) will produce rich information to be considered in the production, planning, marketing or executive level, constituting a strong competitive advantage.

Additive manufactoring

3D Printing tools enable the prototyping of CAD development projects, identifying conceptual or functional errors in an early stage and reducing costs and risk.

Augmented reality

Augmented Reality (AR) finds innumerable applications in manufacturing and production. The usage of cheaper smart glasses or HTML5 development without the need for external plugins are real today and used, for example, to locate an asset on a shelf or in a step-by-step guided work-orders with the help of a headset.

Simulation

Today simulation tools enable the virtual test of a product during its conceptual phase with high accuracy, based on mathematical models. However, the simulation can also be successfully used to test a manufacturing process or advertising campaign as well. This allows, at a very early stage, to fix eventual issues, reducing costs, time-to-market and improving quality.

Cyber-security

Industrial Security, like Security in general, should be seen in a holistic perspective covering physical and logical technology resources but also processes and people as well.

Security enforcement is typically implemented at the Wide Area Network (WAN) and Local Area Network (LAN) protection, and central servers as well, but now should also apply to devices and machines (e.g., PLC). Equipment with non-Ethernet interfaces (e.g., RS-232, RS-485), being more difficult to penetrate, could still represent a security break. Logging, Monitoring and Alarm (LMA) systems must be installed to help in a proactive prevention but also in a forensic analysis.

Embedded intelligence

Manufacturers need to implement built-in "intelligence" and communication capabilities into their products. This characteristic enables self-reconfiguration procedures according to its built-in information, reducing error and improving flexibility. For instance, this will also provide enhanced information for traceability and proactive maintenance. This type of intelligent devices has the capability to generate and store its own data, feeding analytic engines and producing useful information to support decision making processes.

The business

Digital transformation will re-shape the entire macroeconomy. The faster and better is product development, the quicker is time-to-market. Smart products manufacturing, supported by intelligent processes, lead to a global digital business [9].

The competitive advantage does not rely solely on the technological shift but in the business model innovation: digital transformation leverages new and disruptive business models (e.g., the predictability of maintenance).

A simple Business Model (BM) definition is prescribed as the way the organization creates, delivers and captures value [10]. New and enhanced BM can be achieved through:

- vertical value-chain integration – the interconnected workflow of devices, sensors and actuators, with Enterprise Resource Planning (ERP), PLM and Manufacturing Execution System (MES) tools, all generating data to the cloud-based analytical platform;

- horizontal value-chain integration – to enable the interconnection of machines to customers for service level monitoring and data collection, providing meaningful insights to partners and promoting a collaborative environment among the entities involved in the business;

- manufacturing as a service – a new data-driven service where the owner of the asset being produced doesn't operate the manufacturing equipment but pays for the number of produced assets;
- additive manufacturing services – from 3D printing prototypes to the short delivery times;
- predictive manufacturing – devices act as a collaborative community generating data, for continuous processing, to produce information for a better decision.

For the telco service provider new opportunities arise not only on the natural (and obvious) provision of resilient and secure communications but also on the creation of vertical and horizontal solutions, including data science (real-time analytics, machine learning, artificial intelligence), collaborative suites, specialized cloud services for ICT. Consulting Services to start rolling new Industry 4.0 projects is another opportunity where telco can offer services due to the strong knowledge on ICT. Finally, the onset of 5G aware devices can tie the client to convergent solutions provided by the telecom operator. NFV and SDN can also provide flexible and efficient solutions at lower prices to deal with Industrial needs.

The Industry 4.0 evolution path

The business processes digitisation requires a shift thinking in the organization and applies to the small, medium or large corporation because it impacts on processes and people.

The progressive adoption of new technological resources through the use of **Verification and Validation** (V&V) of small PoC projects can be a smart decision to avoid disruption of the traditional processes and the negative impact on business. The running PoC also shortens the necessary learning curve for the organization. The introduction of small changes (and its measurement) is also in line with the Kaizen philosophy largely adopted in manufacturing environments.

The national or regional Industry 4.0 initiatives encourage the creation of **Competence Centers**, with the mission of approaching universities and private companies, promoting the digital transformation through training, awareness, live demos and technical advisory.

The **Digital Innovation Hubs**, comprising Research Centres, Industrial Players, SME, Universities, Clusters, Investors, Associations, Governments and Public Institutions, Incubators and Startups, will enable SME with access to technology.

At present, digital manufacturing **readiness assessment** based on existing tools and specialized consulting services can help on the digitisation transformation.

FITMAN use cases

There are several commercial solutions available in the market claiming to be Industry 4.0 aware. However, the risk of being attached to a single provider, with intrinsic proprietary interfaces, must be considered. That's why standardization is so important in developing vendor interoperable solutions. European Commission runs several projects to address this issue, being Future Internet Technologies for MANufacturing (FITMAN) one of them [11].

FITMAN was one the 5 Use Case Trials projects selected in the 2nd phase of the Future Internet Public-Private Partnership (FI-PPP) program, aiming the provisioning of FI-PPP Core Platform, called FIWARE, with 11 industry-led use case trials which test and assess the **suitability**, **openness** and **flexibility** of FIWARE Generic Enablers while contributing to the Social-Technological-Economical-Environmental-Political (**STEEP**) sustainability of EU Manufacturing Industries. The FITMAN use case trials cover several manufacturing sectors such as automotive, aeronautics, white goods, furniture, textile/clothing, LED lighting, plastic, construction, machinery for wood, and manufacturing assets management.

FITMAN identified three domains of actuation: smart, digital and virtual. Several use cases have been identified for each one of these domains aiming to cover a broad range of manufacturing sectors as shown in **Figure 3**.

Smart Factory trials

- TRW (automotive supplier): industrial health and safety.
- Whirlpool (white goods manufacturer): improved shop-floor decision process within data rich environment.
- Piacenza Cashmere (textile/clothing manufacturer): cloud manufacturing.

Digital Factory trials

- Volkswagen (automotive manufacturer): PLM rump-up for reduced time-to-market.
- AgustaWestland (aeronautics manufacturer): support for management of documentation in the final assembly line (for predictive maintenance purposed due to customized solutions) and training services for blue collar workers.
- Consulgal (construction): efficient collaborative access to information.
- Aidima (furniture manufacturer): trends forecasting and collaborative product design.

Virtual Factory trials

- TANet (manufacturing resource management): networked business innovation.

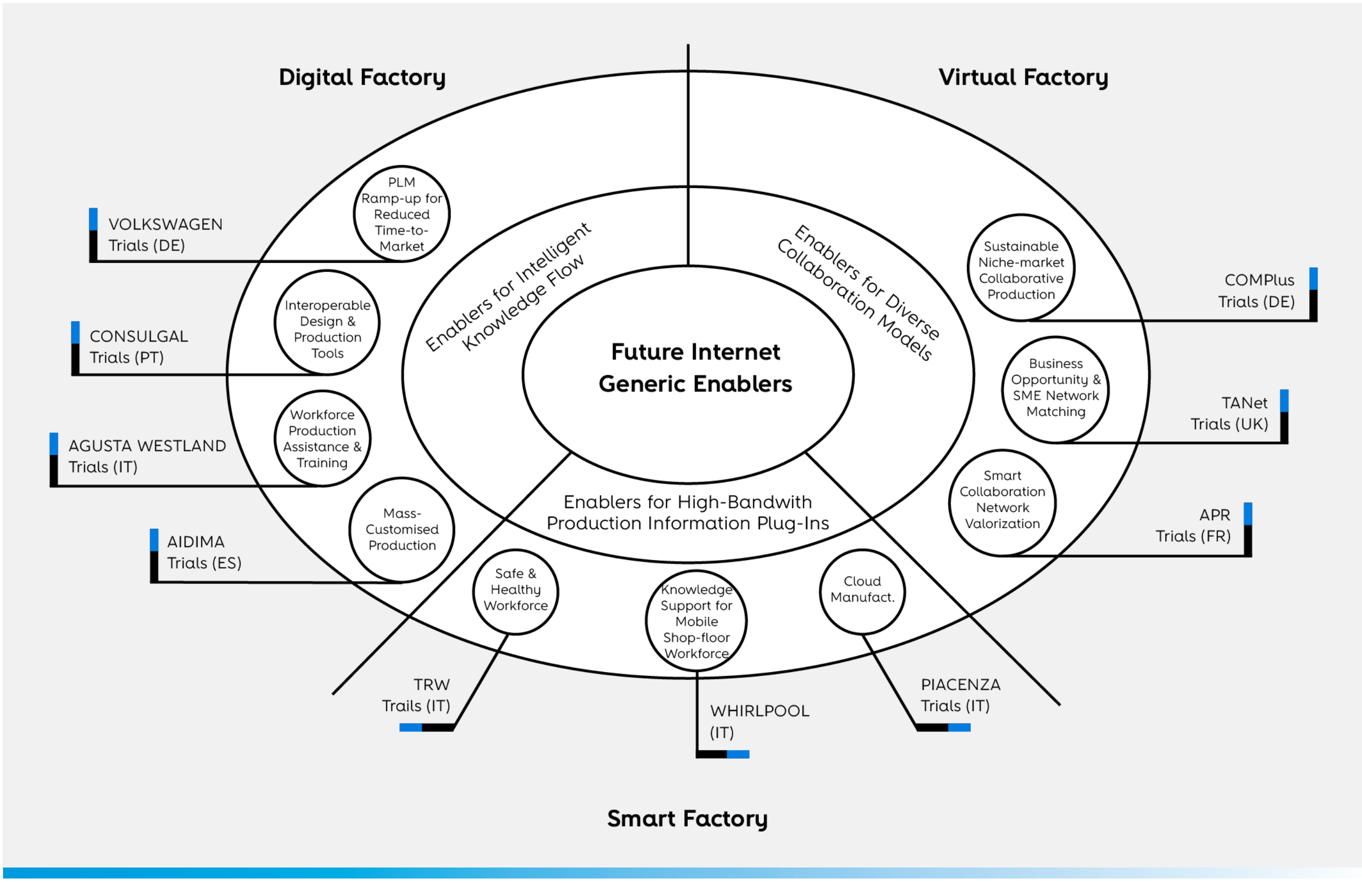


FIGURE 3 – FITMAN use cases

- COMplus (led-lightning): supplier selection, network management.
- APR (plastics): cross-enterprise collaboration.

The analysis of results show clear benefits in all trials, for example, better information is now available to help in the decision making, reduced time-to-market, lower manufacturing costs, identification of defective components during the manufacturing process, improved prevention of occupational accidents and process dematerialization.

The FIWARE generic enablers [12] and FITMAN specific enablers [13] are the foundations for the **FIWARE for Industry 4.0** initiative [6], an open platform that addresses most of the technical and business aspects towards a sustainable manufacturing ecosystem. Altice Labs was a key partner at FITMAN project, namely on developing components, hosting Specific Enablers on its own Cloud platform and promoting a workshop for FITMAN technology dissemination. In the end, the gains for Altice Labs were substantial, namely:

- The massive knowledge acquisition as a result of collaborative work with a large variety of experts from each partner;
- A new, now public, open source framework for the community to use and extend.

Conclusion

The paradigm of Industry 4.0 promises to revolutionize not only the way industrial plants operate but the entire value chains of the manufactured products, creating totally new business models.

Digital transformation enables the dematerialization of inter-department relationship, with orchestrated workflows, leveraging the outsourcing of non-core manufacturing sectors based on secure channels with partners. Since industry will delegate those functions to 3rd party work teams with specialized experts, that will surely add value to those areas. This model will also improve flexibility, where automation can easily be adapted to new tasks, new products, different volumes and diversified products.
From a broad perspective, some challenges for telcos in the Industry 4.0 arena are related with coordination tasks that will be critical because they require multidisciplinary teams working together. The trust between the involved partners and the existence of adequate standardization are also key for the solid emergence of Industry 4.0. Reducing manpower due to automation does not necessarily imply the reduction of production costs given that there will be not only significant investments on technology and licensing (capex) but also on the subscription of support contracts for maintenance (opex) as well. The return of factories to its originating countries could be a reality because the human costs are no longer "the" competitive advantage. Products with short life cycles can prevent automation gains since their requirements (design, production, maintenance, etc) compromise all the investment done requiring new manufacturing machinery.

In the technological core of Industry 4.0 lies a ubiquitous communication network, connecting all devices: not only the production line workstations but also the life cycle management of produced goods, all of them will reside in the global network. The first role of the telco provider will be to provide this connectivity but it does not stop there. The new business models created by Industry 4.0 and the usage of the production infrastructure as a service are an opportunity for telco providers to develop new services on top of their communication infrastructure. These include new services to provide horizontal and vertical integration, manufacturing as a service and data analytics. The new possibilities offered by the new business models in Industry 4.0 and the value they create will be captured by those that understand the changes and move first in the right direction.

Telecom operators realize that OTT players run their business on top of affordable, trusty, resilient and flexible networks because the value is on the offered service and not on the data transport. To engage customer loyalty in a 360-degree perspective and address that market share, telcos should design new and innovative vertical solutions for manufacturing and consider horizontal extensions to fulfil industrial needs.

References

[1] European Commission, "Digitising European Industry," European Commission, 17 August 2017. [Online]. Available: https://ec.europa.eu/digital-single-market/en/policies/digitising-european-industry.

[2] R. L. Hills, Power from Steam: A History of the Stationary Steam Engine, Cambridge: Cambridge University Press, 1989.

[3] H. Kagermann, W. Wahlster and J. Helbig, "Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 - Final Report of the Industrie 4.0 Working Group," Federal Ministry of Education and Research, Germany, 2013.

[4] J. Bechtold, A. Kern, C. Lauenstein and L. Bernhofer, "Industry 4.0 - The Capgemini Consulting View Sharpening the Picture beyond the Hype," CapGemini Consulting, 2014.

[5] B. Sniderman, M. Mahto and M. J. Cotte, "Brenna Sniderman, Monika Mahto, and Mark. J. Cotteleer," 25 May 2016. [Online]. Available: https://dupress.deloitte.com/dup-us-en/multimedia/videos/manufacturing-ecosystems-exploring-world-connected-enterprises.html.

[6] FIWARE, "FIWARE for Industry 4.0," 2016. [Online]. Available: http://www.fiware4industry.com.

[7] OPC Foundation, "Unified Architecture," OPC Foundation, 2017. [Online]. Available: https://opcfoundation.org/about/opc-technologies/opc-ua/.

[8] D. Kiel, C. Arnold, M. Collisi and K.-I. Voigt, "THE IMPACT OF THE INDUSTRIAL INTERNET OF THINGS ON ESTABLISHED BUSINESS MODELS," in *International Association for Management of Technology*, Orlando, Florida, USA, 2016.

[9] River Publishers Series in Communications, Building the Hyperconnected Society: Internet of Things Research and Innovation Value Chains, Ecosystems and Markets, vol. volume 43, O. Vermesan and P. Friess, Eds., Aalborg, Denmark: River Publishers, 2015, p. 331.

[10] European Commission, "FITMAN," European Commission, [Online]. Available: http://cordis.europa.eu/project/rcn/109803_en.html .

[11] S. Kaplan, The Business Model Innovation Factory: How to Stay Relevant When The World is Changing, Hoboken, New Jersey, USA: Wiley, 2012.

[12] FIWARE, "FIWARE Catalogue of Generic Enablers," 2015. [Online]. Available: https://catalogue.fiware.org/enablers.

[13] FITMAN, "FITMAN Catalogue of Specific Enablers," 2015. [Online]. Available: http://catalogue.fitman.atosresearch.eu.

07

Blockchain – a brief introduction

Filipe Relvas Bagulho Cunhal de Almeida, Altice PT
filipe-r-almeida@telecom.pt

Daniel Valente, Altice PT

Fausto de Carvalho, Altice Labs
cfausto@alticelabs.com

Paula Cravo, Altice Labs
pcravo@alticelabs.com

Blockchain requires an understanding of a lot of different domains and it's not easy to explain as a whole. This article explores the theme at a high level and using significant simplifications, aiming to provide readers with a basic understanding which they can use as a basis for exploring further resources to dive deeper in the technical aspects and potential applications.

Keywords

Blockchain; Cryptocurrencies; Bitcoin; Ethereum; Smart Contracts

Introduction to Blockchain

Blockchain is currently one of the most hyped buzzwords, both in technical contexts and in the generic everyday news, mostly for its strong association with the notorious Bitcoin cryptocurrency and for the much anticipated central role in the near future Internet evolution.

Blockchain concept is rather simple, albeit its remarkable possibilities: it allows the existence of tamper-proof distributed ledgers that can be efficiently maintained and kept up-to-date by a network of peers, not requiring any central authority to authenticate or validate transactions ordered by participants and their corresponding balances.

Blockchain requires an understanding of a lot of different domains and it's not easy to explain as a whole. This article explores the theme at a high level and using significant simplifications, aiming to provide readers with a basic understanding which they can use as a basis for exploring further resources to dive deeper in the technical aspects and potential applications.

First, we introduce the key components and features of Blockchain and we address how it works based on its most common application: a Bitcoin transaction example. The next section elaborates on the characteristics that enable other applications beyond cryptocurrencies, in a diversity of areas such as health records, smart contracts, electronic voting and proof of ownership for digital content, highlighting the importance of new approaches such as the more recent Ethereum blockchain. To wrap up, we will present some views on Blockchain impact in the fundamental shift from the Internet of information to the Internet of value, with assets being exchanged instantly in a global economy without intermediaries.

Blockchain key aspects

A concise but perhaps simplistic definition of blockchain is an immutable, distributed ledger visible to the community implementing and using it. Immutable means that the information a blockchain contains cannot be changed. Distributed means that the information is replicated amongst many participants (in Bitcoin terms, nodes). Ledger implies that the blockchain records transactions. Visible to the community means that every transaction recorded in the ledger can be made visible to every participant – user or implementer – of the blockchain [1].

A blockchain is a chain of data blocks, each one containing time-stamped transaction data and a validation of the previous block. This ensures the **integrity** of the whole chain of blocks and also guarantees that the chain cannot be tampered with, for that would invalidate the entire chain.

Besides cryptography, there's also an authorization system, based on digital certificates and public key systems, that prevents unauthorized access and authenticates whoever adds blocks to the chain or simply accesses them. The identity of transacting parties can be masked, so a certain level of **anonymity** can be achieved. The participants in a transaction have access to exactly the same records, so they can validate transactions in real time and verify identities or ownership without the need of a third-party intermediary, which gives the blockchain **transparency** and **auditability** properties.

One of the main features of the blockchain is the absence of a centralized authority to validate transactions/blocks and control the blockchain. In fact, the ledger is shared, updated with every transaction and replicated amongst all participants in near real time. The **distributed** ledger is not owned or controlled by any single organization, and thus the blockchain system is **sustainable** by being independent of controlling entities. All participants must agree to validate each block because the system is distributed and the agreement is achieved through a **consensus** mechanism.

All the described features combined prevent blocks from being deleted or modified and add an **immutability** feature to the blockchain.

The blockchain platform can incorporate business rules and algorithm-based transactions in such a way that business networks can evolve with flexibility. This is a key aspect to support end-to-end business processes and its evolution, in a wide range of activities and scenarios, like votes in a voting system, workflows in a workflow management system, supply chains, decentralized social networks, smart contracts, etc.

All in all, blockchain is a distributed, tamper-proof, append-only database, and one of its greatest features is that it solves the trust gap. In fact, there is no need to make any assumption of trust or involve third parties in order to process transactions and compute blocks because all records are visible and auditable by any member of the blockchain network.

How does it work

There is a number of different approaches and implementations, different technologic and architectural strategies, different purposes and many flavours among blockchains. These different approaches are beyond the scope of this article, as we are focusing only on the basic aspects of the technology.

Generically speaking, a blockchain is mainly comprised of the following components: a database or filesystem to store the blocks, a network for peer-to-peer (P2P) communication between the elements (nodes), a component to build, send and verify transactions to be stored, a distributed hashing algorithm to chain the blocks, and an API so clients can interact with the distributed system. Each node storing the blockchain must interact with other nodes in order to achieve consensus on new blocks to be added. This ecosystem is depicted in **Figure 1**.

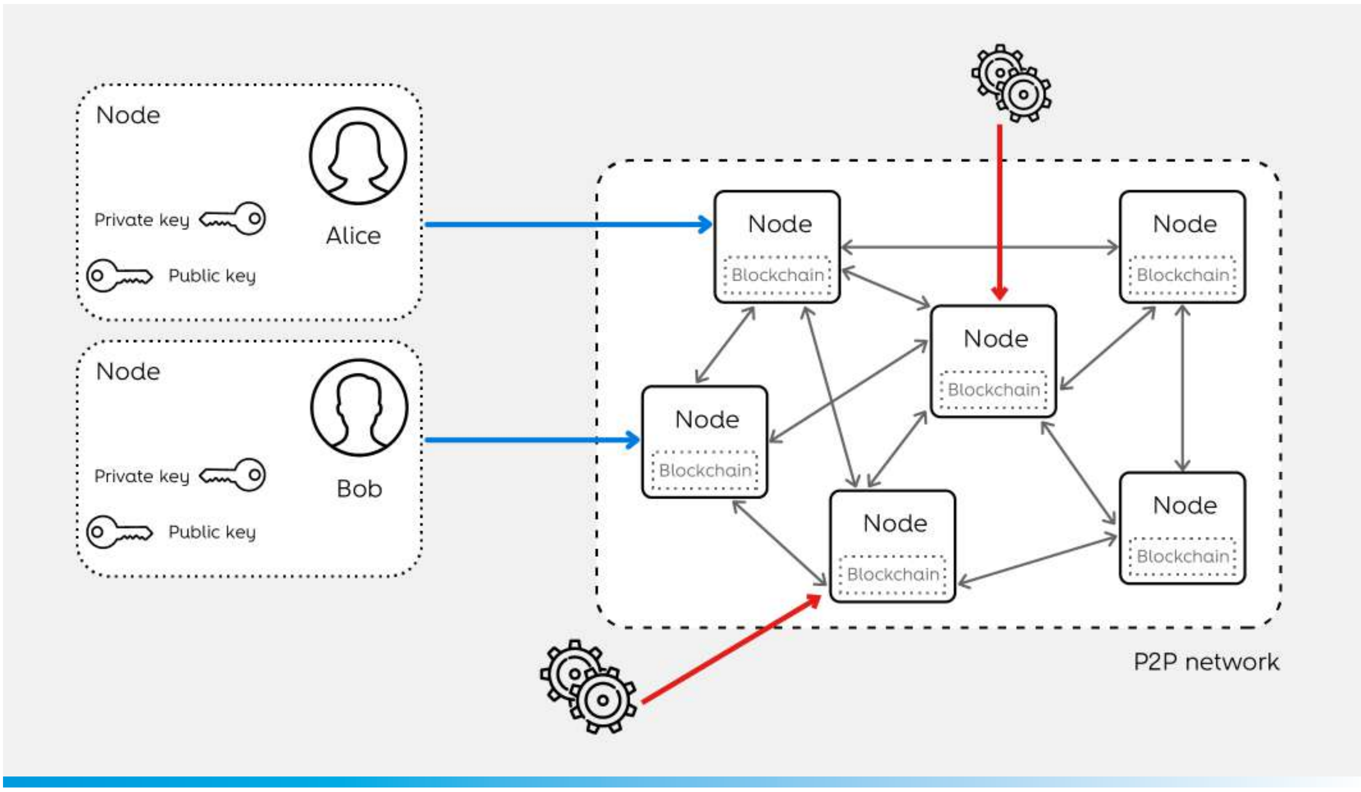


FIGURE 1 – Blockchain: ecosystem overview

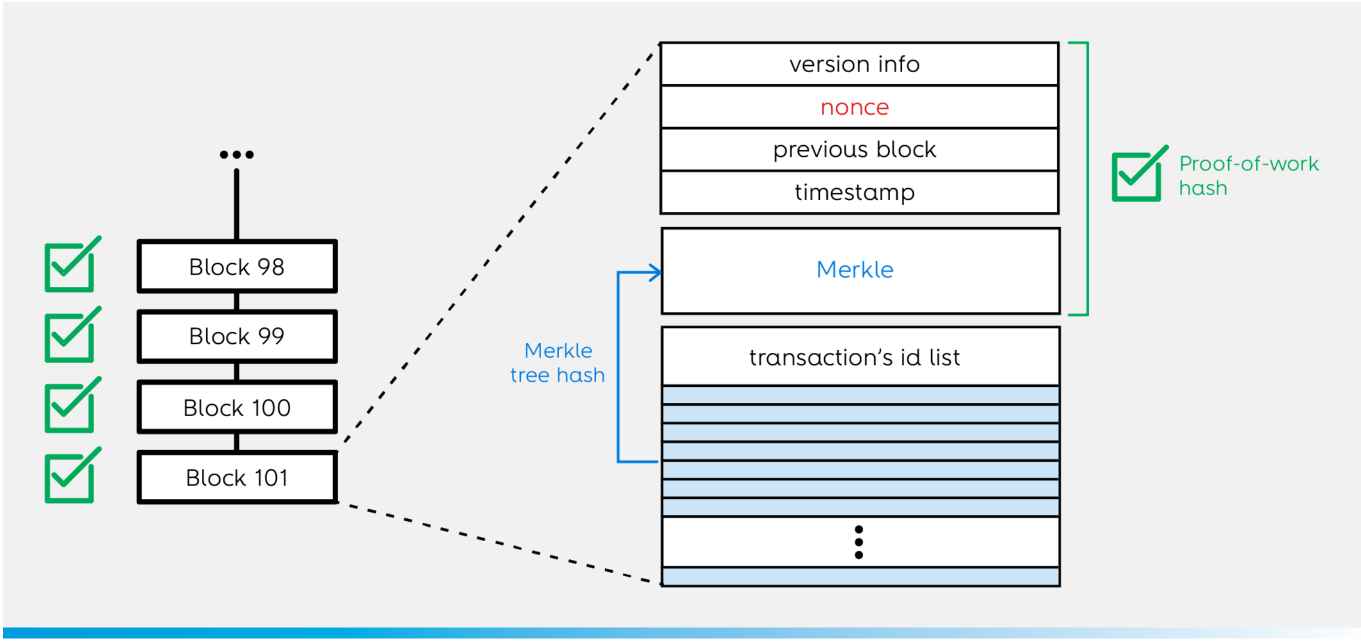


FIGURE 2 – Blockchain: block structure

Each block of the blockchain contains multiple transactions, cryptographically signed. The data model translates the state of the business transactions on the blockchain: the ordered record of transactions done in the system. The consensus amongst peers of the blockchain decides what gets written while guaranteeing the integrity of the data through hashing. That is achieved by participants that calculate new blocks, each including at least four pieces of metadata, as shown in **Figure 2** [2]:

1. The reference to the previous block;
2. The proof of work or proof of stake;
3. The timestamp;
4. The Merkle tree root (of transactions included in the block).

Bitcoin example

Historically, a cryptographic secure chain of blocks was first proposed by Stuart Haber and W. Scott Stornetta in 1991 to provide a means of controlling document modifications and improved in 1992 by incorporating Merkle trees to the blockchain. The proposed structure included the possibility of adding several documents into one block, but the type of robust cryptographic functions needed by the system took some years to be developed. When the mysterious Satoshi Nakamoto published in 2008 the whitepaper "Bitcoin: A Peer-to-Peer Electronic Cash System" [3], aiming to create a currency without a central governmental authority, and released Bitcoin software as open source code on January 2009 [4], a new type of currency was born: cryptocurrency.

Since Bitcoin was the first Blockchain application, let's use it as an example in order to understand how it works.

The Bitcoin data model is the history of all bitcoins produced and transferred between accounts since the beginning of time, Tuesday, September 1st of the year 2009 at 02:54:25 Zulu time (GMT). Transactions are token values transferred between accounts, and the consensus protocol is called proof-of-work: basically finding a digest with a given number of leading zeroes that is a hash of a number of transactions and the previous block hash reference.

Figure 3 shows a Bitcoin transaction between Alice and Bob. Both Alice and Bob have digital wallets associated with their balances, meaning that they each possess a pair of public-private

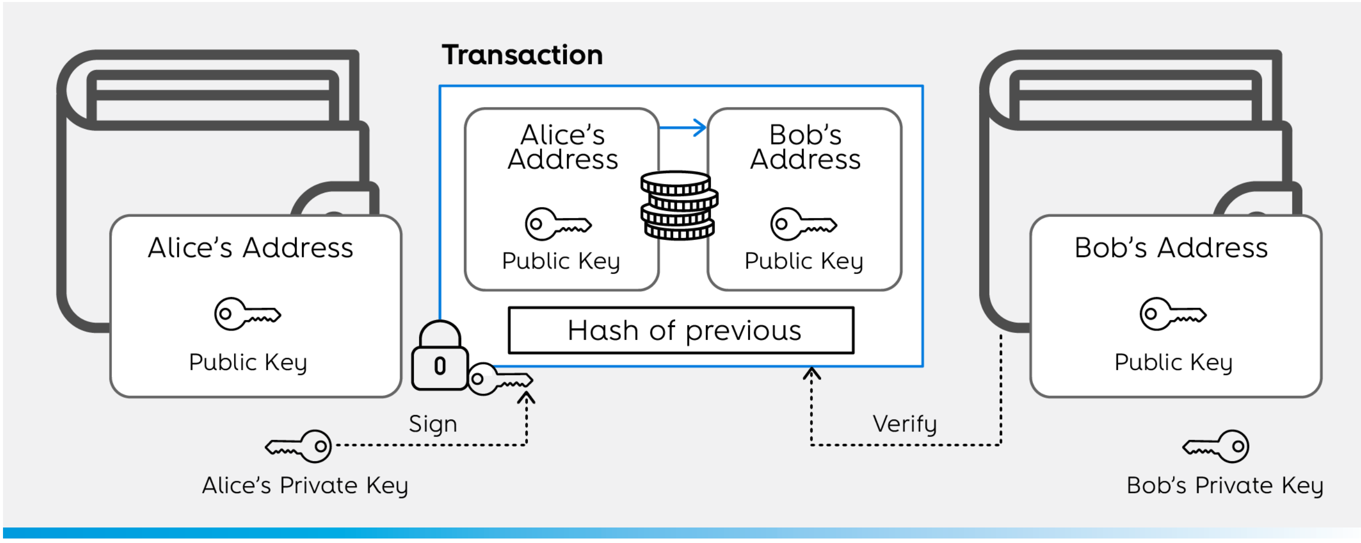


FIGURE 3 – Bitcoin transaction

keys, the public key being known by everyone and constituting the Bitcoin address of each user. Alice wants to send Bob a certain amount of bitcoins, so Alice sends a message from her Bitcoin address to Bob's Bitcoin address with the amount of coins to transfer, and signs the message with her private key. This constitutes a transaction. Each transaction can be verified by everyone with Alice's public key, so Bob can verify that it was Alice who sent him the coins. The transaction also includes a digest (hash) of the previous transaction, which is used as a pointer to where Alice got her coins from so that the flow of transactions can be verified. This signed message is broadcast to the Bitcoin nodes, the peers of the network. The nodes verify, with Alice's public key, that the transaction is valid and if she has credit (using the pointer provided in the message), and they set this transaction to be included in a future block. The transaction is only confirmed and concluded when it is included in a block of the blockchain. Only then can Bob spend the bitcoins Alice sent him.

If Bob wishes to send some amount of bitcoins to Charlie, he repeats the process, but this time, using Charlie's address. And if Charlie wishes to send some bitcoins to David, he can also do so. Transactions can be traced because of the hash of the previous transaction included in each transaction, so the process is transparent to all the parties, as shown in **Figure 4**.

All transactions are broadcast to the Bitcoin network, so all the peers can check what is happening at any time. Before being included in a block, transactions are kept in a pool, shared among all peers. As referred above, transactions are only validated once they are included in a block of the blockchain. The process of adding a block to the blockchain is called mining and goes as follows: each node, or miner, chooses a set of unprocessed transactions from the pool that will compose a block, adds the hash of the last block of the blockchain and tries to solve a puzzle. The puzzle is a resource consuming computation of a hash that has to meet a certain challenge, i.e., the hash value must be lower than a certain target value, with a certain number of leading zeros. The result of the hash computation is unpredictable and depends on the transactions selected, the hash of the previous block and a nonce (a random value, only used once, controlled by the miner), so the solution of the puzzle is a brute-force process where the miner tries different values of the nonce until it results in a hash lower than the target. This will be a solution for the puzzle, and it takes approximately 10 minutes to find (in fact the target is adjusted so that the puzzle solving is always about 10 minutes, long enough for the computation to be complex but not too slow, assuring security and efficiency to the process). The miner then broadcasts the new block, so that

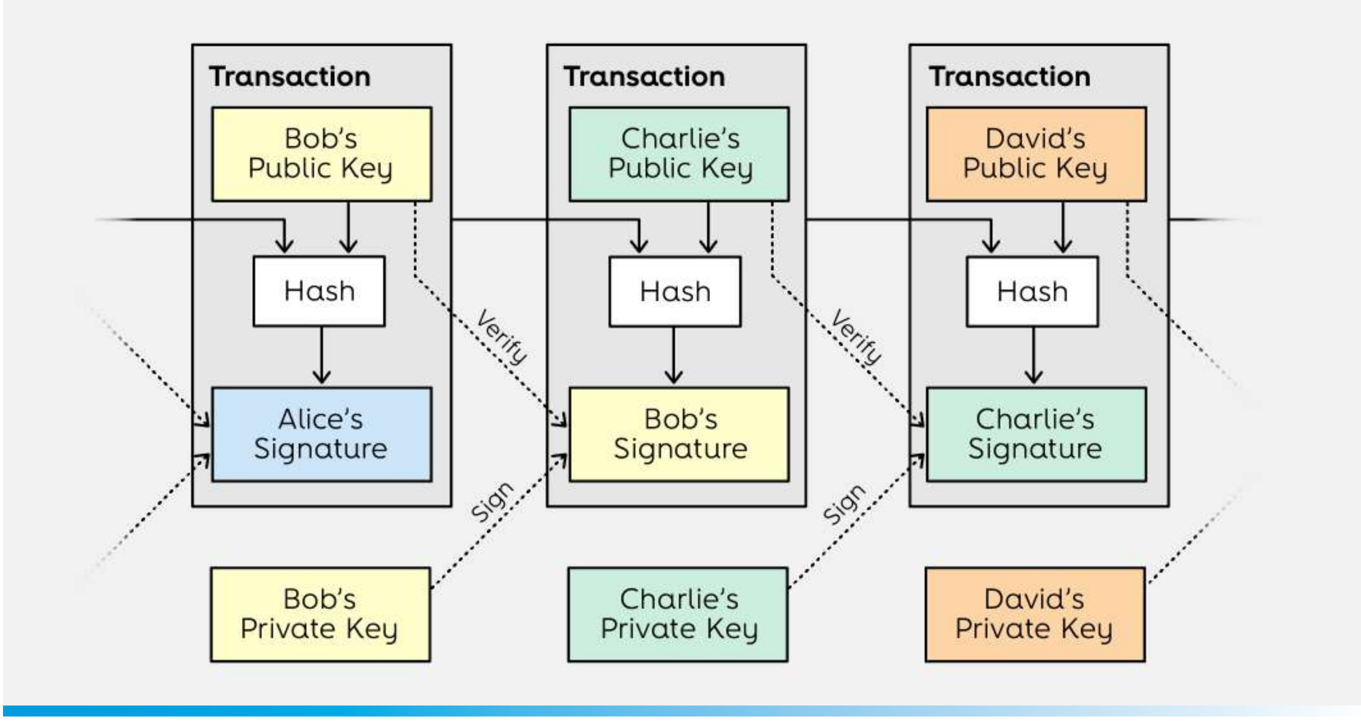


FIGURE 4 – Bitcoin tracing of transactions

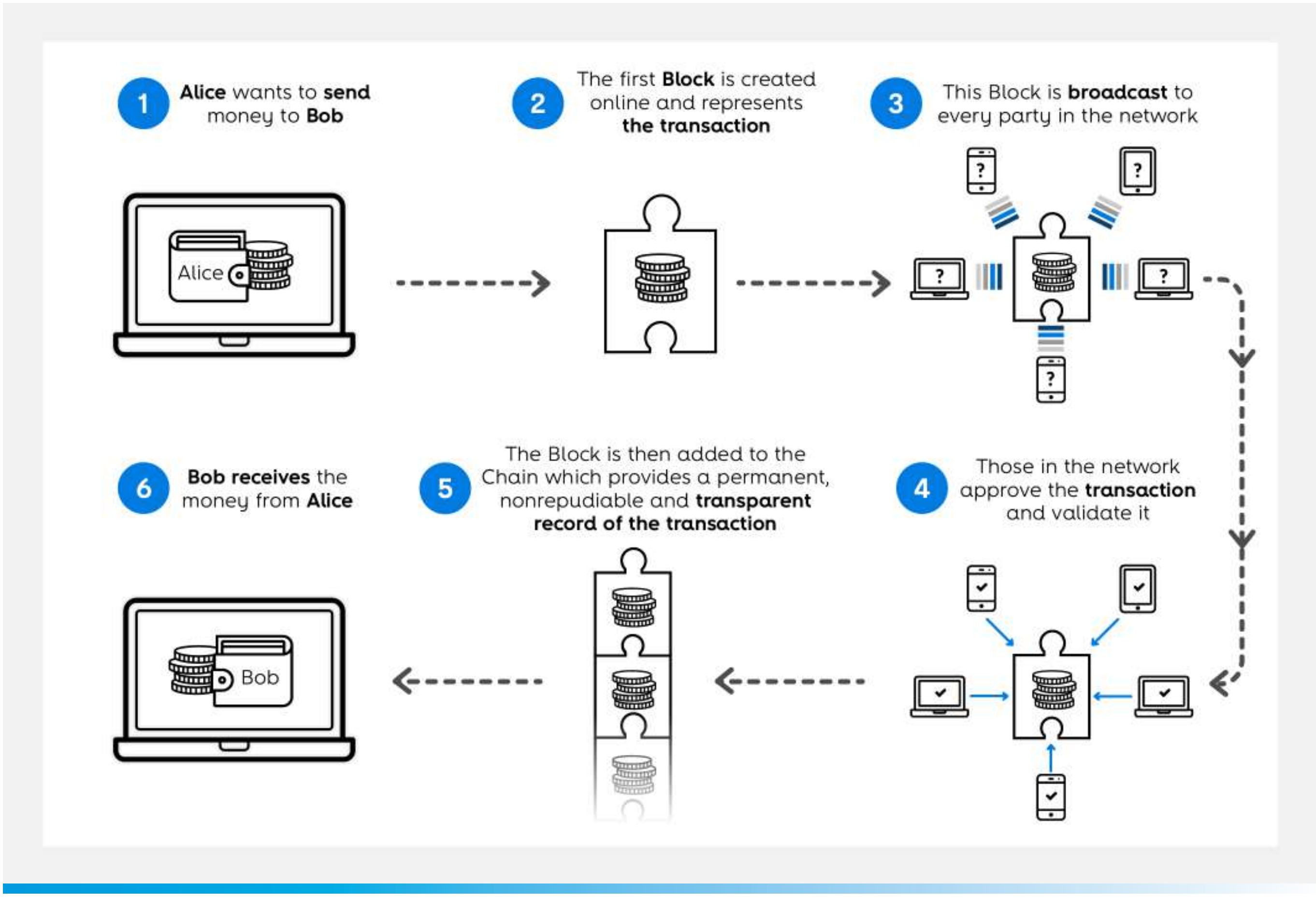


FIGURE 5 – Bitcoin mining process

all the other nodes know there is a new block to be added to the blockchain. In the Bitcoin system, this constitutes the proof-of-work. Bob receives his coins only when the block which contains the transaction is added to the blockchain because consensus occurred only in that moment, and the transaction was accepted by all nodes. **Figure 5** shows the sequence of procedures.

Tampering with a block already appended to the blockchain will be immediately detected because it will fail to validate; the hashes would be different from those on the majority and, even if the node was told to accept that, not being the longest valid chain according to the rest of the nodes, that history (chain) would be rejected by the rest of the network.

As of November 2017, Bitcoin blockchain is over 140 gigabyte, which despite sounding big it's perfectly manageable by a common desktop computer. Miners have local copies of the full blockchain updated via P2P network, while usually users rely on light clients (via their digital wallets) which have simplified mechanisms and trust on third-party servers for transaction validation.

Blockchain impact

Blockchain has been praised as the most disruptive technology since the Internet or the smartphone. Leaving all the technical aspects behind and stating it in a simple way, blockchain is an append-only tamper-proof distributed database, which makes it harder for digital data to go rogue. Blockchains are promising because they can make data trustable, by binding it into a strict framework of rules using cryptography.

From a software engineering point of view, Blockchain is a disruptive computing paradigm suitable to a broad number of use cases, giving the fact that it solves the trust issues on using a database, external service or business. It has the potential to fully represent a business in all its processes and interactions: customers, suppliers, authorities, etc, can be nodes of the distributed network, validating the up-to-date and transparent state of the business. **Figure 6** shows the main disruptive computing paradigms, placing Blockchain in the current decade [5].

In the previous sections, we used a permissionless blockchain implementation as an example, but blockchain can be easily coded with a permission-oriented design. The permissioned blockchain leverages some of the core elements

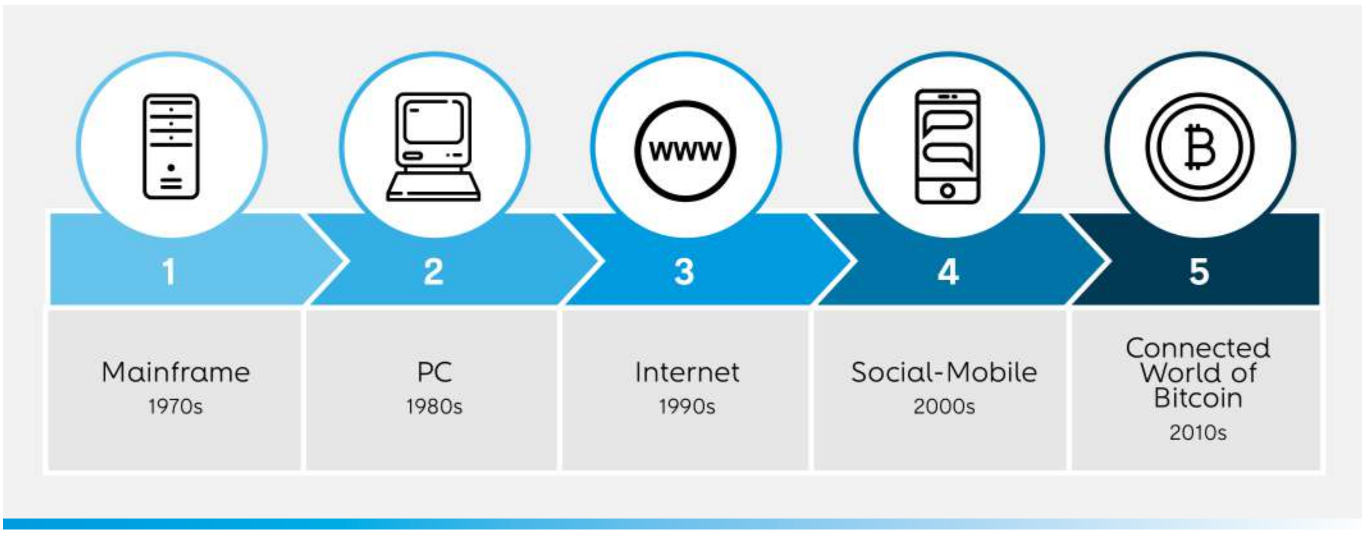


FIGURE 6 – Blockchain, a new computing paradigm

of the Blockchain architecture, like immutability, ability to grant granular permissions, automated data synchronization, rigorous privacy and security capabilities, and process automation, which in turn allows for the creation of new data-driven business models, where data is the primary decision-making tool. A permissioned blockchain restricts the actors who can contribute to the consensus of the system state by only a set of users with the rights to validate the block transactions. Beyond that, a permissioned blockchain has other characteristics that are of interest to many companies: **privacy**, meaning that only allowed actors can view transactions; **scalability**, because there are only allowed actors, there's no burning computational power for the proof-of-work model consensus but instead it uses a proof-of-stake model for consensus [6]; and **fine-grained access control**, for it allows restricted access to the data within the ledger.

In many enterprise use cases, a permissioned blockchain can meet business requirements that are impossible to meet with permissionless blockchain. The Blockchain system can represent a business model by including in its design the business transactions, so that a read-through of the blockchain will result in the current state of the business, for example storing a log on the blockchain and then using machine learning on top of it: rigorous information, unprecedented insight.

In the past eight years, this led to a flourishing ecosystem where we can find the likes of popular Ethereum [6], Waves [7], Linux Foundation's Hyperledger [8] or Polkadot [9] and over a thousand other projects. The most well-known implementations of Blockchain are cryptocurrencies, tackling the obscure slow and expensive transfers of value by banks through the internet. Bitcoin was the first and currently executes around 300 000 transactions daily from Afghanistan to Zambia. Many others exist, such as Zcash, Litecoin, Monero, Ripple or Dash. In fact, at present, there are over 1200 different cryptocurrencies available, although it may be difficult to trade with the less-known ones [10].

Initially, the most common use of cryptocurrencies was the remittance market - transfer of an amount of money across the world - which usually is solved in minutes so that the new owner of bitcoins (or any other) has only to sell the value for fiat currency at an exchange, bank or locally between natives. But right after, cryptocurrencies started banking the unbanked people in undeveloped countries (e.g. Zimbabwe) or under economic struggle (like Venezuela) and people have started adhering to Bitcoin and even living of it.

Cryptocurrencies have also been posing a threat to Venture Capital (VC), a much-discussed theme at the Web Summit this year. People creating startups submit to Initial Coin Offerings (ICO), also known as Token Generation Events (TGE), where a startup generates a digital token and trades it for money or other cryptocurrencies. Not being regulated, the funding is frequently promptly available for the startup to use at will, unlike the more complex and phased VC processes. By October 2017 ICO already yielded over 3 billion USD to startups worldwide [11].

Many of these ICO tokens have been generated over networks like Ethereum or Waves, mentioned above, which provide a tool called Smart Contract. These smart contracts are full-fledged programs that enforce If-Then clauses and transfer value (tokens) according to real user interaction and data. Many implementations already exist, from bricks-and-mortar retailing to insurance or supply chains, for example, AXA's flight ticket insurance. The product is being pitched as a "smart insurance" tool that flyers can use to insure their trips if their flight is delayed by two or more hours. As such, the product makes notable use of smart contracts, a self-executing piece of code that triggers once certain conditions are met on a blockchain [12].

Lastly and most importantly, blockchains are decentralizing businesses and shaping a new form of the Internet, where cloud, usually centralized and controlled by big companies, is expected to decrease in face of blockchain enabled applications, like Decentralized

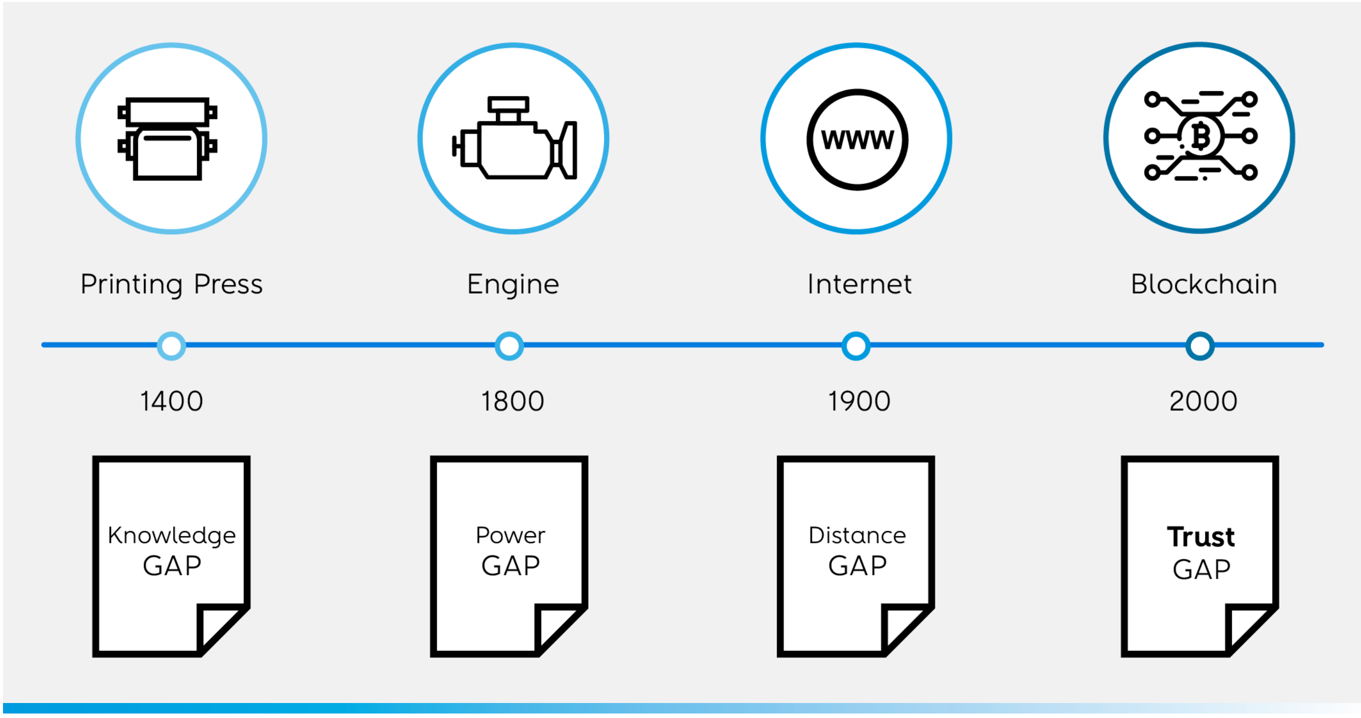


FIGURE 7 – The trust gap

Applications (DApp), distributed, resilient, transparent and incentivized applications which run everywhere and are controlled by no one. Many DApp already exist, most of them Ethereum-based. Ethereum introduces a Turing-complete language on a blockchain which means you can program almost anything on top of it, smart contracts which the whole Ethereum network maintain and execute [13].

Smart contracts enable trust among parties involved in agreements or transactions. They are automated and virtually free of manual error, and they eliminate unnecessary intermediaries. This is a keystone in migrating application-centric ecosystems, to give explicit control of digital assets to end-users and remove the need to trust any third-party servers and infrastructure, and as so, disruption is impending [14].

The history stored in a blockchain implementation will always be the truest possible since no actor can by itself change what has been stored there without the consensus of at least the majority of the players, nor will there be a system or database administrator capable of tampering with the information. With Blockchain, data trust issues are solved and this is considered the most disruptive property of blockchain. In fact, throughout Mankind history, certain inventions had undeniable major roles in the evolution of society, and certainly, Blockchain will have its place in this evolution, as shown in **Figure 7**.

Trends and conclusions

ICO, Bitcoin and cryptocurrencies, in general, have been frequently making news headlines and getting a closer look from financial institutions and regulators, amidst concerns that point to the introduction of some regulatory measures to oversee cryptocurrency exchanges and to control the growth in digital tokens. On the other hand, technology enthusiasts worry that too much regulation may stifle innovation and hamper the exploration of the multiple opportunities envisaged around Blockchain.

The programmability of second-generation platforms such as Ethereum, with powerful characteristics like smart contracts, DApp and permissioned blockchains, broadened its applicability to a diversity of areas beyond currency and payments. There is an ongoing emergence of real-world use cases being discussed or tested, leveraging Blockchain technology with the potential of deeply impacting our lives and lifestyle. Examples include public records (e.g. birth certificates, criminal records, health and safety inspections, property titles, voting), personal records (e.g. certifications, degrees, contracts, wills), health (e.g. genome data, electronic health records), physical asset keys (e.g. hotel rooms, vehicles, houses, lockers, package deliveries) and intangibles (e.g. software licenses, vouchers, DRM licenses, vouchers, copyrights).

Telecom operators naturally have a high interest in Blockchain or, more broadly, Distributed Ledger Technologies (DLT), like many companies in varied sectors already exploring its potential to make common transactional processes cheaper, faster and more transparent. The role they will play in their internal platforms digitalization is expected to rise as they embrace virtualisation of network functions, yet challenges to be addressed for the adoption of DLT by operators involve scalability issues, integration with legacy OSS/BSS, often including interfaces with third parties, increased security risks and compliance, and the need for standardization which will certainly take years.

There are several interesting opportunities to consider and some operators are already working together with vendors, partners and startups, aiming to capitalise on future blockchain-based applications and services. Relevant use cases include internal OSS/BSS processes (e.g. billing and number portability databases), hybrid blockchains for supporting roaming, connectivity provisioning using autonomous Machine-to-Machine (M2M) transactions, digital assets micropayments, mobile money, identity management, secure storage and transmission of health data, and services and applications for smart cities [15].

Blockchain is an information technology still in its pubescence and we are just starting to scratch the surface in terms of the potential outreach of its disruptive features. It is undoubtedly changing the world as we know it, pushing a shift from the Internet of information to the Internet of value, adding an economic layer by which assets can be exchanged instantly, without intermediaries playing a major role in a new global economy of immediate value transfer.

Despite all the hype and although left out of this article for the sake of simplicity, we must be aware that there are several questions, issues and limitations to be addressed in Blockchain technology introduction. For example, a problem recently reported on Ethereum network highlighted the risk of human error, since data stored on a blockchain is not inherently trustworthy – garbage in, garbage out [16]. Another aspect often referred is transactions actual cost and, in that sense, energy figures associated with Bitcoin mining are hardly disregardable [17].

Nevertheless, there are reasons to stay optimistic in face of the ongoing evolution, as new interesting applications are being introduced. One strong direction is Blockchain harnessing the integration and automation of human/machine interaction with M2M and Internet of Things (IoT), enabling a powerful machine economy. Another promising work area is Artificial Intelligence (AI): Blockchain allows distributed coordination of global marketplaces for specialized intelligent agents. Consensus mechanism could even foster a very interesting scenario of a most wanted "friendly AI" enforcement...

And more, much more, so everyone equipped with the knowledge and the right skills can hardly wait for Blockchain tenure.

References

[1] S. Goeringer, “A Simple Overview of Blockchains - Why They Are Important to the Cable Industry,” in *2017 Fall Technical Forum, SCTE-ISBE, NCTA*, CableLabs, Denver, 2017.

[2] Bitcoin Project, “Bitcoin,” Bitcoin Project, 2009-2017. [Online]. Available: https://bitcoin.org/en/.

[3] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2008.

[4] S. Nakamoto, “Original Bitcoin Sourcecode,” Satoshi Nakamoto, 2009. [Online]. Available: https://github.com/trottier/original-bitcoin.

[5] M. Swan, Blockchain, O’Reilly, 2015.

[6] Ethereum Foundation, “Ethereum,” Ethereum Foundation, 2017. [Online]. Available: https://www.ethereum.org/.

[7] Waves Platform, “Waves,” Waves Platform, [Online]. Available: https://wavesplatform.com/.

[8] The Linux Foundation, “Hyperledger,” The Linux Foundation, 2017. [Online]. Available: https://www.hyperledger.org/.

[9] Parity Technologies, “Polkadot,” Parity Technologies, 2017. [Online]. Available: https://polkadot.io/.

[10] CoinMarketCap, “Cryptocurrency Market Capitalizations,” CoinMarketCap, 2017. [Online]. Available: https://coinmarketcap.com/.

[11] Coinschedule, “Cryptocurrency ICO Stats 2017,” Coinschedule, October 2017. [Online]. Available: https://www.coinschedule.com/stats.php.

[12] AXA, “AXA goes blockchain with fizzy,” September 2017. [Online]. Available: https://www.axa.com/en/newsroom/news/axa-goes-blockchain-with-fizzy.

[13] State of the DApps, “State of the DApps,” State of the DApps, [Online]. Available: https://www.stateofthedapps.com/.

[14] M. Ali, “The next wave of computing,” August 2017. [Online]. Available: https://medium.com/@muneeb/the-next-wave-of-computing-743295b4bc73.

[15] E. Velasco-Castillo, “Analysys Mason: Nine blockchain opportunities that telecoms operators should explore,” 13 June 2016. [Online]. Available: http://www.analysysmason.com/Research/Content/Comments/nine-blockchain-opportunities-Jun2016-RDMY0/.

[16] S. Schroeder, “Wallet bug freezes more than $150 million worth of Ethereum,” Mashable, Inc., 08 November 2017. [Online]. Available: http://mashable.com/2017/11/08/ethereum-parity-bug/#fAQAs4J86mqo.

[17] Power Compare UK, “Bitcoin Mining Now Consuming More Electricity Than 159 Countries Including Ireland & Most Countries In Africa,” Power Compare, 20 November 2017. [Online]. Available: https://powercompare.co.uk/bitcoin/.

08

Big data and recommender systems

Luis Cortesão, Altice Labs
luis-m-cortesao@alticelabs.com

Simão Reis, Altice Labs

Within the multitude of content available, allowing an easy and enjoyable way to make a choice is a major advantage a TV provider may provide, but it requires an in-depth knowledge of the user interests and preferences. Recommender systems based on machine learning, whose processes and techniques are presented in this article, have been thoroughly tested and validated in several business areas and, despite some constraints, will have a role in providing solutions allowing accuracy, performance, scalability and timely recommendations.

Keywords

Recommender Systems; Machine Learning; User Profiling; TV Services

Introduction

Big data is quickly becoming a critical and important business success driver across different sectors, but most companies are not yet prepared to make the most of it. One of the key aspects in extracting value from data is the use of smart analytics – advanced algorithms that can add value and make sense out of vast information pools. It has been proven that big data analytics enables substantial profitability and productivity benefits when companies have a clear data strategy aligned to their business objectives. McKinsey [1] reports that the average business investing in big data analytics experiences an initial 6 per cent increase in profits, rising the longer the investment continues.

Digital product consumption is presently growing immensely. It is growing to such extent that those e-commerce systems' consumers are being overwhelmed by the quantity of available products and information (known as the information overload problem) [2]. Similarly, television has evolved greatly in recent years and each day clients have more available channels, more content, and more interactive services. With such amount of available content, it becomes difficult to choose what to see. Without an easy and enjoyable way to make a choice, this multitude of content becomes a disadvantage rather than an advantage. Easing the selection process and focusing on the personal interest of each user is a benefit to any TV provider. But to successfully help the user and recommend targeted content, as well as benefiting the TV provider, it is necessary to really know the user, his interests and preferences.

Data filtering mechanisms emerged as an answer to this problem.

The most well-known mechanisms are recommender systems (or recommender engines). Content recommenders are generally used to help consumers find relevant content (e.g. movies on Netflix or goods on Amazon). Recommender systems seek to predict product preference by a client. It is the purpose of the recommender system to filter and present the consumer only relevant data to him, based on his consumption history. Today's content recommender systems are rather narrow and typically aim at profiling the user and provide recommendations based only on previous consumptions, or through comparison with other users who expressed similar preferences.

In this article, we intend to present relevant techniques and algorithms for user profiling (i.e., for building a client model using his historical consumption data) and how they can be used for analytics and recommendations purposes. We will use a TV recommendation's scenario as a concrete example. We will present some challenges and requirements of this specific case and will follow it throughout the article. In the end, the reader should be able to understand the functioning of a recommender system, how to build a user model and the challenges associated with TV recommendations. Additionally, we will briefly introduce the STREAMLINE project [3], from the European Union (EU) research and innovation program Horizon 2020 (H2020), where we are testing different approaches for recommenders – profile based, machine learning (ML) based and hybrid approaches – based on a common architecture.

User profiling

The main purpose of profiling is to obtain information about users and their interests. In a TV scenario, a viewer may be described by the characteristics of the shows he/she views/follows, like the release year, synopsis keywords and more usual features like actors, director and genres. We call this the User Model (UM), the abstract representation and aggregation of data about a user and his preferences.

There are two main approaches for profiling: explicit and implicit. Explicit user profiling uses data gathered directly from users input, where the system asks explicitly about the data it needs. This approach is losing interest as users are generally not interested in directly giving their inputs. As such, implicit profiling has been gaining interest over the last years. Implicit profiling relies on information learned from the user interaction with the system. For this reason, it is also known as behavioural profiling. Usually, this type of profiling makes use of filtering techniques we will discuss later.

TV recommendations have many particular challenges [4]. Explicit ratings from the viewers are usually not available and implicit ratings must be computed based on behaviour history. Recommendations are targeted, usually not to an individual, but to a whole family, whose preferences may differ greatly. The programs' catalogue is dynamic, each day new shows are available and the recommender must be able to deal with these new shows every day. Lastly, the recommendations are shown on a TV screen with limited interactability, as such, the menus should be simple and intuitive.

User rating

Rating is a relative measurement on how much the user liked an item compared to others. If explicit rating from the user is not available, we may estimate the rating the user would give to an item he already consumed (visualization history). In a TV scenario, the viewing time is the most natural choice for computing this estimation. The number (or frequency) of visualizations may be another relevant indicator.

When we have access to external data sources (social networks or databases) we can improve the accuracy of the rating with explicit data. We could employ the user's Facebook likes on movies, Twitter tweets, Internet Movies Database (IMDb) ratings, etc.

Recommender systems

A recommendation may be viewed as a prediction problem, where given a user and an item the system must return a predicted rating [5].

$$Users \times Items \rightarrow Ratings$$

The main objective of recommender systems is to present interesting content for users in the form of content recommendations by filtering the content less interesting to them. Recommender systems have two main approaches: Content-based Filtering and Collaborative Filtering (CF) [6]. Content-based Filtering aims to use item features to find and recommend other items with similar characteristics to those a user likes. CF assumes people with similar tastes will agree on the future, recommending items that matching users liked.

Figure 1 illustrates one possible data flow architecture for a Hybrid Recommender System. The main component of a recommender system is the Recommendation Engine. The Recommendation Engine receives data about the preferences over items of the target user - and eventually also from other users' preferences - and produces a list of the top recommended items. The illustrated engine makes use of a Hybrid Engine which, in turn, uses multiple other engines, by merging, filtering and sorting the results returned by the different engines (Content-based Engine and CF-based Engine will be explained later). The Rating module converts data of user preference over an item to a numerical rating (explained in section User Rating). The Data Partitioning module optimises the data to be used on the CF-based Engine (in a process that will be explained later). Through the remaining of this text, we will follow the rationale of this Hybrid Recommender System in detail.

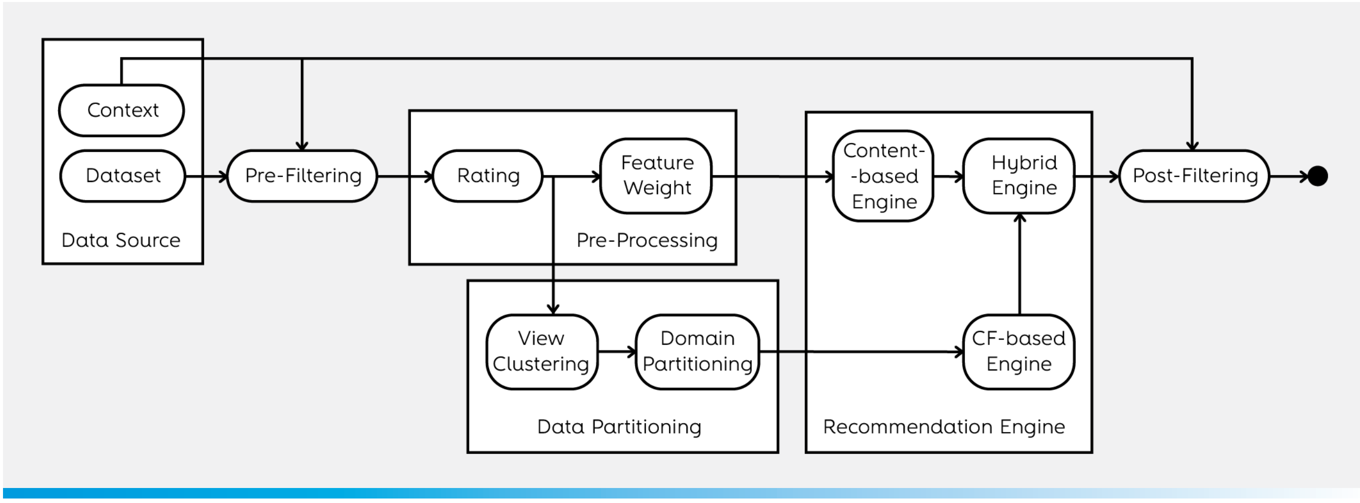


FIGURE 1 – Hybrid Recommender System data flow diagram

Content-based recommendations

The user is expected to have an interest in items similar to those he liked. A Content-based recommender learns the important item features for the user from the items he liked and builds a UM representative of his preferences. Candidate items are compared through their features with items rated positively by the user and the most similar are recommended to him.

A UM is typically a weighted vector of item features. The weight represents the importance of each item feature to the user. These features may be obtained with statistical analysis or may be discovered by a ML technique. Using the feature weight of a user, we may predict the recommendation rating of an item not viewed by the user.

It takes time for the system to learn what's important to the user. Therefore, to avoid the User Cold-start Problem, i.e., no historical data from a new user, some systems obtain initial data through a survey during first use [7]. In addition to the User Cold-start Problem, Content-based Filtering has the disadvantages of requiring information about each item features, does not adapt well to change of interests by the user, and lacks serendipity (i.e., the effect of accidentally finding interesting contents when looking for something entirely unrelated) [8].

Collaborative-based recommendations

CF algorithms are by far the most popular method of recommendation. CF-based recommenders have their basis on the idea that if two users rated items in a similar fashion (similar rating for the same items) then they have similar preferences. Users with similar preferences are grouped and are recommended items they have not rated, but rated positively by other users in their group.

Compared to content-based techniques, CF has the advantage of not needing items description. It is closer to the basic philosophy of ML techniques than content-based techniques, in that features are discovered automatically rather than explicitly given by the user. This allows for the discovery of more subtle aspects that relate users to each other.

There are multiple approaches of CF, namely: User-based Filtering and Item-based Filtering as more traditional approaches, and Context-aware Filtering and Cross-domain Filtering as more modern approaches. We will describe them in the next sections.

User-based collaborative filtering

In a User-based CF, a user's opinion for an item is predicted by looking at users' overlap in opinions for other items, providing a ranking of the most similar users. Items rated highly by similar users are recommended. To increase accuracy, a weight is given to each item based on how similar the other users who rated the item are to the user target of recommendation. This approach is appropriate for TV recommendation scenarios where new TV programs air every day and TV viewers usually maintain their TV service provider for years.

User-based CF has some disadvantages: it suffers from the Sparsity Problem, where each user rates a small fraction of available items, making it hard to match users; it doesn't scale well, as for a large number of item ratings by users, the prediction computations becomes slow and, in big data scenarios, it is unfeasible to take into account every other user's preferences; lastly, it suffers from the Item Cold-start Problem, for it cannot predict ratings for a new item until some users have rated it.

Item-based collaborative filtering

In an Item-based CF, a user's opinion for an item is likely to be the same for similar items. Contrary to content-based recommendations, to find similarity between items we don't consider item features, but rather how other users rated items. K-Nearest Neighbours (KNN) algorithm may be used to find historically similar items to reduce computational costs. Item-based CF was developed by Amazon to deal with their scalability problems of user-based recommendations [9].

This approach scales well when the number of items is usually much more stable than the number of users. It also prevents the User Cold-start Problem, for if a user rates one item we can start searching for similar items. Like the User-based approach, Item-based CF has the disadvantage of suffering from the Item Cold-start Problem, having more impact in this approach.

Context-aware collaborative filtering

Context-aware recommender systems adapt to the specific contextual situation of the user. Some contextual factors may be time, location or purchasing purpose, for example.

The main properties of contextual factors are knowability, observability and their change over time (static or dynamic) [4]. With dynamic factors, it may happen that, over time, one factor is no longer relevant and must be dropped.

Formally, a Context-aware recommender system tries to estimate the rating function

$$\mathit{Users} \times \mathit{Items} \times \mathit{Contexts} \rightarrow \mathit{Ratings}$$

where *Contexts* is a set of contextual factors that constrain the rating of an item by a given user. The context adjustments may be made at two levels: filtering ratings that are irrelevant in each context, or adjusting the ranking of recommendations. For the first case, an example would be a person who likes to watch action movies and could be recommended only to action movies that are airing or will be airing later the same day. Context-aware recommendations may take multiple forms like Contextual Pre-Filtering and Contextual Post-Filtering.

In the Contextual Pre-Filtering approach, context information is used for selecting the relevant dataset subset that will be used as an input for prediction in the traditional CF methodology.

$$(Users, Items) \xrightarrow{Contexts} (Users', Items') \xrightarrow{Recommender} Rating$$

The context serves as a data filter and allows for the use of existing recommender systems. The contextual filtering only transforms the input data. It doesn't interfere with the base behaviour of the engine. A use case would be TV shows recommendations: if a user wants to see a movie on Saturday, only Saturday watching data is provided as an input to the recommender.

In the Contextual Post-Filtering approach, the dataset is processed using traditional CF and the resulting recommendations are altered accordingly to the contextual information.

$$(Users, Items) \xrightarrow{Recommender} (Users', Items') \xrightarrow{Contexts} Rating$$

Cross-domain collaborative filtering

User's similarity is discovered based on the preferences they share. However, users' preferences in a certain domain may not match to preferences on other domains. If two users share the same preferences in literature, that doesn't mean their preferences in movies are similar. Likewise, two users may like the action genre, but one may like horror movies when the other doesn't. Traditional CF compares users in a single domain. Cross-domain systems calculate user similarity on each domain, defining local neighbourhoods of users. Then, the recommendation engine constructs an overall neighbourhood based on local neighbourhoods, making predictions and recommendations based on the overall network [6].

Optimization through clustering

TV recommendations suffer from two main issues. The first being efficiency because data is usually very large therefore has very slow processing. The second being data quality, since data is usually poor due to the number of items being much larger than the number of ratings provided by the users, leading to a very sparse rating dataset. This means that there are few aspects that allow us to correlate users.

The rating dataset is a bi-dimensional one because each user gives one single rate to each item. A simple clustering technique groups similar users and is known as horizontal clustering. The users will become partitioned by their similarity in items they've rated. This will result in an amount of columns being empty. For example, in the case of TV recommendation, the columns related to shows not rated by any viewer in the group will be empty. By cropping those columns will lead to a greater density in the resulting dataset and a smaller volume of data.

Instantiation within STREAMLINE

Altice Labs is involved in STREAMLINE project, a EU H2020 project that aims to reduce complexity, enable faster results, and reduce cost by supporting analysis on "big data at rest" and "fast data in motion" in a single system. In the scope of this project, Altice Labs is providing a use case based on Internet Protocol television (IPTV) recommendation, where it plans to test different approaches for recommendations – profile based, ML based and hybrid approaches – based on a common architecture.

The actual global architecture of the system, presented in **Figure 2**, is divided into two major parts: the inputs and outputs are represented on the left side of the diagram, while the entire data processing pipeline is represented on the right side.

The left topmost block, Web crawling, represents information collected from the web that can bring new value and improve the profiling and

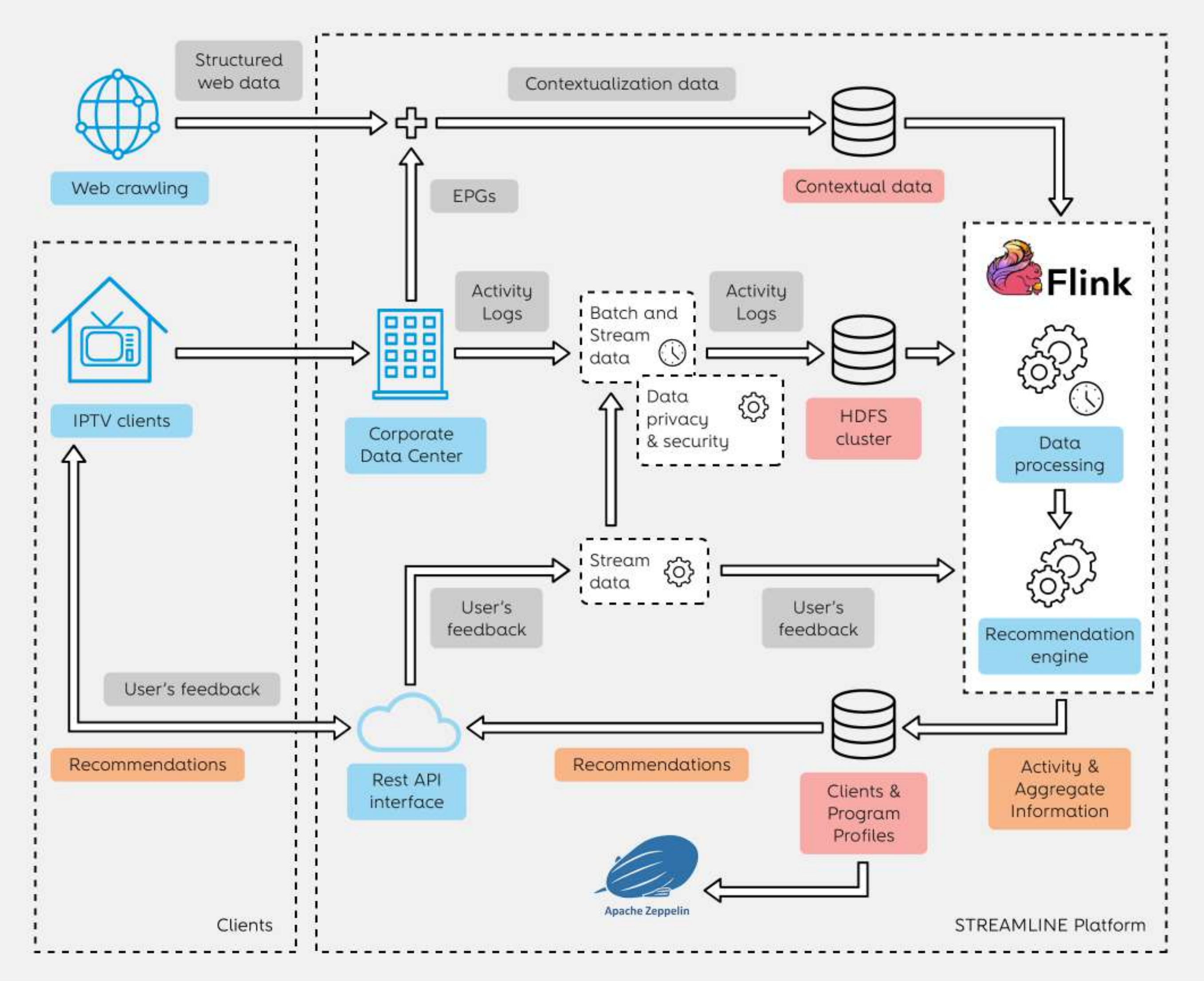


FIGURE 2 – STREAMLINE use case architecture

recommendation engines, i.e., enhanced contextual data. The clients behaviour input is gathered in the form of activity logs produced from their interactions with the set top box and sent to the STREAMLINE platform for storing and analysis. Apart from the activity logs, there are two other data streams represented in **Figure 2**: the Recommendations stream, pertaining to the recommendations sent to each client, and the Users' feedback stream, with the feedback each client provides regarding the received recommendations. To guarantee full protection of data, all the data transferred between the clients' set top box and the platform is encrypted.

The corporate data centre represented in **Figure 2** is the node of the platform responsible for collecting information originating from the clients. As soon as new activity logs are available in the data centre the platform collects, processes and stores that data in a Hadoop Distributed File System (HDFS) cluster, thus allowing the system to have a full history of all clients' activity recorded in the set top boxes. This data subsequently becomes an input to the recommendation engine. Additionally, contextual data is also used as input for this engine. This type of data includes not only external knowledge and information collected from web archives but also detailed information from TV channels and programs obtained from the Electronic Program Guides (EPG).

The development of the recommendation engine is part of an intra-collaboration between project partners, where the goal is to apply and adapt supervised or unsupervised machine learning approaches to model clients' preferences and thus provide them with suitable recommendations.

Conclusions and next steps

IPTV scenarios are one rich field for the exploitation of smart analytics targeting the end user TV consumption experience improvement and therefore adding business value to the TV provider.

Throughout this article, we presented processes and techniques that have been thoroughly tested and validated in several digital product consumption business areas. However, for the IPTV business area, there are some constraints that should be taken into account.

Applicability of CF based recommender processes to TV Streaming business has been validated and documented, but is still impaired by some empirical and business results in IPTV business. For instance, explicit feedback is a relevant input requirement for the presented algorithms, but not entirely feasible given the specific characteristics of the TV ecosystem. Therefore, implicit rating, i.e. feedback originating from the analysis of viewer activity will have increased relevance and might add some bias or error to the implemented processes. In addition, the volatility of the TV catalogue is also a challenge expected to be a factor with relevant impact on the algorithms stability and the overall quality of the results.

The work being executed within STREAMLINE is focused on the usage and optimization of CF based algorithms. Collected results of the implemented solution – namely performance and accuracy – will be evaluated and will allow to identify limitations of those approaches and define alternatives to best cope with the viewer usage circumstances, performance and accuracy expectations, as well as its adequacy to several scenarios (for example, multiple tier operations, investment constraints, cold-start relevance, etc).

Given the above-referred constraints, several alternative approaches are being designed, implemented and tested. Some effort is being applied in an analytical based approach, which will mainly rely on the viewer usage profile, supported on client and/or program relevant characteristics (i.e., genre, timeslots, etc.). It will use data aggregations to identify profiles with common characteristics and will allow timely recommendations based on similar viewer usage. This approach benefits from a reduced computational effort but might skip some relevant composed characteristics that are essential to a more in-depth and accurate profile computations, which are inherent to the ML based processes.

The results from the different approaches will be compared and evaluated, and it is expected that this evaluation will allow identifying their relative adequacy to different business scenarios. More relevantly, it will provide some insights to compose hybrid approaches (for example, analytical with ML), with high accuracy, performance, scalability and timely recommendations, as well as flexibility to adapt or enhance some of the different processes according to the usage.

One additional field of work, which is quite relevant given the business interests of the Altice Group (telecommunications, media, advertising, financial services), may be the effort in extrapolating the profiling and recommendation processes to allow a more thorough characterization of the clients within their used services, from the universe of services provided by the group. Cross service profiling might allow tapping non-evident cross-selling opportunities and the subsequent impact on ARPU and retention. ■

References

[1] J. Bughin, "Big data: Getting a better read on performance," 01 February 2016. [Online]. Available: http://www.mckinsey.com/industries/high-tech/our-insights/big-data-getting-a-better-read-on-performance.

[2] S. Berkvosky, T. Kuflik and F. Ricci, "Distributed Collaborative Filtering with Domain Specialization," in *RecSys '07 Proceedings of the 2007 ACM conference on Recommender systems*, Minneapolis, MN, USA, 2007.

[3] H2020 Streamline Project, "STREAMLINE," European Union's Horizon 2020 , [Online]. Available: http://h2020-streamline-project.eu/.

[4] Z. Çataltepe, M. Uluyagmur and E. Tayfur, "Feature selection for movie recommendation," in *Turkish Journal of Electrical Engineering & Computer Sciences*, Turkey, 2016.

[5] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," in *Recommender Systems Handbook*, New York, NY, USA, Springer, 2015, p. 191–226.

[6] H. Jafarkarimi1, A. Tze Hiang Sim and R. Saadatdoost, "A Naïve Recommendation Model for Large Databases," *International Journal of Information and Education Technology*, vol. 2, no. 3, pp. 216-219, 2012.

[7] D. Asanov, "Algorithms and Methods in Recommender Systems," Berlin Institute of Technology, Berlin, Germany, 2011.

[8] F. Ricci, L. Rokach and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, New York, NY, USA, Springer, 2011, pp. 1-35.

[9] G. Linden, B. Smith and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76-80, 2003.

09

AI: overview and applications

João Pias, Altice Labs
jpias@alticelabs.com

Pedro Miguel Neves, Altice Labs
pedro-m-neves@alticelabs.com

Luís Cortesão, Altice Labs
luis-m-cortesao@alticelabs.com

Janis Zemitis, DFKI
janis.zemitis@dfki.uni-kl.de

César Analide, University of Minho
cesar.analide@di.uminho.pt

Considering that AI “was born” in 1956 and lived for nearly 60 years without known relevant applications, why all this hype around AI now? The fact is that technological advances have brought AI out from the shadows of academia and research, into the business environment. The accelerated pace of technological evolution has paved the way for AI to be implemented at industrial scale and every player willing to have a relevant technology market role must adapt and embrace AI.

Keywords

Artificial Intelligence; Voice Assistants; Machine Learning; Deep Learning; Narrowband Artificial Intelligence; Autonomous Management System; Marketing Campaign Engines; Recommender Systems

Introduction

After years and years of progress with varying degrees of success, artificial intelligence (AI) is spreading out of the research labs and definitely moving into the commercial world, being used in all kinds of areas and taking a leading role under the technology spotlights. Everyone willing to have a relevant technology market role in the near future seems to be already using or planning to use AI shortly.

This article aims to introduce AI technology and applications, beginning with a brief historical overview and an approach to the multiple ways how it is defined and put under perspective by different protagonists of its evolution till nowadays.

Current hype is analyzed next, with a state of the art review of AI technology, highlighting machine learning (ML), deep learning, speech recognition, intelligent apps, natural language processing (NLP), natural language generation (NLG), predictive analytics, virtual customer assistants and computer vision.

Altice Labs is already addressing AI technology in the scope of some key areas and three applications are presented: AI applied to network management, application of AI to marketing campaign engines and automatic digital content recommendation.

This article concludes with a wrap-up restating the importance of AI in the near future and its game-changing nature for digital service providers (DSP).

Historical perspective

It has been more than 60 years since the expression AI was introduced in the common lexicon. The historic milestone symbolizing the first use of the expression "Artificial Intelligence" was established at the summer of 1956 meeting at Dartmouth College in Hanover, New Hampshire, USA, designated "The Dartmouth Summer Research Project on Artificial Intelligence".

However, at the origin of this "invention" is the document entitled "The Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", dated August 31, 1955, by John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon, a group led by the first and that carried out the organization of the meeting in the summer of 1956 [1] [2].

How AI has been defined

It is not surprising that several dictionaries come up with different proposals for AI definitions, being one of them "*the capacity of a computer to perform operations analogous to learning and decision making in humans, as by an expert system, a program for computer assisted design (CAD) or computer-aided manufacturing (CAM), or a program for the perception and recognition of shapes in computer vision systems*" [3]. Nor should it be surprising that within the scientific community there are several approaches to this topic. A few definition examples are provided below:

- John McCarthy, the "inventor" of the expression and considered by many the "father" of AI: "*[AI] is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable*";

- Elaine Rich and Kevin Knight, in their book "Artificial Intelligence" [4], present a simple definition: "*AI is the study of how to make computers perform things that, at the moment, people are better*";

- In Stottler Henke's glossary for AI [5], one may find a statement explaining that "*The great practical benefits of AI applications and even the existence of AI in many software products*

> *go largely unnoticed by many despite the already widespread use of AI techniques in software. This is the AI effect*".

This is, in our opinion, the noblest objective one can hope for in working in the research, development and innovation (R&D+I) field of AI: to expect that the results of our work will go unnoticed to its users, by interfering so non-intrusively in their lives that its benefits are intrinsically unconsciously enjoyed.

Until the centennial anniversary

The accelerated pace of technological evolution has been evidencing the feeling that a technological singularity (simply put, the creation of super-intelligent computer agents, capable of surpassing every human intelligence) will occur, approaching at a vertiginous cadence. Will it happen until the centennial anniversary? Some say so. Others contradict.

Raymond Kurzweil, known by the epithets of inventor, futurist and computer scientist, points to the year 2029 when a man-programmed device will be able to pass the Turing test [6] and defines 2045 as the year of the singularity: *"[...] I set the date for the Singularity - representing a profound and disruptive transformation in human capability - as 2045."* [7]

Paul Allen, Bill Gates' partner in the foundation of Microsoft, rejects this proximity of the singularity proposed by Kurzweil based on "The Law of Accelerating Returns" [8] where the latter states: *"[...] we won't experience 100 years of progress in the 21st century - it will be more like 20 thousand years of progress (at today's rate)."*

Not denying the possibility that such a singularity will ever occur, Allen does not accept Kurzweil's postulate for his foresight as it is based on past rates of technological growth to predict future technological growth rates. The co-founder of Microsoft argues that the assumptions set forth in "The Law of Accelerating Returns" are not like the laws of physics and, therefore, "*these 'laws' only work until they don't.[...]*" [9].

The rise of AI

Considering that AI was born in 1956, at least as a concept, and lived for nearly 60 years without known relevant applications, besides the massive production of science fiction contents of all kinds, why all this hype around AI now? The fact is that technological advances have brought AI out from the shadows of academia and research, into the business environment. The massive adoption of cloud computing and big data technologies has paved the way for AI to be implemented at industrial scale.

Whilst during the last decade an appropriate online presence became mandatory for organizations and their business perspectives, nowadays experts predict that an appropriate AI strategy is even more crucial and "*the digital disruption caused by these technologies will be significant*" [10]. Some even foresee this to be an even bigger technological shift than the observed during the "*Internet revolution*" or "*smartphone revolution*" [11].

This has been readily understood by the major tech corporations (Apple, Microsoft, IBM, Google, Amazon, Facebook, Baidu) which is reflected in their aggressive investment strategies into the AI sector. Many offer their results as powerful AI toolkits and libraries, often in the form of open-source projects to encourage involvement and further speed up their advancements with the help of spontaneous communities teaming around the projects. With powerful tools and libraries at our disposal, the main challenge at the corporate level is, in one hand, the recruitment of capable AI talent, as 80% – 90% are being already employed at the before mentioned major tech corporations and recruitment is happening straight out of academia [12] and, in the other hand, overcome the fear of society in adopting AI due to the implicit potential of job losses making them believe that another benefit of AI is the regained time to let engineers focus on creativity and innovation as "*AI automates some of the repetitive tasks of the engineer*" [11].

Contrary to the science fiction like achievements portrayed in media - autonomous driving, augmented reality (AR) and intelligent bots – the real disruptiveness of AI will be more subtle. It is the change that will affect all disciplines on a smaller scale – faster and better understanding and analysis of the available data in the most diverse contexts. Vast quantities of data are being generated every minute by all kinds of systems and it has long exceeded the amount that humans can comprehend or analyze effectively. With more and more internet of things (IoT) devices appearing in the market, the industry is following the trend, switching to digital, with interconnected cyber-physical systems targeting smart factories (Industry 4.0). There is a lot of added value hidden in this data and waiting to be extracted. This is where AI will establish itself, as not only are such data amounts unproblematic, for some ML techniques such as reinforcement learning, they are a necessity to reach satisfying end results.

The challenge lies in transforming the raw operational data into a format that ML system can consume and learn from. By utilizing the appropriate ML techniques new insights can be gained which then can be used by an AI system for diverse use cases: optimize company expenditure; let an AI agent control the load balancing on systems thus improving performance and saving electricity costs at the same time; faster and improved decision making; better and more precise predictive models (e.g. in the scope of maintenance, or weather). The possibilities are endless and the gain tremendous for companies as well as individuals. Finding the various use cases and preparing the necessary data and tools to address them is the current AI frontier.

Apart from the growing influence in data analysis, AI improves existing and opens up new communication channels. Advancements in object recognition (OR) have opened the doors to autonomous driving and AR, whereas advancements in NLP gave birth to "voice search" and enabled personal assistants. Actually, voice search skyrocketed since 2015 and is now available in most smartphones and a core component of the AI-based personal assistants (e.g. Siri, Google Now, Cortana, Alexa). The benefit is not only the simplified access to the search functionality but also the simplified usability, making it faster and more intuitive than typing. Chief scientist at Baidu, Andrew Ng, predicts that by the year 2020 we should see 50% of search requests to be made by voice or image [13].

Businesses have adopted this technology by automating customer support in the form of AI assistants or chatbots. Nevertheless, there will still be a need for human support due to factors like urgency or emotion or, as currently still is the case, higher level decision making. Further improvements in AI, with the increase of data to learn from, this experience will become more and more personalized and natural, ultimately considering the user satisfaction during interaction and responding accordingly (e.g. by reacting on cues in the person's style of speech or writing). Same as a big part of our society has switched from the traditional paper press, to reading the newest headline in their smart-device on the move, it is not too far off to get personalized news being told to us by our smart-device upon request, be it the built-in AI personal assistant talking or an extension of our favourite news app.

After the release of ARKit in iOS 11, Amazon has already taken a step forward by enabling an experimental version of an image search for products via the smartphone camera. Pre-trained OR models are being used to distinguish the visible objects and, after classification, suggestions from their products on Amazon in that category are being shown. Despite the economic spin that Amazon has put on it, it highlights the current capabilities in OR and one possible use case. Combine OR with NLP and you can create an accessibility app that helps blind people identify objects and even explain or describe them.

But these are just the first use cases we get to experience representing the tip of the iceberg. Due to the machine's capability to outperform humans in certain tasks, AI will continue to gain acceptance. By now there are first versions of AI systems that create new AI by themselves without the need for human guidance and coaching. In these self-play approaches, the machine has free reign about how it should interpret the available data. It competes in a task against an older version of itself in millions of encounters. Only a goal condition is being set. This leads to new, innovative approaches to certain problems and brings us a step closer to simplify the tedious data preparation process. The downside to this approach is the "intelligence" that the AI derives can become so complex that it is hard for a human to follow why certain decisions were made. The latest example of this is DeepMind's AlphaGo Zero. Compared to its predecessor AlphaGo, which in 2016 defeated the grandmaster Lee Sedol in the ancient board game "Go", AlphaGo Zero is completely self-trained and overcame its predecessor after training for only a fraction of the time AlphaGo needed. However, the game moves made by AlphaGo Zero are often described by game experts and enthusiasts as "alien" as they often don't follow the established tactics.

The available hardware, data and expertise in AI will improve and open up new yet undiscovered possibilities for its use, ultimately becoming part of our everyday lives the same way as smartphones have by now. There will be aspects of AI that we will not even notice and take for granted as they will be enhancing existing computation processes in the background. Those could be improved weather reports, accurate on the minute, or that your lawyer is actually a piece of code running somewhere in the cloud. Other aspects, like self-driving public transport and cars, will be more visible but become the norm and turn into an accepted tool like the Internet has over time.

One can agree that AI has the required capabilities and will revolutionize our systems across all existing fields. The foundation for implementation at an industrial scale is set and many use cases have already matured and shown the added value AI provides. Now it is up to the companies to decide on their AI strategy, which technologies best suits them, or otherwise risk being left behind.

AI technologies – an overview

Intending to resume the vast universe of technologies that are generally classified under the umbrella of AI technologies and techniques, Gartner published its "Priority Matrix for Artificial Intelligence, 2017" [14] – **Figure 1**. This map exhaustively sums up all the relevant AI themes known today and also prioritizes them according to their benefit (or impact) and estimates the time for the full adoption of each one of them

Within the scope of this article, we will focus on AI themes which will both have a particular interest for communication service provider (CSP) and DSP and an expected shorter-term impact. More concretely, in the next sections we will focus our attention briefly on the description of the following AI technologies:

- ML;
- Deep Learning;
- Speech Recognition;
- Intelligent Apps;
- NLP;
- NLG;
- Predictive Analytics;
- Virtual Customer Assistants;
- Computer Vision.

Just a quick note on VR and AR: even though Gartner considers them as AI topics [14], we consider that those are subject matters in a field

benefit	years to mainstream adoption: less than 2 years	2 to 5 years	5 to 10 years	more than 10 years
transformational		AI-Related C&SI Services Cognitive Expert Advisors Deep Learning Intelligent Apps Machine Learning	Algorithm Marketplaces Artificial Intelligence for IT Operations (AIOps) Platforms Cognitive Computing Conversational User Interfaces Deep Reinforcement Learning NLP Virtual Assistants	Artificial General Intelligence Autonomous Vehicles Human-in-the-Loop Crowdsourcing Level 4 Vehicle Autonomy
high	Ensemble Learning Speech Recognition	Commercial UAVs (Drones) Deep Neural Network ASICs GPU Accelerator Natural-Language Generation Predictive Analytics Virtual Customer Assistants	Augmented Reality Bots Computer Vision Digital Ethics Graph Analytics Level 3 Vehicle Autonomy Neuromorphic Hardware Prescriptive Analytics Smart Robots	
moderate		Consumer Smart Appliances FPGA Accelerator Learning BPO Virtual Reality	Knowledge Management Tools	
low				

As of July 2017

FIGURE 1 – Gartner's priority matrix for AI, 2017 [14]

of their own, which nevertheless will probably use many support AI techniques in the future. Therefore we chose not to address them here.

Machine learning

ML is a technical discipline that aims to extract knowledge and pattern information from a typically large series of observations or sample data points. The discipline of ML can be further sub-divided into three areas, namely:

- Supervised learning (input/output data pairs available and the task is to find the more accurate mapping);
- Unsupervised learning (only input data is available, without any classification, categorization or labelling, and the task is to infer a function able to describe hidden structures from the data);
- Reinforcement learning (where evaluations are given of how good or bad is a policy leading to a sequence of actions).

ML is one of the hottest concepts in AI technology at the moment, given its extensive range of effects on business, namely on all business functions that involve finding patterns of interest (relevant information) in huge amounts of data being generated daily by all kinds of information systems, applications and devices. Combined with big data technologies, ML is making all the difference in the way companies manage their businesses.

Deep learning

Deep learning expands standard ML by allowing intermediate representations to be discovered. These intermediate representations allow more complex problems to be tackled and others to be potentially solved with higher accuracy, fewer observations and less cumbersome manual fine-tuning. Seen by some authors as a sub-branch of ML, deep learning is getting a lot of attention lately because it is capable of addressing previously human-only capabilities, such as image recognition, text understanding and audio recognition.

Deep learning discovers intricate structure in large datasets by using the back-propagation algorithm to indicate how a software machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolution nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shined a light on sequential data such as text and speech [15].

Hundreds of vendors are exploring the applicability of deep learning to a range of fields, such as computer vision, conversational systems and bioinformatics. Researchers are steadily publishing surprising new papers on this topic. Heavyweights like Google, Apple, Microsoft, Facebook and Baidu are increasing their deep-learning R&D, for example, deep learning is behind Apple's Siri, Google's Google Now, Microsoft's Cortana and Amazon's Alexa. Hardware manufacturers are intensifying delivery of new, high-performance computing systems for training deep neural networks [14].

Speech recognition

Speech recognition technology (sometimes referred to as speech to text - STT) is a set of technologies capable of recognizing human speech and then translate phrases into text strings which can then be further processed by computer means.

Speech recognition performance has rapidly accelerated in the last year. At the research level, companies such as IBM, Microsoft, Google, Amazon and Baidu all demonstrated rapid scientific improvements in 2016-2017. Application of deep learning techniques like convolution neural networks, recurrent neural networks and end-to-end neural architectures using connectionist temporal classification loss (championed by Baidu) has dramatically advanced the state of the art, accuracy and reduced the time to train models. Microsoft and Google have both achieved parity with human error rates with 4.8% errors while humans are at 4.9% [14]. In face of the recent accuracy results achieved and the widespread use of the technology on smartphones, personal assistants and driver-car user interfaces, speech recognition has effectively reached the status of mainstream technology.

Intelligent apps

Intelligent apps (or smart mobile applications) are hyper-customizable and self-learning, based on the application of big data and ML techniques. Intelligent apps allow the app provider to collect more personal and commercial information about consumers and progressively adapt the app to the user.

Intelligent app strategies are becoming extremely popular with marketers and IT leaders in charge of customer experience. These IT leaders are moving towards the first three stages of the intelligent app strategy through the increased use of data ingested from mobile apps, and wearable devices and behavioural mobile app analytics that collect and sync information about users, their whereabouts and their social graph. Increasingly, providers are also actively looking at taking the personalized customer experience toward a predictive experience and thus are moving towards the fourth stage of the intelligent app strategy. In addition, in the next two to five years, IoT and big data will meet analytics, and more data will make systems smarter and ML will be a big factor in how this evolves [14].

Natural language processing

The term NLP is usually applied to the use of several natural language tools (more than 20 different tool classes), such as knowledge graphs, STT, machine translation, automatic summarization, entity recognition, question answering, NLG, sentiment analysis and text analytics. NLP helps to make the interaction between humans and computers easier and more natural to humans.

Natural language generation

NLG is a technology capable of generating natural language descriptions (e.g. complete English text descriptions). The technology combines NLP with ML in order to generate contextualized insights from large volumes of data.

This technology is becoming common with large information analysis platforms, where information like trends, data relationships and correlations are dynamically identified and explained in plain English, as the users interact with business intelligence and analytics applications. NLG is used to synthesize textual content by combining analytic output with contextualized narratives, helping less expert users to derive knowledge from complex data, without the help of experts.

Predictive analytics

Predictive analytics is a form of advanced analytics that examines data or content to answer the question “what will happen?” or more precisely “what is likely to happen?”. It is characterized by techniques such as regression analysis, multivariate statistics, pattern matching, predictive modelling and forecasting.

The market is demonstrating greater practical readiness as more organizations move from talking about predictive analytics to adopting and executing, pushing the technology forward. Interest is also driven by the improved availability of data, lower-cost computer processing (especially in the cloud) and proven real-world use cases. Predictive models are no longer just produced by data science platforms; predictive analytics is being embedded in more business applications than ever before.

Virtual customer assistants

Virtual assistants (VA) help users or enterprises with a set of tasks previously made only by humans. VA use AI and ML (e.g., NLP, prediction models, recommendations and personalization) to assist users or automate tasks. VA listen and observe behaviours, build and maintain data models, and predict and recommend actions. They may act for the user, forming a relationship with the user over time. VA shift responsibility for understanding the business process from the user to the system by corresponding with the user.

The VA space is currently dominated by conversational interfaces such as Apple Siri, Google Assistant, Microsoft Cortana, IPsoft Amelia, Nuance Nina, Amazon Alexa, Kore.ai and SAP CoPilot. Increasingly, image recognition, behaviour and event triggers will enhance VA that may be deployed in several use cases, including virtual personal assistants, virtual customer assistants and virtual employee assistants. VA can act on behalf of consumers, employees and businesses, but the use cases are all based on the same AI technologies.

VA adoption grows as users get more comfortable with them, technologies improve and the variety of implementations multiply, namely [14]:

- Unobtrusive, VA-like features, such as Gmail’s Smart Inbox with recommended replies and

Microsoft's Delve that finds unknown resources embedded in existing products;

- Narrow-purpose VA have also emerged (such as personal financial advisors, health and wellness coaches, and calendaring agents);
- VA are increasingly used to answer customer questions about products and services.

Computer vision

Computer vision is a process that involves capturing, processing and analyzing real-world images and video to allow machines to extract meaningful, contextual information from the physical world. Today there are numerous different and important computer vision technology areas, including machine vision, optical character recognition, image recognition, pattern recognition, facial recognition, edge detection and motion detection.

The convergence of enabling technologies — such as deep learning, neural networks, large swaths of data and massive parallel processors — has brought new life into the significantly advancing field of computer vision. For example, 30 years ago, object classification was a relatively difficult task. However, today, current computer vision implementations are able to classify millions of individual objects with more than 95% accuracy. These advancements allow computer vision algorithms to make quicker and more accurate visual identifications. In turn, this has led to the rise of "new players" in this field (outside of academia) such as Baidu, Microsoft and Google. Although computer vision is still in its adolescence, with a limited adoption, it is beginning to be recognized for its broad applicability in AI, being expected to ramp up quickly [14].

Additional thoughts on AI tools

As mentioned above, access to AI technology has been reserved to the real experts until recently when a new range of AI expertise companies started to provide the platforms and tools for non-experts to use alone and even some problem-domain assistance in the application of those tools to specific business problems. A popular topic like "strong AI" or general-purpose machine intelligence, now more often called "Artificial general intelligence" or AGI, which means a form of AI capable of addressing a large variety of problems or use cases in a human-like fashion, without any assistance, does not exist yet and experts say it will take decades before it actually exists in a useful form, that is beyond science fiction movies and trade show demonstration. What we presented in this section fits the kind of AI generally called "weak AI" or "narrowband AI", that is, a set of tools and techniques which no matter being extremely powerful in their areas of application can only deal with specific areas of automation, not general problems of any possible kind.

In the next section, we will present use cases of the above described AI technologies within the context of Altice Labs' projects.

AI applications at Altice Labs

AI applied to network management

For CSP, the transformation that is currently taking place in the network domain, mostly driven by the virtualization of the network functions (NFV), as well as by the network "softwarisation" (SDN), is revolutionizing the management domain which will have to evolve in order to accommodate a more flexible and dynamic network environment. This is challenging for the traditional operational management systems which have been, for decades, rigid and static. On the other hand, it unveils new opportunities for service providers which will be able to be more efficient from an operational and business perspective. On the operational side, due to the ability to dynamically deploy virtualized sensors and/or actuators in specific and potentially problematic locations, service providers will have the possibility to detect (from a sensing/assurance perspective) and solve (from an actuation/fulfilment perspective) potential network and service problems much quicker than before. Furthermore, with these new sensing and actuation capabilities, applying AI techniques within the operational management systems is finally a realistic scenario and therefore solving or mitigating network and service problems in advance and in a formless prone to human mistakes is now possible. In the end, this will increase the user quality of experience (QoE) and satisfaction and indirectly impact the business revenues. For this reason, self organizing networks (SON) features for 5G networks, based on AI mechanisms will soon become a reality and shift the network management paradigm from a human-based, machine-assisted methodology towards a machine-based, human-assisted paradigm.

CSP are already moving towards this new management approach in which they will be able to provide autonomic capabilities to respond to detected or predicted network problems in proactive or reactive manners. Two main methods are utilized to provide autonomic functionalities:

- The first approach, which is probably the one that will be implemented in the first place by service providers, relies on direct reaction to identified network problems by means of the definition and application of autonomic policies that specify the tactics to guide the automatic reactions. This autonomic management approach, also known as **policy-driven autonomic management**, suits the regular network problems or events that have been recognized by the system or network administrators;

- For emerging network problems that have never been identified due to the novelty and dynamicity of 5G networks, the intelligent algorithms in the field of ML and deep learning, such as classification and clustering, can be applied. This is the second approach, also known as **AI-driven autonomic management**, which is complementary to the first one and will probably be implemented in real life scenarios in a 2nd stage.

Figure 2 depicts a very high-level perspective of an autonomic management system, which provides self-organizing management for software-defined and virtualized 5G networks. Apart from the underlying SDN and NFV-based network infrastructure, a closed-loop workflow referred to as an autonomic management loop, starting from the monitor and terminating at the orchestrator, is highlighted. When a network problem is detected or predicted, an autonomic management loop is initiated. The network problem is analyzed and diagnosed for deciding on a countermeasure over the underlying infrastructure. To verify the effectiveness of an action, the monitoring modules derive the achieved performance and resources usage and notify the autonomic system.

On the left part of **Figure 2** is represented the **monitoring & pre-processing framework**, which is responsible for sensing the managed domain through heterogeneous sensors (virtualized, programmable, physical), persisting the large volumes of collected data and applying pre-processing/aggregation techniques to reduce the amount of data, and in parallel, produce network indicators to be analyzed upwards.

On the top part of **Figure 2** is represented the **autonomic management framework**, internally composed by the **analyzer**, **diagnoser** and **decision-maker** components:

- The **analyzer** is in charge of deriving a symptom for each detected or predicted network concern based on key perform indicators (KPI) metrics and sensor data;

- The **diagnoser** takes advantage of AI, data mining and stochastic algorithms to provide network intelligence in order to diagnose the root cause of network problems. The diagnoser will process the reported symptom

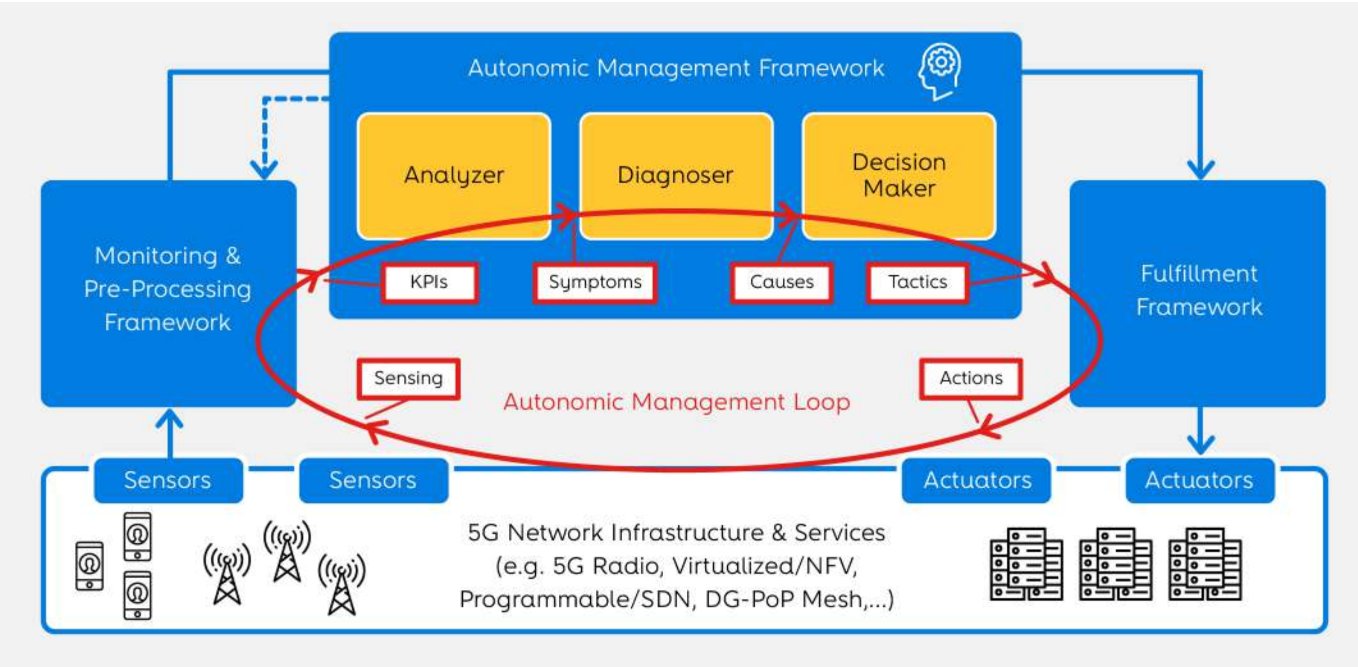


FIGURE 2 – AI applied to network management

and notify the decision-maker of the root cause. If the information contained in a symptom is not enough, the diagnoser needs to directly access the monitoring database to get the necessary information, such as specific sensor data and aggregated metrics. For identified symptoms, a policy-based diagnosis can be utilized to provide a direct and fast processing, which is suitable for real-time applications. For unidentified symptoms, the diagnosis will be achieved by means of intelligent algorithms;

- The **decision-maker** is in charge of deciding a set of corrective and preventive tactics to deal with detected and emerging network problems, respectively, in reactive or proactive manners, based on the incoming diagnostic information. This module makes use of two different approaches for providing a high-level response. One approach is a direct application of tactics specified through policies. Another is the integration of AI algorithms guided by such tactical strategies in order to determine tactic actions to be taken on the network, even for the cases where there are not completely defined tactical strategies about how to react;

- Finally, represented on the right side of **Figure 2** is the **fulfilment framework** which is responsible to orchestrate the tactics received from the autonomic framework in the managed domain in order to optimize the service delivery and improve QoE. This closes the autonomic management loop, also known as management control continuum.

Altice Labs is currently running three projects in the area of AI applied to network management: two H2020 projects named Selfnet [16] and SliceNet [17] as well an academic partnership research project.

Application of AI to marketing campaign engines

In the recent years, we saw an increase in one-to-one real-time marketing campaigns throughout several business areas, including CSP and DSP. This proved to bring good business results when properly implemented and when using advanced real-time tools such as the active campaign manager (ACM): revenue increased; customer satisfaction has improved, and churn has diminished. However, a closer look at the performance of this type of campaigns shows that there is a huge opportunity to increase its effectiveness by:

- Better defining the campaigns target audience by previously identifying the customers that will not accept it;
- Knowing what is the best campaign for each customer; which campaign will have the higher probability of acceptance by each customer.

Marketing team's biggest challenges includes targeting as many customers as possible with the most suitable campaign without eroding customer's relationship by proposing a campaign he doesn't want. In the next paragraphs, we shed some light on how AI can be used to tackle this challenge, the techniques used and the results obtained and, for the sake of simplicity, in this context, the term campaign includes product offers that are communicated to customers.

Affinity matrix

Under the umbrella of the innovation roadmap of the Altice Labs' product ACM, the use of AI techniques is based on the knowledge about past events, as a way to predict how a customer will react to a campaign. From a wide set of AI and ML techniques and algorithms available today, the use of the k-nearest neighbours (kNN) technique fitted our purpose mainly due to its simplicity, robustness with regard to the search space, easy to understand by humans, adequate performance and few parameters to tune. Also, kNN allows a probabilistic interpretation of results. Other techniques, such as deep learning, were evaluated but their inability to explain the obtained results/predictions and the computational effort necessary to retrain the models ruled them out.

kNNii, in a few words, is a non-parametric, lazy learning method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space – in our case the examples are the historical data consisting of customer properties and acceptance / not acceptance of campaigns. The output depends on whether kNN is used for classification or regression:

- In kNN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour;
- In kNN regression, the output is the property value for the object. This value is the average or weighted average of the values of its k nearest neighbours.

When the kNN algorithm is used with historical data from one campaign, an acceptance score for new customers can be obtained. The same principle applied to each campaign results in an acceptance score for each campaign of each customer. This can be represented as the affinity matrix - **Table 1**. For example:

	3-Play 100Gb offer	Sports TV bundle	Prime phone
Jamie	80	60	20
John	30	70	20
Sara	20	20	20
Ana	30	80	10

TABLE 1 – Affinity matrix example

Note: The Affinity values in the matrix do not formally mean probability of acceptance but are normalized between 0 and 100 and higher values mean a higher probability of acceptance.

According to the above example (**Table 1**) from the affinity matrix it is possible to extract highly valuable information such as:

- Who are the top 50% customers with higher probability to accept sports TV channels bundle? Ana and John;
- Which customers don't have campaigns with a score greater than 30? Sara. Are new campaigns needed for customers like Sara?
- Which campaigns don´t have customers with a score greater than 30? Prime phone. Should this campaign be replaced?

However, because the values in the matrix do not represent the probability of acceptance, we know that Ana is the customer with higher probability to accept sports TV channels bundle, but we don't know the exact probability. And this is an important limitation since the probability can be 90% or 10% for instance. An example of how to overcome this limitation is explored bellow - target audience optimization.

Target audience optimization

Generally, in a campaign life cycle, where a product is proposed to a customer, effectiveness (of the campaign) is defined as the ratio between the number of customers that accepted the campaign and the number of customers the campaign was proposed to:

Campaign effectiveness calculation:

- Customers that were targeted and have accepted: targeted positive (TP);
- Customers that were targeted and have not accepted: targeted negative (TN);
- Targeted customers (TC) = TP + TN = **target audience**;
- Effectiveness = TP / TC.

Using the example provided in **Table 1**, if we proposed sports TV channels bundle to 3 customers (Jamie, John, Ana) and only one accepted (Ana), then we have and effectiveness of 1/3=0,3(3) or 33,(3)%. It is clear that we have wasted the other 66,(6)% proposals.

As we broaden the number of targeted customers for a campaign starting at the most probable to accept according to the affinity matrix, the targeted positive will increase, the effectiveness tends to decrease because of the number of

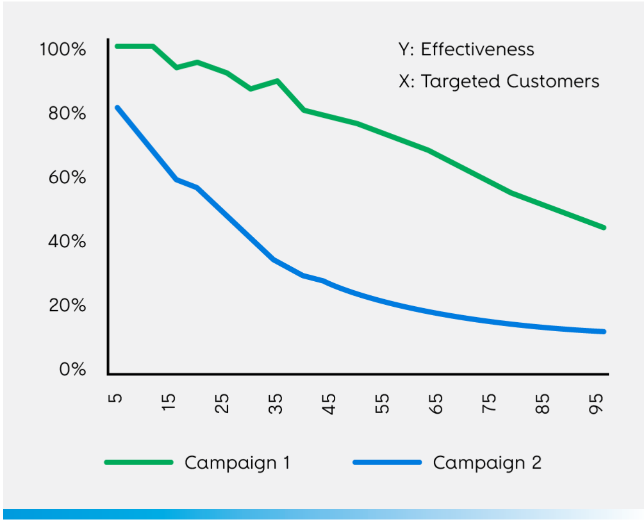


FIGURE 3 – Examples of targeted customers (X-axis) vs effectiveness (Y-axis) using the affinity matrix to select the customers starting with customers with higher affinity

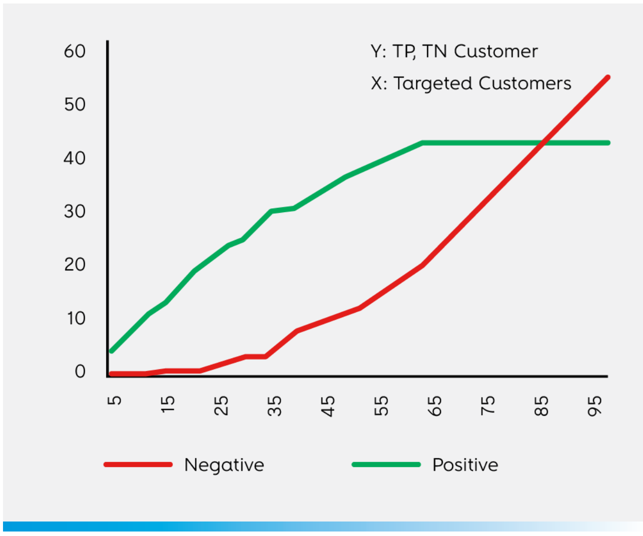


FIGURE 4– Example of targeted positive and targeted negative as targeted customers (X-axis) increases using the affinity matrix to select the customers starting with customers with higher affinity

targeted negative increases faster than targeted positive. If the targeted customers set is small a high effectiveness is achieved and a low number of targeted negative will be received, but then some customers that would have accepted the campaign are left outside the targeted customers. **Figures 3 and 4** depict generic examples of the relation between targeted customers and effectiveness. The depicted relation is highly dependent on the campaign details.

The optimum number of targeted customers depends on the campaign goals. For a set of campaigns, it may be interesting to have a very high effectiveness, minimizing the number of targeted negatives. On another set of campaigns it may be interesting to have a very high number of targeted positives regardless of effectiveness and so on. With the use of the affinity matrix for a given campaign and starting at a predefined target customers set with historical results for targeted positive, targeted negative and effectiveness it was possible to predict the behaviour of the campaign and plot the expected targeted customers VS effectiveness behaviour; the same was achieved for the campaign behaviour regarding target positive and target negative as a function of targeted customers set.

Going back to the sports TV channels bundle example it is possible to predict that if the campaign targeted customers with an affinity greater or equal than 70 (Ana, John) the effectiveness will increase from 33,(3)% to 50% without losing any targeted positive (Ana). In this example, Jamie would be available for another campaign. Testing in real-world scenarios, a high increase of effectiveness was possible (>100%, more than doubled), reducing the number of targeted customers (<50%, less than half) while losing a limited number of targeted positives (~10%). Knowing the expected behaviour of the campaign in terms of effectiveness, while increasing or decreasing the targeted customers set according to each customer campaign affinity, is a powerful tool to better achieve the campaign goals and better define the optimum number of targeted customers.

Final remarks

When applying kNN to create the Affinity Matrix and to target audience optimization, it was clear that different campaigns required a different configuration of kNN to achieve good results: different customer properties were chosen (dimension reduction) depending on the campaign. Future work will include the automatic dimension reduction thus allowing the complete process to be automatic and thus easily encapsulated in a software product (ACM).

The affinity values are not fully comparable in two different campaigns, which prevent us from directly answering the question: which is the best campaign for a given customer? Future work will overcome this limitation, where the focus will include customer optimization besides campaign optimization.

Using the adequate customer properties was important but equally important was the quality of such properties. Much better results were obtained when customer properties were captured in the moment when the customer is targeted by a campaign. kNN implementation details:

- Datasets ranged from 6000 to 325 000 records of campaigns acceptance / not acceptance;
- Initial customer properties ranged from 90 to 200 before dimension reduction and around 40 after dimension reduction;
- Distance Function used was Euclidean and K ranged from 3 to square root of training dataset size;
- Non-numeric customer profiles were normalized using a random function;
- The basic rules of using different datasets to train and evaluate the results were followed and the R programming language was used for implementation.

From the obtained results it is clear that valuable improvements are achieved using AI/ML to optimize current campaigns and that bottlenecks regarding low effectiveness can be tackled (if the campaign goals demand so). One of the main challenges is to be able to encapsulate the process of using AI in a way that it becomes straightforward to use by non-AI specialists and without the complexity of the AI technology implementation details.

Automatic digital content recommendation

Digital product consumption (digital services and content) is growing immensely in the last few years. It is growing to such extent that consumers are being overwhelmed by the number of available products and associated options being lost in the vast amount of information from where they must make their choices (known as the information overload problem) [18].

One of the major objectives of a DSP is to enhance its customer usage experience but, with such an amount of available content, it becomes difficult for a client to choose. Without an easy and enjoyable way to make a choice, this multitude of content becomes a disadvantage rather than an advantage for the DSP. Easing the content selection process and focusing on the personal interest of each user is a benefit (rapidly becoming mandatory) for any service provided by a DSP. But to successfully help the user and recommend targeted content, it is indispensable to know the user interests and preferences. Besides its use to promote customer satisfaction and thus service usage - increase client fidelity - in-depth client knowledge may also prove crucial in reducing DSP acquisition and operational costs, by allowing it to focus on negotiating the most relevant content.

In the specific usage scenario of IPTV services, and in particular when being provided through a STB, the DSP is constrained by [19]:

- Uncertainty: which person (or persons) in the house is (are) watching TV;
- Technology: the list of proposed items has to be small because of the limited screen definition and the reduced navigation capabilities; the generation of the suggested items must respect very strict time constraints because TV's customers are used to a very responsive system;
- Scaling: the system needs to scale up in a successful manner with both the number of customers and items in the catalogue;
- Dynamism: the live broadcast channels parts of the catalogue are subject to last minute changes.

Nowadays, as IPTV services may also be provided through different devices, the content choice is also impacted by the specific consuming context – e.g. the same person may experience different consuming behaviour depending on the device and environment/context constraints (outdoor, in a waiting room, with a companion).

ML technologies may help the DSP in the process of providing the consumers with accurate, timely and context adapted contents. For instance, the identification of client preferences similarities can be supported on unsupervised learning techniques, by processing datasets with multiple features describing clients' behaviour and drawing inferences between them and therefore creating clusters of clients with similar preferences or identify hidden patterns between the features describing the clients.

More specifically, content recommender systems may help the DSP in providing personalized recommendations by trying to predict what the most suitable products or services are, based on the past user's preferences and constraints. It might use information of users preferences, either explicitly expressed - e.g., as ratings for products – or inferred from user actions interpretation [20] – e.g. in the IPTV scenario, completely viewing a specific content may be considered an implicit sign of preference for a TV program.

The most common recommender systems implementations are [20]:

- Content-based: the system learns to recommend items that are similar to the ones that the user liked in the past; the similarity of items is calculated based on the features associated with the compared items;
- Collaborative filtering: recommends to the active user the items that other users with similar tastes liked in the past; the similarity in taste of two users is calculated based on the similarity in the rating history of the users;
- Hybrid approach: by mixing both techniques in order to circumvent each intrinsic limitations and constraints of all implementations.

Deep learning techniques may also prove useful given their adequacy to process huge datasets (vertically and horizontally) and automatically incorporate feedback on the model evolution (e.g. recurrent model networks). Additionally, neural network implementations might be quite helpful in identifying feature correlation that may enhance feature selection and design, therefore allowing its usage with ML algorithms not suited for high feature dimensionality.

One additional field of work, which is quite relevant given the business interests of the Altice Group (telecom, media & content, data & advertising), may be the effort of extending the profiling and recommendation processes to allow a more thorough characterization of the users/customers, by taking full advantage of their usage patterns within the universe of services provided by the above-mentioned group. Cross-service profiling might allow tapping non-evident cross-selling opportunities and the subsequent impact on average revenue per user/unit (ARPU) and retention.

Altice Labs is currently running two projects in the area of digital content recommender systems: an H2020 project named STREAMLINE [21] and an internal project aiming to provide recommendations to Altice PT service users.

Key takeaways

AI has definitely come to stay and should be part of most of the CSP/DSP solutions developed in the near future. Actually, these players are well positioned in this field since one of their key assets - abundance of data (big data) - is the primary fuel for AI technologies applications. However, due to the extent of these applications, many challenges and opportunities are foreseen, namely in:

- Customer services and support: chatbots and AI agent-assisted customer support;
- Home services and entertainment: home automation, entertainment automation (voice control, intelligent content recommenders);
- Network optimization: self organizing networks and agent-assisted network management, capacity optimization, service assurance;
- Marketing engagement: AI assisted product bundles, recommendations, personalized customer profiling, upsell and cross-sell.

As shown in this article, Altice Labs is already addressing some key areas of AI application and expects to be investing more in this area in the near future, recognizing the “game-changing” nature of AI for DSPs. ■

References

[1] J. McCarthy, M. L. Minsky and N. Rochest, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” *AI Magazine*, vol. 4, no. 27, pp. 12-14, 2006.

[2] R. S. Morison, “Proposal for the Dartmouth summer research project on artificial intelligence,” 100 Years: The Rockefeller Foundation, 1955.

[3] Dictionary.com, LLC., “Dictionary.com,” 2017. [Online]. Available: http://www.dictionary.com/browse/artificial-intelligence?s=t.

[4] E. Rich and K. Knight, Artificial Intelligence, 2nd ed., New York: McGraw-Hill, 1991.

[5] Stottler Henke - Smarter Software Solutions, “Artificial Intelligence Glossary,” Stottler Henke Associates, Inc., 2017. [Online]. Available: https://www.stottlerhenke.com/artificial-intelligence/glossary/.

[6] A. M. Turing, “Computing Machinery and Intelligence,” vol. LIX, no. 236, pp. 433-460, 1950.

[7] R. Kurzweil, The Singularity is Near: When Humans Transcend Biology, Viking Penguin Group, 2005.

[8] R. Kurzweil, “The Law of Accelerating Returns,” Kurzweil Network - accelerating intelligence, 7 March 2001. [Online]. Available: http://www.kurzweilai.net/the-law-of-accelerating-returns.

[9] P. G. Allen, “The Singularity Isn’t Near,” MIT Technology Review, 12 October 2011. [Online]. Available: https://www.technologyreview.com/s/425733/paul-allen-the-singularity-isnt-near/.

[10] D. Schubmehl, “Worldwide Cognitive Systems and Artificial Intelligence Revenues Forecast to Surge Past $47 Billion in 2020, According to New IDC Spending Guide,” IDC - Analyze the Future, 28 October 2016. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS41878616.

[11] C. Stancombe, A.-L. Thieullent, S. KVJ, A. Chandna, R. Tolido, J. Buvat and A. Khadikar, “Turning AI into concrete value: the successful implementers’ toolkit,” Capgemini, 2017.

[12] R. C. Waite, “This Is Why All Companies Need An AI Strategy Today,” CB Insights, 14 September 2017. [Online]. Available: https://www.cbinsights.com/research/artificial-intelligence-strategy/.

[13] KPCB - Kleiner Perkins Caufield Byers, “Internet Trends 2016 - Code Conference,” KPCB - Kleiner Perkins Caufield Byers, 2016.

[14] K. F. Brant and T. Austin, “Hype Cycle for Artificial Intelligence,” Gartner, 2017.

[15] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” *Nature*, no. 521, pp. 435-444, 2015.

[16] 5G PPP, “Selfnet,” 2014. [Online]. Available: https://5g-ppp.eu/selfnet/.

[17] 5G PPP, “SliceNet,” 2017. [Online]. Available: https://slicenet.eu/.

[18] S. Berkovsky, T. Kuflik and F. Ricci, “Distributed Collaborative Filtering with Domain Specialization,” in *RecSys ‘07 Proceedings of the 2007 ACM Conference on Recommender Systems*, Minneapolis, Minnesota, 2007.

[19] R. Bambini, P. Cremonesi and R. Turrin, “A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment,” in *Recommender Systems Handbook*, Springer, 2010, pp. 299-331.

[20] F. Ricci, L. Rokach and B. Shapira, “Introduction to ecommender Systems Handbook,” in Recommender Systems Handbook, Springer, 2010, pp. 1-35.

[21] H2020 Streamline Project, “STREAMLINE,” European Union’s Horizon 2020 , [Online]. Available: http://h2020-streamline-project.eu/.

10

Tackling security and privacy in a digital world

Luis Cortesão, Altice Labs
luis-m-cortesao@alticelabs.com

Paula Cravo, Altice Labs
pcravo@alticelabs.com

The need of providing and guaranteeing both technical and regulatory means of protecting critical information, either personal, industrial or national, can no longer be devalued. Under the vision of a hyper-connected world, security and privacy must be properly tackled by holistic approaches, since they are considered as key factors for a large-scale deployment of new IoT-enabled services and applications.

Keywords

Privacy; Security; GDPR; IoT; Smart Cities

Introduction

One of the 21st century undeniable facts is the existence of a digital world ruling all things, data, information and knowledge. But this new world came with many challenges: new ways of communication, huge amounts of information to store, and a new “digital reality” paradigm, translated in the dissemination of social networks, digital services and digital personas, that frequently co-exist with and complement traditional store-front services and identities, extending or even replacing their physical counterparts. Even objects become “smart”, as the Internet of Things (IoT) is intended to transform our daily lives through the convergence of significant efforts from different disciplines. But IoT and Smart Objects are not only revolutionizing the life and experience of consumers but also the way companies learn about their consumers through the collection of data from smart devices. The availability and completeness of such data allows for more accurate profiling of consumers, which, in turn, may lead to more efficient direct marketing. This poses the problem of IoT data privacy, data anonymization, and the consumer’s rights to decide on the use of his own data.

The need for providing and guaranteeing both technical and regulatory means of protecting critical information, either personal, industrial or national, can no longer be devalued. Under the vision of a hyper-connected world, security and privacy must be properly tackled by holistic approaches, since they are considered as key factors for a large-scale deployment of new IoT-enabled services and applications.

This article will briefly present the European Union (EU) General Data Protection Regulation (GDPR), one of the “legal tools” developed by EU in an effort to achieve the above-mentioned protection. We will also present a list of technical security and privacy recommendations, of general application to any system. Specific guidelines for the IoT world will also be covered, along with the brief presentation of the EU Framework Programme 7 (FP7) SMARTIE project, where some of these guidelines were developed and applied.

Outstanding attacks and data breaches of the near past

The past two years broke records after records on cyber security attacks and exposure of sensitive information. The following list presents a short timeline of some of the most relevant attacks on security and user’s privacy. While some of the attacks are difficult to prevent, most of them could have been avoided had the companies under attack invested more in its costumers’ security and privacy. Other attacks could have had lesser impact if a correct security process had been implemented and followed by all parts. Other cases simply highlight how some companies completely disregarded the importance of protecting their costumers’ private information.

- 2010 – Malicious worm Stuxnet successfully attacks Iran’s nuclear program, on what some consider an act of cyber war [1];
- 2011 – PlayStation suffered an external intrusion that compromised personal information of 77 million users and may have exposed their credit card numbers, security answers, physical addresses and more [2];
- 2013, 2014 – Yahoo unattended data breaches compromise more than 1.5 billion user accounts, including emails and credentials. The breaches were not reported until December 2016 [3];
- 2016 – Russian suspected of interference with USA state-elections systems [4];
- 2016, September – Sports Direct, a United Kingdom sports retailer, was hacked. Unencrypted information of 30 000 employees was accessed, but the company decided not to notify the staff [5];

- 2016, October – on October, 21, Mirai botnet malware, affecting Linux systems on common IoT devices, launches a series of huge Distributed Deny of Service (DDoS) attacks all over the world, affecting Netflix, Twitter, Spotify, Reddit, CNN, PayPal, Pinterest, FoxNews and many others. Two months later, a variant of Mirai attacked and crashed approximately 900 000 routers from Deutsche Telekom, as well as other Internet Service Provider (ISP) companies [6][7];
- 2017, March – WikiLeaks reveals that the US Central Intelligence Agency (CIA) is snooping common citizens using everyday gadgets [8];
- 2017, May – WannaCry ransomware exploits a vulnerability of Microsoft (MS) systems and affects many hundreds of companies on 150 countries. Old Microsoft systems or those not running the latest updates, including hospital medical equipments were locked [9];
- 2017, June – Petya/NotPetya ransomware, exploiting the same vulnerability as WannaCry, targets facilities in Ukraine and spreads across the world [10];
- 2017, June – Sensitive personal details of 200 million US citizens are accidentally exposed by a US Republican Party's contractor. The extremely detailed information was gathered from a large variety of sources, and the included subsequent data analysis wasn't protected by even a password [11].

The list of cyber attacks and incidents with private information exposure and compromise of personal and national security is endless. The previous examples cover security incidents and attacks that include industrial facilities damage, risks to the independence of nations, attempts to control unwary citizens by intelligence agencies, and the obvious attainment of illicit profit. The methods used also varied and ranged from worms infecting air-gapped facilities, ransomware, DDoS, smart and network devices snooping, data collecting and analytics, etc. In some cases, the attack was a simple uncontrolled access to internal systems.

According to Symantec's 2017 Internet Security Threat Report [12], *"in the last 8 years, more than 7.1 billion identities have been exposed in data breaches"* and there was an increase in ransomware attacks. IoT devices are extremely vulnerable, taking, on average, 2 minutes to be attacked. *"The targeted attack landscape shifted considerably during 2016, with several groups emerging from the shadows and engaging in more public, politically subversive activities"*, and this trend is likely to continue.

The McAfee Labs 2017 Threats Predictions Report [13] states that threats will range from ransomware, sophisticated hardware and firmware attacks, attacks on smart home IoT devices, the use of machine learning to enhance social engineering attacks, to hacktivism. Cyber espionage will become as common in private sectors as it is among nation states, and an increase in cooperation between industry and law enforcement will be observed.

It's a cliché to say that information is power. And power has always had very high value. Because data is the basis for information,being private data the more valuable, several nations became very aware of the need to protect data and are working on legislation to prevent the repetition of some of the aforementioned attacks. One of the most recent legislation in the EU is the new GDPR, very briefly presented next.

EU General Data Protection Regulation

On May 25th, 2018, the EU's regulation GDPR takes effect and becomes enforceable. This new regulation, technically known as EU 2016/679, replaces the 1995 Data Protection Directive [14].

GDPR introduces so many changes, with so many rules and consequences in a multiplicity of areas that, by itself, is driving growth in software-based privacy solutions. Some of the most important ones are highlighted next:

- GDPR is a regulation designed to enhance the protection of individuals residing in the EU as well as addressing the export of "personal data" outside the EU;

- It establishes a single law to enforce European data protection rules and regulation and the right to personal data protection. It legislates on common sense data security ideas, especially from the Privacy by Design school of thought;

- It applies to all organizations that do business in the EU and any organization outside the EU handling EU citizens' personal data, which means that any company that works with information relating to EU citizens will have to comply with this regulation, even though they don't have any servers or offices in EU;

- It broadens the definition of "personal data": any information relating to an identifiable natural person, meaning someone who can be identified directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

- The rules for obtaining valid consent to using personal information are demanding. All organizations collecting personal data must be able to prove clear and affirmative consent to collect and process the data, and consumers silence or inactivity is no longer valid consent. Specific conditions apply to the validity of consent given by children, with requirements to obtain and verify parental consent below certain age limits;

- GDPR introduces mandatory Data Protection Impact Assessments (DPIA) in the cases where a processing is *"likely to result in a high risk to the rights and freedoms of natural persons"* [15];

- Some organizations may have to appoint a Data Protection Officer (DPO), whose primary objective is ensuring compliance with the GDPR. His tasks include: advising their colleagues and monitoring their organization's GDPR/privacy law/policy compliance, including via training and awareness raising, running audits, advising regarding DPIA and cooperating with supervisory authorities;

- The right to erasure and to be forgotten are also included in this regulation, in line with other EU past decisions on the internet protection of individuals. These rights allow individuals to request, at any time, that their data be deleted, extending the application of the regulation to include data published on the web;

- GDPR requires that organizations must not hold data for any longer than absolutely necessary, must not change the use of data from the purpose for which it was collected originally, and, at the same time, they must delete any data at the request of the data subject, as long as this request complies with legal obligations. This meets the foundation principles of Privacy by Design, presented later;

- Breach notifications are also regulated. They must include information on data categories, records touched and the approximate number of data subjects affected, among other. This means some sort of detailed intelligence is needed on what hackers and insiders are doing;

- Liability is extended to all organizations that touch personal data, which means that organizations that are merely service providers must also comply with the regulation;

- The one-stop shop privacy rule in GDPR is intended to ensure that organizations and individuals can deal with cross-border privacy related issues from their home base and that such issues can be addressed consistently across the EU. Organizations that operate in multiple EU countries will deal with only one supervisory authority rather than a different one for each EU state, and individuals will only have to approach one single data protection authority to lodge complaints.

Summing up, this regulation is about consumer's rights. It aims to protect consumer data that will impact not only businesses in the EU but the rest of the world as well. **Figure 1** summarizes consumer's rights.

On the other hand, GDPR is also about obligations for organizations. The regulation's rules are based on Privacy by Design principles. **Figure 2** summarizes organization's obligations.

And what will happen when a data breach occurs? **Figure 3** summarizes some of the steps an organization must follow to address data breaches when complying with GDPR.

Consumers have the right to...

give explicit consent to collect his personal data	correct inaccurate data	transfer personal data from one system into another	withdraw consent and ask for personal data to be erased	know what is being collected and how is being processed

FIGURE 1 – Consumer's rights under GDPR

Organizations must...

limit data collection	process data only for the purpose for which data was collected	conduct prior assessments when processing sensitive data	limit access to the consumer's data only to authorized personnnel	record processing activities	implement continuous assessments to check for compliance

FIGURE 2 – Organization's obligations under GDPR

Step 1	Step 2	Step 3
Organizations must notify Data Protection Authorities (DPA) within 72 hours after discovering the breach detection	Organizations must provide detailed information, including: • number of records exposed • type of data exposed • the consequences of the breach • measures taken to mitigate exposure	Organizations must notify the data subject if the breach is of high risk to consumer's right

FIGURE 3 – Data breach processing

What if the organization does not comply with GDPR? Fines are potentially hefty: up to €20 million or up to 4% of the organization's global annual revenue for that year (whichever is larger). GDPR's heavy fines make a stand that EU citizens' privacy is to be taken seriously, and are meant to play a major role on organization's compliance.

GDPR accomplishes a better personal data security because it not only requires privacy by design but also embraces security by design. This means that security must be considered present at every phase of the development process and for the entire data lifecycle, in an embedded way. This way, all aspects of the digital reality will be included: the network, the cloud, all IT infrastructures, the data storages, all databases, the applications, etc.

But complying with the regulation also means that organizations will probably have to redesign their cyber security infrastructure, to review their security policies, to conduct regular testing, including social engineering, and to study mechanisms and solutions to prevent identity fraud.

All this work will be incomplete without constant systems monitoring and auditing, accompanied with the proper notification and alerting mechanisms [16].

The digital new world

Defining and trying to enforce security on private data based on laws is only a part of the work. In fact, forecasts predict the digital world will have over 20 billion devices by 2020, with up to 63% of them dealing with personal information [17]. The widespread use of mobile devices and the mainstream adoption of IoT technologies have opened up whole new platforms and users for attackers to target, and in 2016 a number of emerging threats against these three increasingly high-profile areas could be observed. Infected IoT devices could also be used as a stepping-stone to attack other devices in a private network, and this could also mean that a device could participate in a global botnet that plays a role in taking down websites or services.

Hence, in the digital new world, laws can fulfil their goals only when properly assisted by technologic developments in security and people with the right skills.

Therefore, it is essential to correctly implement all the core pillars of information security: **confidentiality**, to ensure only allowed users access data; **integrity**, to ensure data is not tampered or altered by unauthorized users; and **availability**, to ensure that systems and data are available to authorized users when they need them. These security pillars are supported by the Security by Design Principles, as recommended by the Open Web Application Security Project (OWASP), and listed next.

Privacy by Design Principles, advocated by GDPR and also listed in next section, complement Security by Design Principles whenever dealing with private or otherwise sensible information. These principles may be considered the technical basis of any security and privacy approach in the digital world.

IoT devices, due to their intrinsic characteristics that enable them to exchange and analyse data in an automated manner, have more specific security and privacy requirements that will be addressed in the final sections of this chapter.

Security by design principles

The following principles are meant as guidance in order to produce secure applications by design [18]:

- **Minimize the attack surface area:** every feature added to a system also adds a certain amount of risk to it because every point of interaction is a potential part of a surface attack. So, closing unnecessarily open network ports, limiting the number of resources available to distrusted users, reducing the amount of running code, controlling physical access to critical devices and ensuring the secure destruction of outdated devices are examples of means of reducing the attack surface area;

- **Establish secure defaults:** by default, the user experience should be secure, and it should be up to him to reduce his security options, if allowed;
- **Follow the principle of least privilege:** all accounts must have the least amount of privilege required to perform their tasks. This encompasses user rights, resource permissions such as processing limits, memory, network and file system permissions;
- **Implement defence in depth:** multiple layers of security must be placed in the system, for if one security control is reasonable, more controls that approach risks in different fashions are better and can make eventual vulnerabilities much more difficult to exploit. For example, requiring for user authorization on all accesses, logging all accesses and using external auditing systems makes a system unlikely to be vulnerable to anonymous attacks;
- **Fail securely:** handling errors securely is a key aspect of secure coding. In general, a failure should follow the same execution path as disallowing the operation;
- **Don't trust services:** trust on external systems should not be implicitly assumed. All external systems should be treated taking this precaution;
- **Implement separation of duties:** as a rule of thumb, the entity that approves an action, the entity that carries out an action, and the entity that monitors that action must be separate. The goal is to eliminate the possibility of a single user from carrying out and concealing a prohibited action;
- **Avoid security by obscurity:** it is a weak security control and nearly always fails when it is the only control;
- **Keep security simple:** avoid complex approaches to security. Simplicity is often more straightforward and helps in controlling the attack surface area;
- **Fix security issues correctly:** test security issues to understand their root cause and develop fixes that don't introduce regression problems.

Privacy by design principles

Privacy by design or, more broadly, data protection by design, is a framework based on proactively embedding privacy into the design and operation of internet systems, network infrastructures and business practices. This means building privacy into the design, operation and management of a given system, business process, or design specification. The framework is based on Seven Foundation Principles [19]:

1. **Proactive not reactive – preventive not remedial:** Anticipate, identify, and prevent invasive events before they happen; this means taking action before the fact, not afterwards;
2. **Lead with privacy as the default setting:** Ensure personal data is automatically protected in all systems or business practices, with no added action required by any individual. For companies and organizations, this may mean using explicit opt-ins, implement safeguards to protect consumer data, assure restricted sharing, minimize data collection, and correctly implement retention

policies. This will reduce security risks for the fewer data collected and stored, and the less damaging a breach will be;

3. **Embed privacy into design:** privacy measures should not be add-ons, but fully integrated components of the system. So, developers should think about privacy as a core feature of the product;
4. **Retain full functionality (positive-sum, not zero-sum):** privacy by design employs a win-win approach to all legitimate system design goals; that is, both privacy and security are important, and no unnecessary trade-offs need to be made to achieve both;
5. **Ensure end-to-end security:** data lifecycle security means all data should be securely retained as needed and destroyed when no longer needed. Privacy protections follow the data, wherever it goes. This applies when the data is first created, shared with others, and then finally archived. Appropriate encryption and authentication should protect the data till the very end when it finally gets deleted;
6. **Maintain visibility and transparency—keep it open:** assure stakeholders that business practices and technologies are operating according to objectives and subject to independent verification. This is the principle that helps build trust with consumers. Information about privacy practices should be out in the open and written in non-legalese;
7. **Respect user privacy—keep it user-centric:** keep things user-centric; individual privacy interests must be supported by strong privacy defaults, appropriate notice, and user-friendly options. Consumers own the data. The data held by the organization must be accurate, and the consumer must be given the power to make corrections. The consumer is also the only one who can grant and revoke consent on the use of the data.

IoT, smart objects & smart cities privacy and security threats

IoT offers immense potential for empowering citizens, making government transparent, and broadening information access. In the context of smart cities, IoT devices are used to run a city or a building. However, privacy threats are enormous, as is the potential for social control and political manipulation [20]. Most of the technical security issues are similar to those of conventional servers, workstations and smartphones, but the firewall, security updates and anti-malware systems used for those are generally unsuitable for the much smaller, less capable, IoT devices.

The SMARTIE (EU FP7 project) project focused on the application of security and privacy mechanisms to different use cases under the umbrella of a high-level architecture with a strong emphasis on security and privacy. The instantiation of this architecture aimed to facilitate the creation of valuable and innovative services for citizens in the upcoming generation of smart cities, while at the same time security and privacy are preserved. Another outcome of this project was that the proposed architecture must be very modular, with highly independent modules that can be reused per se on other type of solutions or platform architectures. **Figure 4** shows an example of smart city infrastructure [21].

In this infrastructure, a data platform (the Smart City's IoT platform) receives data from various sensors and sends commands to actuators in the city. The platform offers interfaces for various kinds of services in the city, such as metering and control of energy consumption, transportation links, or traffic.

A smart city platform, monitoring and controlling large parts of the city, introduces several new security and privacy threats to the city's infrastructure and to its citizens. The previously mentioned fundamental principles of information

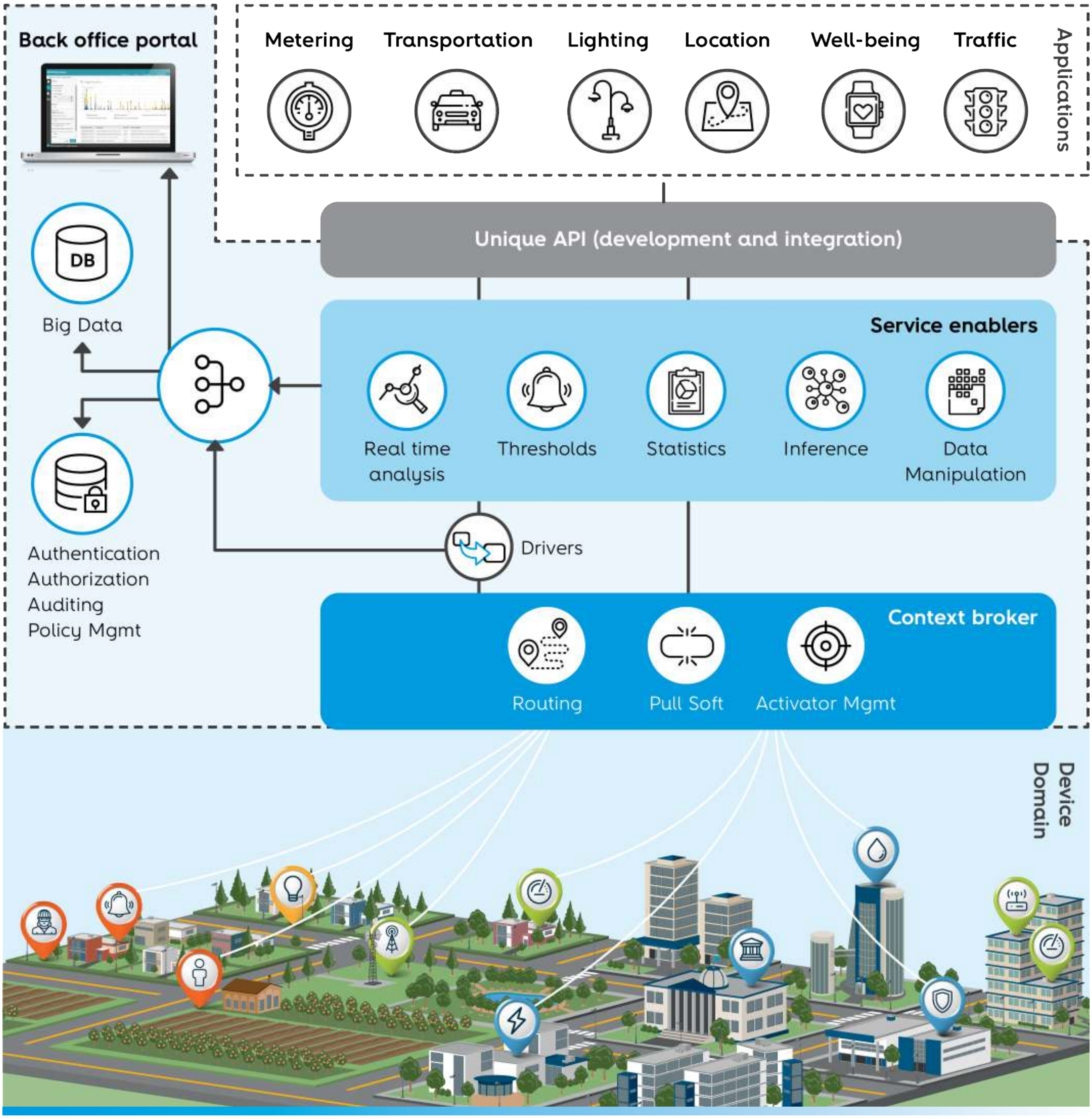


FIGURE 4 – Visualization of a city infrastructure with sensors, IoT platform and applications

security - confidentiality, integrity and availability - need also to be applied to several specific aspects of a smart city.

Every critical city infrastructure must be protected against malicious attacks with security mechanisms in the IoT platform. Furthermore, the platform must control the access to private information of users and subsystems. In general, attacks can target the IoT infrastructure at any point, from devices in the field, to communication channels, or servers. The attack might try to sabotage or compromise subsystems to take over control of certain aspects of the city. Another target may be the data in the IoT platform, which could compromise the privacy of the stakeholders and citizens.

The following paragraphs take a closer look at the threats, differentiating between external and internal threats.

External threats

As a smart city IoT platform grants access to critical infrastructure and confidential data, it is likely to become a target of external attacks. The smart city information platform will have to be resistant to external attackers. These can attack devices or communication channels. In particular the following issues need to be taken into account:

- **Unauthorized external data access:** External attackers may try to access private data from users, components or subsystems of the IoT environment. For instance, the energy consumption of city areas or even single houses is potentially interesting for unwanted commercial use cases. The platform stores the information sent by the sensors, so there is the risk that an attacker tries to access or corrupt private data;

- **Unauthorized device control:** Several actors are integrated in the smart city environment and are controlled automatically or via remote control. These could be display units, traffic lights, heating systems or even fire doors. Mis-utilization of these devices by external attackers must be prevented by the platform under all circumstances;

- **Hacking and Sabotage:** There are several conceivable scenarios where the city infrastructure is sabotaged by external attackers. The smart city platform will have to fend off denial-of-service or man-in-the-middle attacks. The platform must offer suitable mechanisms for intrusion prevention and data encryption to prevent failure of subsystems caused by unauthorized outsiders. Once parts of the system are compromised by the attacker, they can be used to attack the full smart city infrastructure, now acting as an internal attacker.

Internal threats

The complexity and multiplicity of components, actors and users of a smart city environment are huge. Several internal security issues have to be considered. Internal adversaries are especially dangerous as they may have detailed knowledge about the infrastructure, direct access to or control of the systems of the city infrastructure, or of the IoT platform. They may also have access to some of the system's access credentials, like digital certificates, passwords, etc. Insider attackers can be either legitimate but malicious subsystems users or administrators, or hackers who have already compromised parts of the system.

Recent reports on malware and backdoors in various systems show that hackers can easily manage to compromise at least parts of a system. In a IoT system, in particular, internal IoT devices or subsystems, with limited/restricted access from the exterior, are more difficult to physically protect and are often the weakest link.

Once a part or a subsystem is compromised, hackers can act as internal attackers to the rest of the infrastructure. The defence in depth principle, mentioned before, must be followed to avoid that the compromise of subsystems poses a major threat to the full infrastructure. It is important to follow security by design principles when building a complex infrastructure with multiple layers of security. The following security aspects must also be taken into consideration in an IoT environment:

- **Unauthorized internal data access:** Internal adversaries may have the possibility to bypass certain access control mechanisms and therefore have access to raw data. If the data itself is protected by cryptographic means, the access to meaningful plaintext is still difficult;

- **Violation of data and device integrity:** Data integrity is the maintenance and assurance of data accuracy and consistency over its entire lifecycle. Device integrity consists on the assurance that the device software/firmware is updated and has not been otherwise tampered with. These are of paramount importance because the integration of several subsystems into one platform threatens data integrity of components with unexpected side effects. Software errors or hardware failures should not influence data or communication of other components.

IoT security good practices

In the next sections we present a list of good practices and recomendations from the IT domain adapted to the IoT reality. The proposed security good practices are separated into the three phases of the lifecycle of devices and services: the development of IoT devices and services, their integration into the IoT network, and the use of the devices and services until disposal [22]. These good practices do not replace security and privacy by design, presented earlier, but are meant as an extension of them.

Good practices for the development of IoT devices and services: security of the development lifecycle

These recommendations are to be applied by device vendors and service providers.

Design phase

a) Separate security functions from other functions: security functions should have clear and limited interfaces to the other functionalities;

b) Make assumptions about the security requirements explicit: consider known limitations on the device usage, as well as constraints from the environment;

c) Consider third-party review by security specialists for developers with limited security experience;

d) Prepare user interactions with the products or services: developers must prototype the user interface as soon as possible in order to identify ways to help users on security issues.

Development phase

a) Use configuration management tools, and leverage upon development environments such as compilers or static analysers to highlight possible vulnerabilities within static source code, and guarantee the application and maintenance of security policies on systems, applications and network devices;

b) Take security into account when choosing your programming language; when available, leverage upon the operating system security functions;

c) Use standard, secure frameworks or stacks whenever possible – do not re-implement security functions;

d) Ensure team security awareness training in order to increase their understanding and practical implementation of security best practices.

Testing phase

a) Test security functions compliance;

b) Perform additional security audits and penetration testing;

c) Perform a privacy impact assessment.

Good practices for the development of IoT devices and services: security functions for hardware and software

a) **Security audit:** security events must be logged, and users should be notified whenever needed;

b) Communication protection: communication should be protected against disclosure, modification, replay and denial of service;

c) Cryptography: confidentiality, integrity and authenticity must be protected by using strong and standard cryptography. Keys must be managed securely, and the use of a trust infrastructure, such as Public Key Infrastructure (PKI), is encouraged;

d) User data protection: the integrity, confidentiality and authenticity of user data must be protected. Confidentiality protection must be defined with regards to privacy issues. Apply Privacy by Design principles;

e) Identification, authentication and authorization: strong authentication methods must be used, as well as access control mechanisms. Passwords and sessions should be managed accordingly;

f) Self-protection: hardware and software self-protection measures should be in place to protect previous security functions. Data used to enforce these security functions should be protected, and hardening should be used to reduce the attack surface.

Good practices for integration of devices in the IoT network

a) Minimum reliability: hardware must provide basic reliability measures to resist outages and jamming; Software components must handle data changes without failure, errors and improper functioning;

b) Trust relationships: use a trust infrastructure; use secure pairing for devices; check the security assumptions at installation time;

c) Network security: Introduce network elements to mitigate the propagation of attacks: gateways, proxy-firewalls, etc; use network segregation as an additional security measure.

Good practices for the usage until end-of-life

Protection of data exchanges

a) Ensure access rights;

b) Leverage on gateways to reduce the network exposure of the weaker devices.

Operational security and maintenance

a) Vulnerability survey - Perform vulnerability survey; check the security assumptions regularly during lifetime;

b) Security updates - Protect the software update mechanism;

c) Remote interfaces protection - Provide user-friendly interfaces for device and services security management; protect remote monitoring interfaces;

d) Security management system for support infrastructures - Rely on existing sources for security good practices - ISO 27001, OWASP and others - in order to secure infrastructures.

Control of user data

Provide secure backup and/or deletion of the data stored/processed by the device (and by associated cloud services) during the operation and at end-of-life.

Conclusions

Better security and privacy are long due requirements for the digital world. Any organization must account for market value based on credibility and trust considering their compliance to security regulations.

GDPR addresses only one type of data, personal data, but the principles that form this regulation's basis can also be applied to other scenarios: IoT data, industrial data, national security data, etc. Furthermore, GDPR compliance will force organizations into rethinking and reorganizing their information security governance models as this is implied by the mandatory DPIA.

Meanwhile, the number of IoT devices will continue to grow, as well as their applications in critical infrastructures, like those needed to run a city. This will raise the level of risk of those infrastructures. IoT devices are an attractive target for botnets because security has not been a priority for the device manufacturers. Also, their vulnerabilities tend to be left unprotected due to the lack of built-in mechanisms to receive automatic updates, and they are often forgotten once installed. And yet, the 2017 Irdeto Global Consumer IoT Security Survey [23] found that 90% of consumers believe it is important that a connected device has security built into the product, which indicates that consumers are aware of cyber threats that may target their connected devices and that security of those devices is essential to keep their personal data safe. The potential for profit in attacking IoT devices will increase and, as a result, many organizations will need to enhance security strategies to protect against larger-scale attacks targeting IoT technologies. Therefore, security must be a core consideration in the design and manufacture of IoT devices.

A list of recommendations should state that:

1. All stakeholders should reach a consensus on minimum security requirements;
2. Industry actors should support security-driven business models;
3. All actors should contribute to raising security awareness;
4. Industry actors should develop security assessment methods or frameworks;
5. Policy makers should clarify the legal aspects of smart home and IoT environments;
6. Industry research and publicly-funded initiatives should integrate cyber security in R&D projects related to smart home and IoT.

Secure IoT data management for smart cities is very important in order to make smart city IoT successful. Depending on the context, several approaches may be followed:

- Constraints of devices and infrastructures;
- Scale of required solutions;
- Trust in IoT system.

The list of security and privacy principles, good practices, and recommendations presented here reflect a shift-left paradigm, or a DevSecOps approach [24]. This can be defined as a mindset where everyone is responsible for security: instead of operating security in a traditional waterfall process model, where the security defects of a system are to be determined by security staff after the design and development phases, and to be corrected before the system is released, security is taken into consideration in all phases and iterations of the development and delivery process, and promoted by the communication and collaboration between product management, software development, security and operations professionals. This paradigm allows security and privacy features to keep up with the speed, flexibility, scalability and capabilities of the increasingly complex digital world.

One thing is certain: there is no one-size-fits-all solution and there's yet a lot of work that needs to be done, but if organizations had followed these guidelines and good practice recommendations in the past years, most certainly many of the threats mentioned earlier in this article wouldn't have been successful, or, if they had, would have been detected earlier, their impact would have been lower, and organizations would not have been in the headlines for the wrong reasons.

Finally, following the presented guidelines and with a little push from legislators to increase the motivation of manufacturers and organizations in pursuing them, we can aim for a better security and privacy in this new digital world.

References

[1] N. Falliere, L. O. Murchu and E. Chien, "W32.Stuxnet Dossier," Symantec Security Response, Cupertino, CA - USA, 2011.

[2] Admin CyberScout, "Two years on, lessons learned from the playstation data breach," CyberScout, LLC., 12 May 2013. [Online]. Available: http://cyberscout.com/education/blog/two-years-on-lessons-learned-from-the-playstation-data-breach.

[3] M. Kan, "Yahoo execs botched its response to 2014 breach, investigation finds," CSO, from IDG Communications, Inc., 2 March 2017. [Online]. Available: http://www.csoonline.com/article/3176181/security/yahoo-execs-botched-its-response-to-2014-breach-investigation-finds.html.

[4] B. T. Lee, "Russia's attempt to hack US election officials, explained," Vox, 6 June 2017. [Online]. Available: https://www.vox.com/new-money/2017/6/6/15745888/russia-election-hacking-leak.

[5] Z. Zorz, "UK sports retailer Sports Direct hacked," Help Net Security, 9 February 2017. [Online]. Available: https://www.helpnetsecurity.com/2017/02/09/sports-direct-hacked/.

[6] S. Thielman and C. Johnston, "Major cyber attack disrupts internet service across Europe and US," The Guardian, 21 October 2016. [Online]. Available: https://www.theguardian.com/technology/2016/oct/21/ddos-attack-dyn-internet-denial-service.

[7] J. Leyden, "Sh... IoT just got real: Mirai botnet attacks targeting multiple ISPs," The Register, 2 December 2016. [Online]. Available: https://www.theregister.co.uk/2016/12/02/broadband_mirai_takedown_analysis/.

[8] A. Jesdanun, "WikiLeaks: CIA has targeted everyday gadgets for snooping," The Seattle Times, 7 March 2017. [Online]. Available: http://www.seattletimes.com/nation-world/wikileaks-cia-has-targeted-everyday-gadgets-for-snooping/?_sm_au_=isM6s61Nq5QT7Pf5.

[9] J. Kastrenakes, R. Brandom, C. Lecher, R. McCormick, A. Liptak and T. Warren, "WannaCry Ransomware: all the updates on the cyberattack," The Verge, 14 June 2017. [Online]. Available: https://www.theverge.com/2017/5/14/15638026/wannacry-ransomware-updates-cyberattack-cybersecurity?_sm_au_=isM6s61Nq5QT7Pf5.

[10] R. Brandom, "A new ransomware attack is infecting airlines, banks, and utilities across Europe," The Verge, 27 June 2017. [Online]. Available: https://www.theverge.com/2017/6/27/15879480/petrwrap-virus-ukraine-ransomware-attack-europe-wannacry.

[11] Technology, BBC News, "Personal details of nearly 200 million US citizens exposed," BBC, 19 June 2017. [Online]. Available: http://www.bbc.com/news/technology-40331215?ocid=socialflow_facebook&ns_mchannel=social&ns_campaign=bbcnews&ns_source=facebook&_sm_au_=isM6s61Nq5QT7Pf5.

[12] Symantec, "Internet Security Threat Report - Volume 22," Symantec Corporation, Mountain View, CA, USA, 2017.

[13] McAfee Labs, "McAfee Labs 2017 Threats Preditctions," Intel Security, Santa Clara, CA, USA, November 2016.

[14] Council of the European Union , European Parliament, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC," *Official Journal of the European Union*, p. 88, 27 April 2016.

[15] The Article 29 Data Protection Working Party, "Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679," European Commission Article 29 Working Party, Brussels, 2017.

[16] i-SCOOP, "GDPR: online guide to the EU GDPR and the compliance steps to take," i-SCOOP BVBA, 26 March 2017. [Online]. Available: https://www.i-scoop.eu/gdpr/?_sm_au_=isM6s61Nq5QT7Pf5.

[17] P. Middleton, J. F. Hines, B. Tratz-Ryan, E. Goodness, D. Freeman, M. Yamaji, A. McIntyre, A. Gupta, D. Rueb and T. Tsai, "Forecast: Internet of Things — Endpoints and Associated Services, Worldwide," Gartner, Inc., Stamford, CT, USA, 2016.

[18] OWASP Foundation, "Security by Design Principles," OWASP Foundation, 3 August 2016. [Online]. Available: https://www.owasp.org/index.php/Security_by_Design_Principles.

[19] A. Cavoukian, "Privacy by Design - Strong Privacy Protection - Now, and Well into the Future," Information & Privacy Commissioner, Ontario, Canada, 2011.

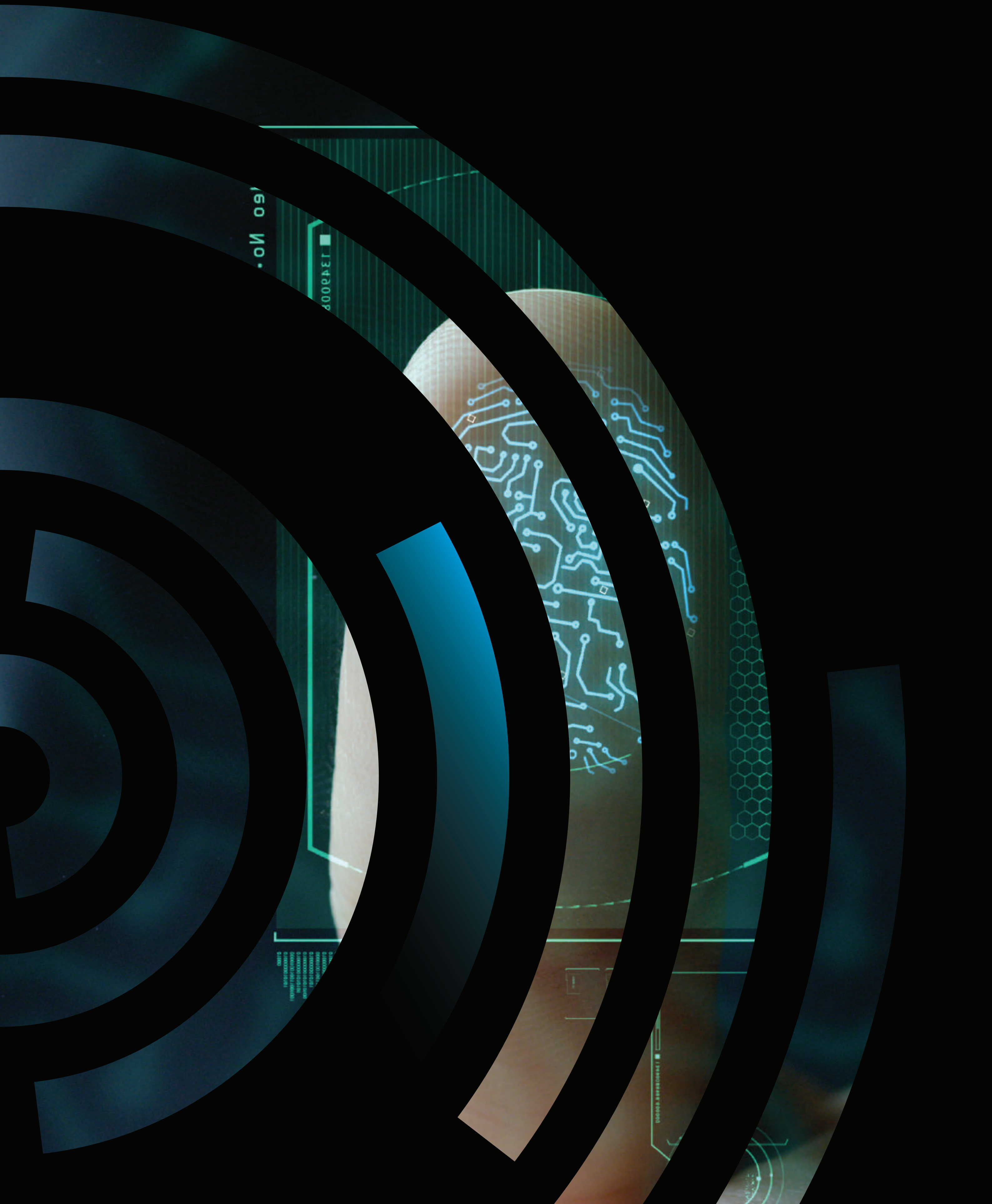
[20] P. N. Howard, Pax Technica: How the Internet of Things May Set Us Free Or Lock Us Up, New Haven, Connecticut, United States: Yale University Press, 2015.

[21] J. M. Bohli, A. Skarmeta, M. Moreno, D. García and P. Langendörfer, "SMARTIE Project: Secure IoT Data Management for Smart Cities," in *International Conference on Recent Advances in Internet of Things (RioT),* Singapore, 2015.

[22] C. Lévy-Bencheton, E. Darra, G. Tétu, G. Dufay and M. Alattar, "Security and Resilience of Smart Home Environments - Good Practices and Recommendations," European Union Agency For Network And Information Security, Heraklion Crete, Greece, 2015.

[23] Irdeto, "Irdeto Global Consumer IoT Security Survey: Insights to Help Secure Today's Connected World," Irdeto, Hoofddorp, The Netherlands, 2017.

[24] D. Shackleford, A DevSecOps Playbook, S. I. I. R. Room, Ed., SANS Institute, 2016, p. 18.



11

A holistic study of cable and mobile identity credentials

Isabel Borges, Altice Labs
isabel@alticelabs.com

Fernando Santiago, Altice Labs
fernando-m-santiago@alticelabs.com

Ralph Brown, CableLabs
r.brown@cablelabs.com

José Paulo Pires, Altice PT
jose-pires@telecom.pt

The article herein results from a joint collaboration project between Altice Labs and CableLabs having as main goals the exploration of possible synergies that can exist between the SIM-based and PKI-based approaches and the identification of possible convergence areas between those security forms, enabling gains in simplicity, flexibility and efficiency for both telco and customers.

Keywords

eSIM; eUICC; DOCSIS; Digital Certificates; Cable Networks; Authentication; Privacy; Identity Credentials

Overview

Today secure digital identity is implemented using different technologies to enable authentication and authorization to specific resources access. Mobile Network Operators (MNO) have already a strong security infrastructure supporting digital mobile subscription identity, having the Subscriber Identity Module (SIM) as a core element. Cable Service Providers have in place security infrastructures based on Public Key Infrastructure (PKI) to manage digital certificates and authentication of hardware devices, such as cable modems.

With the evolution of traditional SIM Cards to embedded SIM (eSIM), new application scenarios are being addressed, namely those concerned with Machine-to-Machine (M2M) and Consumer Electronics. In these scenarios, the subscriber identity might be used for identifying their devices.

Digital certificates in Wi-Fi Hotspot 2.0 are being used to provide subscriber identification that associates the device with the user and account.

The following diagram (**Figure 1**) presents a possible trend and an eventual evolution movement. SIM-based security on mobile networks is extended from individual (subscriber) to device identification and PKI-based security on cable networks is extending from device to individual identification.

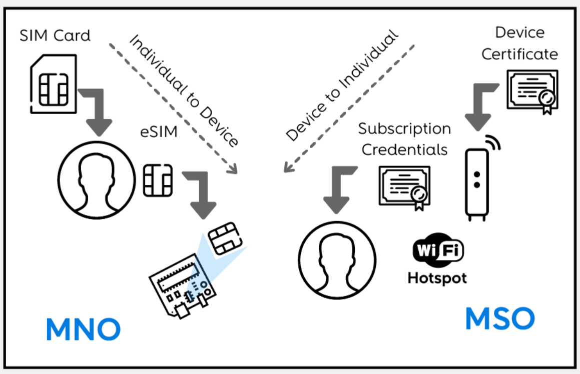


FIGURE 1 – Identity convergence

Convergent network operators will have to deal with provisioning and management of multiple forms of identity from users, devices and applications. In this Altice Labs and CableLabs joint study [1], several scenarios have been explored regarding the evolution trends and convergence of digital identity management on future ecosystems of network service providers.

SIM-based security on mobile networks

The first generation of mobile handsets had no SIM card or support for authentication and encryption but all the subsequent generations require the presence of a SIM card to grant the user access to the mobile network and services. The SIM contains the International Mobile Subscriber Identity (IMSI) as well as credentials that are required for the identification and authentication of the subscriber, and for contracted services access authorization. A SIM is a detachable microprocessor card, resilient to data manipulation, containing its own operating system, storage and built-in security features that prevent unauthorized users to access, retrieve, copy or modify the subscriber IMSI and credentials.

A recent evolution of the SIM has been driven by M2M needs and use cases. The main new characteristics are:

- Embedded: physically integrated into the device and cannot be removed;
- Remotely provisioned: updatable over the air with support for storing one or more operator profiles.

The evolution is also driven by consumer market, mainly on wearables and handset technologies.

eUICC

An embedded Universal Integrated Circuit Card (eUICC), also known as eSIM, is a new generation of SIM that supports Over-The-Air (OTA) provisioning of the operator profile. The GSMA [2] defines the eUICC as "*A small trusted hardware component, which may be soldered into mobile devices, to run the secure network access application(s) and enable the secure changing of subscription identity and other subscription data*". It is an agnostic hardware, which may be a re-programmable SIM application contained in a secure element, a surface mounted M2M SIM or even a removable SIM card.

The internal high-level architecture of the eUICC is represented in **Figure 2**. The operator profiles are isolated within the eUICC and it is impossible for any profile to access the applications or data of any other profile stored on the eUICC.

The ISD-P is the on-card representative of the MNO, or Subscription Manager – Data Preparation (SM-DP) if delegated by the MNO. The ISD-R is the on-card representative of the Subscription Manager - Secure Routing (SM-SR). The SM-SR ensures the secure transport of both the eUICC platform and eUICC profile management commands. A profile includes the file system; NAA, like SIM, USIM, ISIM; Applets, such as electronic purse, multimedia, payment or ticketing applications; SSD, that is used when there are agreements between operators; and CASD, that provides services only to security domains of the Profile and only when the Profile is in the enabled state.

Remote provisioning

According to GSMA [3], Remote SIM Provision (RSP) is "*the ability to remotely change the SIM profile on a deployed SIM card without having to physically change the SIM card itself. This capability is hardware agnostic and can be deployed on removable and non-removable UICCs*".

GSMA together with ETSI defined a set of use cases regarding RSP on M2M and Consumer scenarios [4] [5] [6] - see **Table 1**.

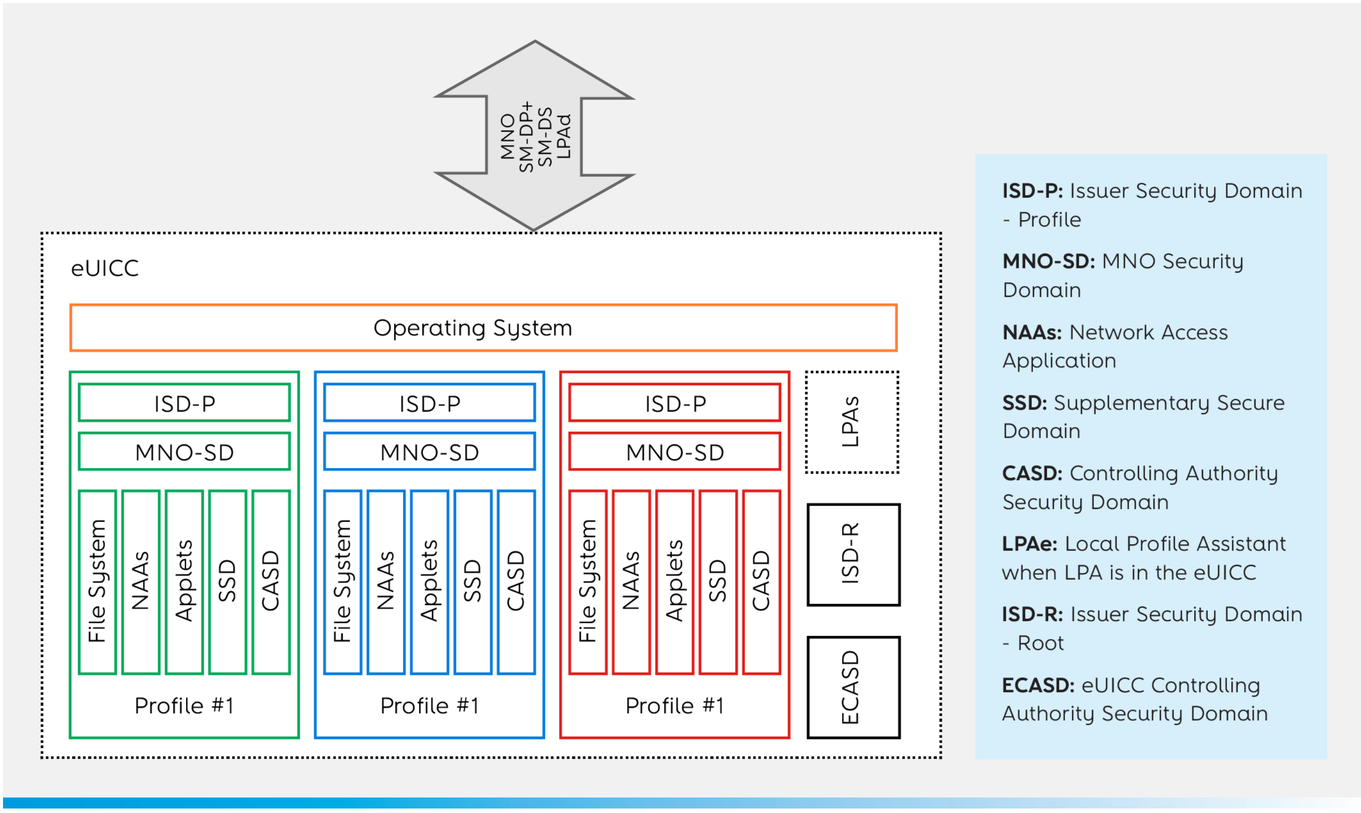


FIGURE 2 – eUICC high level architecture

M2M		
#	Use Case	Description
1	Provisioning of multiple eUICCs	M2M Service Provider sets up subscriptions for a number of connected M2M devices to start telecommunication services with a first MNO
2	Provisioning of an eUICC for a first subscription with a new connected device	An end user purchases a new type of communications or connected device from an OEM together with a subscription to provide first services to this device
3	Change of subscription for a device	A subscriber changes the contract and thus subscription for the device to stop services with the current MNO and start services with a new MNO.
4	Stop subscription	The M2M customer sells his device and stops the subscription for services from the current mobile MNO.
5	Transfer subscription	The M2M customer transfer subscription between devices
Consumer		
#	Use Case	Description
1	First purchase of a subscription to a new device	The subscriber purchases a device from a device manufacturer and requests a subscription to gain network access
2	Purchase an additional subscription for a device	The subscriber wants to purchase an addition subscription to the same device, possibly with a different MNO
3	Switch subscription	The subscriber wishes to switch from one subscription to another subscription on the same device
4	Terminate a subscription	The subscriber wants to terminate a subscription with a MNO
5	MNO swapping	In a swapping model (national and international), a subscriber is able to download multiple profiles onto the eUICC and then swap between profiles as required.
6	Replacing a device but not Service Provider	The subscriber is replacing a device but wants to keep the subscriber relationship with the MNO unchanged
7	Adding a device to an existing contract	The subscriber wants to add an additional device to an existing contract, using the same contractual terms for the new device
8	MNO swapping	In a swapping model (national and international), a subscriber is able to download multiple profiles onto the eUICC and then swap between profiles as required.

TABLE 1 – M2M and Consumer use cases for RSP

GSMA considered two architectures for RSP:

- **M2M** - often used in unsupervised locations, where devices could be under external conditions, such as weather, temperature, vibration, among others, and usually in push operations initiated from the MNO/service provider side. M2M architecture is also very often associated with corporate market, where the management of a large fleet of devices is requested;
- **Consumer primary/companion devices** - including wearables, smartphones, tablets, etc, where the eUICC solution will optimize the device space, simplifying the logistics

and distribution processes. The aim of this architecture is to give to the user the ability to easily choose and change the MNO/service provider.

The RSP M2M architecture, depicted in **Figure 3**, comprises 3 main elements and entities.

- The **eUICC** - capable of supporting remote provisioning;
- The **SM-DP** - Responsible to prepare the profiles and manage the secure download and installation of these profiles onto the eUICC. The SM-DP and MNO are the only entities allowed to establish a secure and authenticated channel to the eUICC to manage a profile;
- The **SM-SR** - interfaces with eUICC and its main role is to deliver the encrypted operator profile and credentials to the eUICC and manage the eUICC (enable, disable, delete) during the product's lifetime.

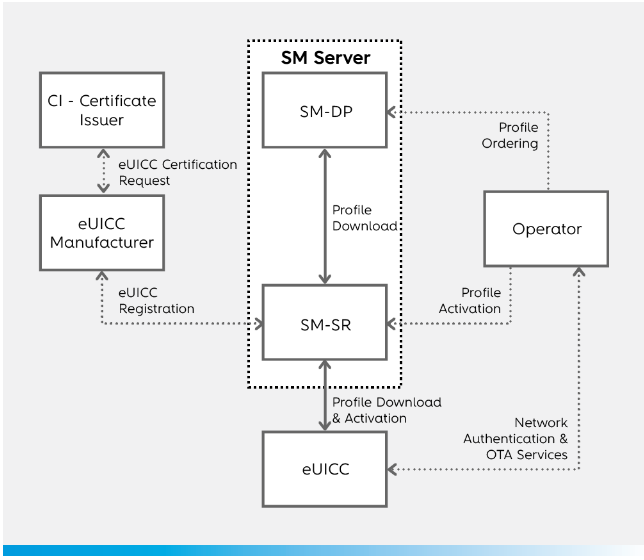


FIGURE 3 – eUICC RSP architecture

The consumer architecture is similar to the M2M architecture with the following main differences:

- The functions of the SM-SR are undertaken by the SM-DP+ and the Local Profile Assistant (LPA) within the device, and a device is always reachable by any qualified SM-DP+;
- This architecture is based in a PKI model instead of pre-shared keys. The secure channel is initiated based on challenge and signed certificate public keys;
- The Subscription Manager - Discovery Service (SM-DS) is responsible for providing addresses of one or more SM-DP+ to a local discovery service.
- There are options to use efficient connections to the SM-DP+ to improve the scalability of server infrastructure.

PKI-based security on cable networks

The CableLabs Data Over Cable System Interface Specification (DOCSIS) specifies the equipment interface requirements for broadband internet access services. The interface requirements are specified between the Cable Modem Termination System (CMTS), which is in the cable operator's network, and the Cable Modem (CM), which is located at the customer premise and interfaces to the Customer Premise Equipment (CPE). This architecture is shown in **Figure 4**.

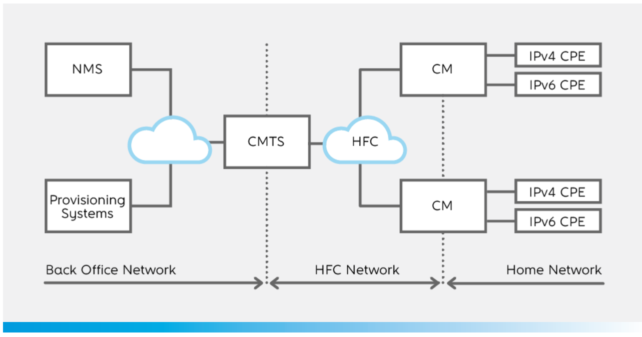


FIGURE 4 – DOCSIS architecture

There have been five generations of DOCSIS specifications issued by CableLabs [7] [8] [9] [10] [11]. These specifications have also been adopted as both national and international standards.

Of these specifications, the DOCSIS 3.1 Security Specification is most relevant to this discussion. The primary goals of DOCSIS security are to provide CM users with data privacy across the cable network, to prevent unauthorized users from gaining access to the cable network, and to provide privacy of customer traffic, the integrity of software downloads, and prevent theft of service.

DOCSIS provides a number of tools to support these goals such as:

- Traffic encryption for privacy/confidentiality, with Baseline Privacy Interface Plus (BPI+);
- Secure software download to assure a valid CM image;
- Configuration file authentication to help secure the provisioning process.

DOCSIS BPI+ begins immediately after the CM finishes registering if Early Authentication and Encryption (EAE) is not enabled. BPI+ provides the method to secure the link between the CMTS and the CM. The three main steps within BPI+ are CM authentication, key exchange and traffic encryption.

Digital certificates

CM Authentication is performed using RSA public cryptography and X.509 certificates. Each CM contains a unique certificate and corresponding private key, the certificate of the Certification Authority (CA) that issued it, and additional information identifying the CM.

The CM sends these certificates and information to the CMTS for validation. The CMTS validates them by verifying certificate parameters and that they chain to the Root CA. Once these certificates have been validated, the CM is considered trusted enough to continue the BPI+ initialization process. BPI is used to exchange keying material between the CMTS and the CM for purposes of traffic encryption. BPI uses two interdependent state machines: the authorization state machine and the Traffic Encryption Key (TEK) state machine.

The DOCSIS PKIs use x509 v3 certificates for authentication and software downloads. The hierarchy of the DOCSIS PKI includes the DOCSIS Root CA, a number of sub CA, and end entity certificates. The Root CA is installed in both the CMTS and CM. The sub CA and end entity device certificate are installed in the CM.

The organization of the CableLabs PKI hierarchy has two versions: a legacy version established for DOCSIS 3.0 and earlier, PacketCable, OpenCable, and DPoE; and a subsequent version for DOCSIS 3.1, Remote PHY and subsequent CableLabs projects.

The new PKI established for DOCSIS 3.1 and subsequent CableLabs projects unifies much of the PKI hierarchy. It uses one master Root CA, with multiple Intermediate CA that can be used for multiple projects; the certificates have nested validity periods and use a centralized intermediate CA model (all Intermediate CA are operated by CableLabs). This PKI is used for end entity device certificates, server certificates, and Code Verification Certificates (CVC) used to sign secure software downloads. **Figure 5** shows the PKI hierarchy.

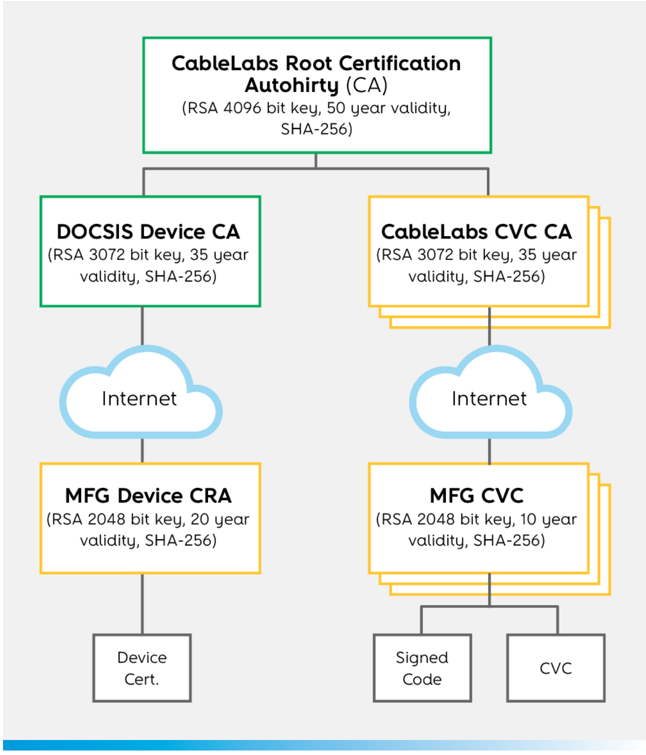


FIGURE 5 – New DOCSIS PKI

Provisioning process

The DOCSIS provisioning systems consist of the following elements:

- The Dynamic Host Configuration Protocol (DHCP) server - provides the CM with initial configuration information, which includes IP addresses when the CM boots;
- Config file server - a Trivial File Transfer Protocol (TFTP) server used to download configuration files when a CM boots and permits the configuration of CM parameters;
- Software download server - used to download software to the CM;
- Time protocol server - provides current time of day to CM;
- Certificate revocation server - provides certificate status;
- The Simple Network Management Protocol (SNMP) manager - allows operator to configure and monitor SNMP agents within the system (CMTS and CM) and provides CM with initial configuration information that includes IP addresses when the CM boots;
- Syslog Server - collects operation messages of devices;
- Internet Protocol Detail Record (IPDR) collector - used to collect bulk statistics.

The DOCSIS cable modem initialization and provisioning is based on there being at least two IP addresses allocated to the customer premise equipment. The first IP address is private and is allocated to the CM itself for management purposes and the second (third, fourth, etc.) are allocated to the CPE behind the cable modem (computer, router, etc.).

The following steps are taken at the CM initialization:

- The CM scans for a downstream channel on the cable network;
- Once the downstream channel is acquired begin timing and power level ranging on the upstream channel;
- Perform EAE if enabled;
- Establish IP connectivity for the CM;
- Initialize Baseline Privacy if EAE disabled and BPI is enabled.

Hotspot 2.0 security on Wi-Fi networks

The Hotspot 2.0, also known as Passpoint 2.0, and related technology is supported and specified collectively by the Wi-Fi Alliance (WFA) and the Wireless Broadband Alliance (WBA). Hotspot 2.0 is intended to enable automatic network selection and secured attachment for Wi-Fi networks.
WBA runs the Next Generation Hotspot (NGH) program covering end-to-end Wi-Fi architecture and testing, including Hotspot 2.0, roaming & accounting, location, user experience and device compliance program.

Authentication models

Hotspot 2.0 security has three forms of authentication:

- EAP-TLS – Device Certificate
- EAP-TTLS – Username/Password
- EAP-SIM, EAP-AKA, EAP-AKA' – based on SIM cards

Passpoint subscriptions are installed on devices either pre-sale or via the Online Sign-Up (OSU) server.

WFA certified devices come with Passpoint 2.0 trust root certificates that are used by devices to authenticate OSU servers. The OSU server can provision devices with CA certificates used by the device to authenticate the home AAA server. The OSU server can provision devices with user certificates used by the network to authenticate the user.

On-line sign up process

The Hotspot 2.0 OSU process is as follows:

- A mobile device scans the available Wi-Fi networks for those for which it is already subscribed and attaches to those that match its subscription policies;
- If the mobile device does not have a subscription already for the available Wi-Fi networks, it begins the OSU process:
    - It provides a list of the available Passpoint Wi-Fi networks to the user;
    - The user then selects the Passpoint Wi-Fi network he/she would like to use;
    - The Network Operator of the selected Wi-Fi network may have preferences for how the user identifies himself and what subscription options may be offered, e.g. per hour or per day fee, free access under a cable subscription, etc;
    - Once the user has completed this information, the subscription is downloaded and the user is connected;
- The OSU process is only required once, the first time the user activates the mobile device on a particular network operator's Passpoint network. After this, no user interaction is required and the mobile device connects automatically to any access point advertising this particular Passpoint network.

Security approaches comparison

A direct comparison of features and functions between SIM-based and PKI-based security systems can be made across several dimensions. These include differences and similarities in physical architecture, actual security keys used, how ephemeral keys are negotiated and assigned, and the way security functions (authentication, authorization, and encryption) are executed.

Next paragraphs describe the comparison between SIM-based and PKI-based systems. Specifically, Hotspot2.0 is not covered since it is a solution that already uses PKI and digital certificates.

Generally, the eSIM processes specified by GSMA appears, right now, to be highly optimized towards enabling 4G/LTE services for consumers and M2M. In eUICC RSP architecture, PKI and digital certificates are used to authenticate all involved entities as part of a single, secure and certified system. In the consumer eUICC RSP architecture, PKI is the security approach used in the download of operator profiles from SM-DP+ to eUICC, while in the M2M eUICC RSP architecture, in addition to using PKI to encrypt the operator profiles from SM-DP to eUICC, the SM-SR uses a shared-key encryption generated at the time of eUICC manufacture. Shared-key encryption is considered stronger, although it requires a more complex implementation and the existence of SM-SR, so it was recommended by the GSMA Security and Fraud Group for the M2M architecture. In the consumer architecture, there is a simplification, by using only PKI.

DOCSIS does not have an equivalent capability to support access profiles to/from multiple access providers much less other trust ecosystems. The GSMA eSIM model seems to do that well. WFA Hotspot 2.0 would allow leveraging multiple network providers.

While LTE, Wi-Fi, DOCSIS, and PON all are, at least, partially shared access infrastructures, however, DOCSIS and PON are wired plant and as such do tether to specific geographical locations.

Also, in DOCSIS systems, the public/private key pairs and their associated PKI certificate is used to implement, to at least some degree, integrity functions. Firmware updates in DOCSIS are signed using PKI, and so integrity and authentication are provided. However, configurations are not. On the other hand, eUICC/SIM ecosystem provides additional security mechanisms for protecting the application layer that can be used by non-telco providers (e.g. banking, automakers, etc). Such capabilities include security domains that can store private information (for instance certificates) in a tamper-resistant manner. The eSIM allows the existence of Java applets, that could be used for securely execute updates and configurations downloads.

Both SIM-based and PKI-based approaches implement strong confidentiality measures to secure communications channels. DOCSIS devices do not typically house any data and so do not encrypt data at rest. However, some sophisticated LTE devices (aka, smartphones) house significant user data. Encryption of user data, however, may not leverage the GSMA eSIM keys. The eUICC allows IoT service providers to leverage the MNO infrastructure and personalization chain to provide non-network operator applications with specific credentials stored in SIM cards. These credentials can enable the use of the SIM card as a trust anchor to verify the integrity of other components and to facilitate the deployment and management of public/private key pairs (as needed e.g. for application layer security) to other components within the IoT service architecture. Generic Bootstrapping Architecture (GBA) allows applications to authenticate using a time-limited session key created from SIM card stored credentials that can be used for setting up a secure tunnel.

Individual versus device identification

Individual identity

Personal identifications in today's society can take many different forms. However, in the scope of our study, we are focused on SIM-based and PKI-based processes of identification. Identity is the digital representation of data relating to a specific individual, allowing that he/she could be authenticated and authorized to access a set of services.

SIM-based

Individual authentication in mobile scenarios is challenging in the sense that the physical location of the device cannot be used as a trusted part of the authentication process, devices can be cloned, PINs guessed, etc. Individual authentication can be achieved by two Factor Authentication (2FA) or even Multi-Factor Authentication (MFA) mechanisms, like PIN + fingerprint or PIN + certificates + retina reading.

PKI-based

Individual authentication and authorization on systems relying heavily on the physical location of the device (such as leased lines or cable services) is usually not very common. Usually, individuals are allowed access once the device is validated. Two different users can access the service undifferentiated of their preferences. Individual authentication can be done by OTT services.

Device identity

Device identity provides for the unique and verifiable identification of devices, most often referring to end devices. The device identity provides the foundation for implementing security functions such as authentication, authorization, digitally signing data (usually for integrity and non-repudiation purposes), encryption, and policy enforcement. The identity may include attributes that describe how the device may be used or to govern its functionality. The aggregate of this information may be digitally signed. This signature allows verification of the integrity of the identity. Ultimately, this relies on a secret. Most commonly, this information in aggregate is referred to as a certificate.

SIM-based

eUICCs are uniquely identified by the eUICC-ID (EID). The EID with the required information to identify the mobile device will be stored in the ECASD. The ECASD will also include several other security fields.

PKI-based

Traditional cable modem identification includes MAC address, RSA public encryption key, serial number, manufacturer ID, and other attributes (such as supported cryptographic functions). To ensure strong binding to a device, these are incorporated in an X.509 digital certificate bound to the device MAC address and signed by a sub-CA. Certificates are expected to last the lifetime of the device.

Collaborative work

Since the beginning of 2017, Altice Labs and CableLabs have been working together on a common project to address the potential synergies that may exist on the security convergence paradigm between individual and device identity management. This work has resulted in the first phase on a deliverable about the state-of-the-art of both, SIM-based and PKI-based architectures. On the second phase, the work comprised the definition of a set of scenarios to be further studied. In a nutshell the scenarios are:

- Use of eSIM as a secure element in the STB and Home Gateway (HGW), allowing remote provisioning of device and user credentials, supporting device authentication and service authorization, providing communication and application security and allowing secure storage of security credentials and data provisioned by one or more service providers;
- Use of digital certificates to provide devices (HGW, STB, mobile handsets, IoT devices, ...) with security services, to address equipment authentication, communication encryption, firmware images digital signature and user or device mobility;
- Design a convergent identity management and provisioning ecosystem architecture which provides a single and unified security management of the different security assets (Digital Certificates, SIM/eUICC, etc). In short, this ecosystem will provide security mediation between the customer and the service, unified customer identification, secure convergent authentication, bootstrapping trust to enable higher confidence profile or certificate deployment and bridging trust across ecosystems;
- Design a cloud-based Identity-as-a-Service (IDaaS) solution to provide strong security services to 3rd parties (OTT providers, IoT service providers, banking, web stores, ...), taking advantage of SIM/eSIM;
- Retail IoT Application. Some ecosystem consortiums for IoT will deploy PKI certificates, WFA 2.0 certificates/profiles, or GSMA profiles into devices. This is essential for a strong trust basis to be established to IoT endpoints. Once this starts to happen, access networks need to be able to determine that a device has a certificate/profile and apply the corresponding security controls. The services to be provided are authentication, privacy management, encryption, network segmentation/isolation and secure configuration, management and updating. Digital service providers will be able to address security for targeted verticals to include health care, wellness, automotive, security, and home automation. Better network hygiene will be achieved reducing cybersecurity operational costs. IoT security that just works well on supported access networks may create better customer loyalty.

From the above scenarios, the most relevant were chosen based on a priority criterion for both organizations.

CableLabs has prioritized the retail IoT application as CableLabs and its members believe that retail IoT presents the greatest cybersecurity risk to the cable network, cable operators, and their customers. Standardizing good security practices among IoT devices, for example through the Open Connectivity Foundation (OCF), will reduce the cybersecurity risks associated with retail IoT devices.

For Altice Labs and Altice PT, it is important to apply this knowledge in the specification and implementation of a PoC by applying a secure element (eSIM) in the Fiber Gateway. Altice Labs aims to validate the concept of having a centralized home platform able to fully control home automation in a secure way.

Conclusions

Whilst the observation that identity credentials in both the SIM and PKI ecosystems are evolving as originally hypothesized, the transition of current evolutions of these trust ecosystems is not yet complete. While GSMA is driving eUICC adoption in multiple kinds of devices (traditional handsets, M2M, and consumer devices) adoption beyond handsets remains limited today. Adoption of PKI credentials in Wi-Fi and consumer devices is similarly limited. Consequently, the evolution of these ecosystems is still nascent. CableLabs is considering the use of a secure element as part of the DOCSIS security architecture to increase the robustness of implementations with hardware. Meanwhile, Altice Labs is willing to validate the introduction of the eSIM in a future evolution of its Fiber Gateway product to provide security capabilities not only in smart home scenarios but also for enterprise customers. Convergence across the SIM and PKI ecosystems is immature and adoption of recent levels is limited. There is a trend towards PKI in the consumer marketplace in both the SIM (although limited to some operations related to management of operator profiles in the eUICC) and PKI ecosystems, so this may be the clearest opportunity for convergence or to evolve to a future-proof security solution.

In conclusion, the evolution in security and particularly in identity credentials can promote business opportunities that telco cannot afford to lose.

Acknowledgements

This paper was based on the state of the art report that was done with the contributions of the following staff:

- Mike Glenn, Steve Goeringer, Brionna Lopez, Brian Scriber, Massimiliano Pala and Luther Smith from CableLabs
- Luís Cortesão, Paula Cravo, Luís Miguel Silva and Filipe C. Pinto from Altice Labs

References

[1] CableLabs and Altice Labs, *Analysis and Summary on eSIM and PKI Architecture: Report on state of the art*, Version 1.0 ed., 2017.

[2] GSM Association, *Embedded SIM Task Force Requirements and Use Cases v1.0*, GSMA, 2011.

[3] GSMA Intelligence, *The future of the SIM: potential market and technology implications for the mobile ecosystem*, London, United Kingdom: GSMA Intelligence, 2017.

[4] GSM Association, *Embedded SIM Specification: Remote SIM Provisioning for M2M*, GSM Association, 2014.

[5] ETSI, *ETSI TS 103 383 V13.2.0 (2016-05) Smart Cards; Embedded UICC; Requirements Specification (Release 13)*, ETSI, 2016.

[6] Ofcom, *Reprogrammable SIMs: Technology, Evolution and Implications - Final Report*, CSMG, 2012.

[7] CableLabs, *DOCSIS 1.0 - Cable Modem to Customer Premise Equipment Interface Specification (2008); Baseline Privacy Interface Specification (2001); Operations Support System Interface Specification: RF Interface (2001); Baseline Privacy Interface MIB (2001); [...]*, CableLabs.

[8] CableLabs, DOCSIS 1.1 - Radio Frequency Interface Specification; Operations Support System Interface Specification, CabelLabs, 2005.

[9] CableLabs, *DOCSIS 2.0 - Radio Frequency Interface Specification (2009); Operations Support System Interface Specification (2008); Baseline Privacy Plus Interface Specification (2008)*, CableLabs.

[10] CableLabs, *DOCSIS 3.0 - Operations Support System Interface Specification; MAC and Upper Layer Protocols Interface Specification; Cable Modem to CPE Interface Specification; Physical Layer Specification; Security Specification*, CableLabs, 2017.

[11] CableLabs, *DOCSIS 3.1 - Physical Layer Specification; MAC and Upper Layer Protocols Interface Specification; Cable Modem Operations Support System Interface Specification; CCAP Operations Support System Interface Specification; Security Specification*, CableLabs, 2017.

12

OTT multimedia content delivery: a study

Carlos Parada, Altice Labs
carlos-f-parada@alticelabs.com

João Nogueira, Altice Labs

Bernardo Cardoso, Altice Labs
bernardo@alticelabs.com

In the last few years, the world has experienced an explosion of video content delivery over the Internet, using the so-called over-the-top model. More efficient video encoding, the wide spreading of broadband and the proliferation of Internet flat-rate plans are the main reasons for this evolution. Netflix is the major symbol of this model, with more than 100 million subscribers worldwide, and many others are willing to follow this example.

Keywords

OTT; Multicast; Software-Defined TV; CDN; Live Streaming; VOD; Timeshift TV

Introduction

Traditional TV services have substantially changed in recent years. Today, subscription TV services reach a large number of households, and IP-based delivery is gaining ground to traditional cable. Video-On-Demand (VOD) and timeshift capabilities are now widespread, allowing customers to get access to Digital Video Recorder (DVR) features and watch past TV programmes. More recently, global players like Netflix and others are operating a revolution, by providing TV contents (mostly on-demand) and leveraging their business in an OTT model strategy. Using the Internet and regional caching systems, they are able to reach customers in the global market, increasing dramatically their business at a very low cost. Simple delivery models combined with competitive subscription plans make services extremely attractive.

Today, each Altice operation runs its own TV solution, constraining synergies and precluding the group from taking advantage of scale. In a global environment, Altice faces strong competition from OTT players. In such a challenging environment, it is of vital importance to understand how a traditional telco can take advantage of OTT-like technologies to deliver TV/multimedia services, reducing costs and increasing profits. Furthermore, a large telecom group like Altice must take advantage of scale to find synergies, by unifying technologies and managing multiple operations in a convergent manner.

With this context in mind, and considering the fact that some Altice operations are experiencing exceptional transformations, it is important to understand the current trends of the TV and multimedia content delivery and define a unified strategy for the entire group. Recently, Altice Labs has been engaged in a study to elaborate on how we envision the TV/multimedia services of the future, considering two- and five-year timeframes, and proposing a strategy to lead this transformation.

To meet the requirements of future TV services, a new palette of emerging technologies needs to be addressed. Special remarks go to the cloud and software-defined technologies, which provide the flexibility and agility to efficiently run TV components, instantiating and scaling them, where and when required, adjusting the capacity to the current load. Other important topics are IP, unicast, open Application Programming Interfaces (API), micro-services, multi-access, or multi-operation.

This article starts by introducing the reader to the topic of TV/multimedia content. In the first section, it highlights the state-of-the-art and major future trends. Next, briefly identifies the relevant players in the delivery of TV/multimedia contents, comprising traditional telcos, technology vendors and OTT providers. In sequence, the proposed strategy is described, presenting the options layer by layer: presentation, control plane and media plane, for two- and five-year timeframes. Then, the article analyses pros and cons of using IP multicast or IP unicast to deliver live contents, using Portuguese and USA inspired scenarios. Finally, a brief conclusion is presented.

State-of-the-art and trends

In the last few years, the TV and multimedia sector has quickly evolved, changing radically the way people watch TV. New and richer user experiences, with higher quality, multiple screens, Internet-based, or cloud-based solutions, are some of the main trends.

Firstly, the **user interface** (UI) is changing and a minimalist look-and-feel is emerging, where content itself is the user interface. Big images and titles, large background areas and video overlay, are the preferred user interactions.

Multi-screen and companion devices are becoming increasingly popular, allowing users to watch TV on any screen (e.g. mobile, tablets, PCs). Users expect similar look-and-feel and the same set of features, no matter the screen they are using. Advanced features like moving the viewing from one device to another is a nice-to-have ability. The momentum for companion devices is also high, working as personal assistants for complementary actions like search or recommendations.

Users also expect ever increasing **video quality improvement**. 4K resolution is quickly becoming an industry standard, making it mandatory for top content providers to support this technology. 4K contents and devices are becoming available very quickly (e.g. PlayStation, Xbox One S, GoPro, iPhone). Soon, the 8K evolution will come to the market, as the technology is poking out.

In order to respond to these challenges, **TV platforms are already improving**, increasing performance, reducing costs and power consumption. Cloud solutions are becoming popular, supporting large libraries and massive streaming, even from long tail contents. Open TV platforms, like Reference Design Kit (RDK) or Android TV, are also rising, promising to change the vendor landscape in the near future.

New **business models** are emerging. Most of the traditional telcos started to provide OTT TV offers of their own (e.g. MEO Go, Verizon Go90) or reselling the service from OTT content providers (e.g. Vodafone/Netflix). Although OTT services deliver essentially VOD contents (e.g. Series, Movies), most of them are looking to introduce live streaming as the natural evolution (e.g. Apple, YouTube, Hulu). The Subscription VOD (SVOD) model with full access to all contents is becoming the rule.

In this context, **contents play a central role** in the ecosystem. Traditional TV providers are acquiring expensive rights for live sports. In the future, OTT players may strive for the same, although for example, Netflix has recently stated they will not make this move. The establishment of partnerships with big content providers becomes a strategy (e.g. Netflix with Disney), whereas in-house production is gaining momentum (e.g. Netflix, Altice Israel).

Legal changes can influence the way players act in the market, in particular in the European Union (EU) context. In February 2017, the EU approved the Portability of Services, meaning that subscribers from a member state must be able to access TV/multimedia services when travelling as if they were in their origin country (today, multimedia services cannot be watched abroad). This norm will take effect in the first quarter of 2018. Furthermore, the EU is pursuing the so-called “Open Audiovisual Market in Europe”, in order to break the Country Of Origin (COO) principle, allowing any provider from any member state to deliver TV/multimedia services to customers in other member states, overcoming the current legal barriers on content rights.

Relevant players

The TV/multimedia content delivery, traditionally managed by cable TV providers, involves today a set of players fighting for a piece on the global cake. The three major players involved in this market are traditional TV providers, OTT content providers and technology vendors.

TV providers

TV providers include the most traditional cable TV operators, who deliver TV channels using a Hybrid Fiber-Coaxial (HFC) infrastructure, and traditional telcos, delivering IPTV on copper (ADSL) or fibre technologies.

Most of the traditional TV providers are introducing mobile OTT operations, providing their customers with multi-screen experiences, either by developing their own solutions, powered by technology vendors, or in partnership. The service, look-and-feel and content (e.g. VOD, channels) and features (e.g. live, DVR, timeshift TV), try to be unified with the traditional residential TV, although in most cases, there are some gaps.

Some TV providers open the service to external customers, but the vast majority only provide the service for their residential customers as an add-on. In a few cases the service is free, supported by advertising, but in most of the cases, there is a monthly fee.

When telco operators provide TV/multimedia services, usually they do not charge for the traffic for the mobile bandwidth consumed, thus promoting the use of their own services against competitors, like Netflix.

Examples of traditional TV providers that created their own OTT service as a complement of their residential offer are Altice PT with MEO Go, SFR with SFR TV/PLAY, Swisscom with TV Air, Verizon with Go90.

Technology vendors

There are several traditional TV/multimedia technology vendors well positioned in the market, advertising very similar selling points. All of them announce the support of high video quality (4K), TV anytime and anywhere (timeshifted TV on any terminal), long-tail VOD contents, multi-device, uniform UI, etc.

From a technological perspective, those solutions propose a new paradigm, using the cloud as infrastructure and taking several advantages of that. Firstly, cloud infrastructure is cheaper, more agile and more efficient, as resources can be dynamically created, scaled or removed. It reduces the service creation time, as using Commercial Off-The-Shelf (COTS) hardware, there is no need for testing and certification for specific hardware. The cloud also leverages the Software as a Service (SaaS) model, and the use of micro-services increases the reusability and efficiency. The cloud helps telcos to focus on their business: the service. Overall, vendors claim for more operational efficiency, faster time to market, and increasing of agility and elasticity.

Examples of vendors well positioned to provide telcos with the technology above: Ericsson with Video Storage and Processing Platform (VSPP) and Cisco with Infinite Video Platform (IVP).

OTT content providers: the Netflix case study

There are several OTT content providers in this exponentially growing market. This section focuses on how the best provider of the class works: Netflix.

Netflix is a stunning case of success and the prominent leader of OTT multimedia content delivery. Today, Netflix is present in more than 190 countries (the biggest exception is China) and has over 100 million customers worldwide, around half of which in the USA, delivering millions of hours of video daily. In 2016, Netflix had revenues of USD 8.8 billion with a profit of USD 186 million.

Netflix provides flat-rate plans to its customers (€7.99, €9.99 and €11.99) and all customers have free access to all contents (SVOD). Multiple devices can be used to access contents, from STB, smart TV, PC, tablets, phones, to gaming consoles or sticks. Associated with VOD contents, Netflix provides other capabilities like search, offline view, ratings, reviews, recommendations, favourites, etc. In 2013, Netflix became a content producer, creating very successful exclusive contents.

Technology-wise, Netflix relies on two different technologies. The first is the Amazon Web Services (AWS) technology, supporting the presentation layer (UI), the Control Plane (CP) and the Full Content Library. The second is an in-house developed system called Netflix Open Connect, supporting content delivery (streaming) through Content Delivery Networks (CDN).

AWS makes use of the Elastic Compute Cloud (EC2), the compute part of the Amazon's offer. It is used to support the servers of the UI and CP, allowing the management of the Virtual Machines (VM) (instantiation, scaling, etc.). It controls the geographical location of instances and implements (anti-)affinity policies. The Amazon Simple Storage Service (S3), comprises the storage part, and it is used to store the full content library, by taking advantage of a Distributed File System (note: Netflix has developed the s3mper software to fix some S3 consistency problems). The Amazon's Service is spread across 16 regions worldwide and 44 availability zones (for affinity purposes).

Netflix Open Connect is a proprietary CDN system (like Youtube or Google have), based on specialized hardware made by Netflix: the Open Connect Appliances (OCA) – see **Figure 1**. Netflix decided for a proprietary solution for several reasons: (1) Netflix represents today a large amount of traffic for Internet Service Providers (ISP), so it is important to have a close relationship with them; (2) Netflix would be too big for Akamai (at the beginning Netflix used Akamai CDN, but then moved to Open connect); and, (3) Netflix claims that OCA are single-purpose and extremely efficient for its operation.

FIGURE 1 – Netflix OCA devices

The OCA hardware has several rack configurations. Those are mixed servers (compute + storage), with 2/4 10Gbps in a Link Aggregation Group (LAG), 14-280 TB storage, and are able to support around 10-20K simultaneous streams. OCA are powered by a FreeBSD system, using NGINX for load balancing in addition to HTTP service and BIRD for routing protocol implementation, namely, for Border Gateway Protocol (BGP).

OCA can be deployed in two different manners: embedded and Settlement-Free Interconnection (SFI) with a Private Network Interconnection (PNI) possibility. In the embedded case, the hardware is shipped to the ISP and installed there, doing BGP-peering in the telco premises. Netflix provides the OCA at no cost and helps with the configurations. The ISP provides space, power, cooling and infrastructure maintenance. In the SFI case, the OCA are deployed in Internet eXchanges (IX), where they can BGP-peer with local ISP there, while in the PNI case, OCA are deployed somewhere else and interconnected via a private link (PNI). The embedded case has advantages both for the ISP and Netflix, minimizing the Internet traffic, while improving the quality experienced by customers, respectively.

Netflix uses different Autonomous System Numbers (ASN) for BGP-peering, the ASN#40027 for embedded scenarios and the ASN#2906 for SFI/PNI scenarios. ISP announce the prefixes of the customers they want to be served by the OCA. This way, OCA know what customers they can serve and pass this information to the control plane. When a particular customer requests contents, this information will be used to select a particular OCA.

Netflix does the pre-positioning of the most popular contents (up to 40% storage capacity) during off-peak hours (fill window). This push strategy allows ISP to save bandwidth at peak-hours (delivery window), smoothing the Internet traffic profile. Fill windows are usually 12-hour long and can be shifted -+2 hours as desired by the ISP. OCA are simply content delivery boxes. All other aspects, like subscribers, Digital Rights Management (DRM), etc., are managed at the AWS.

Figure 2 depicts a basic flow diagram with the Netflix combined operation between the AWS (presentation/control plane) and the Open Connect (caching): (1) OCA periodically (5 min.) report available contents and routability information; (2) a customer presses the "Play" button; (3) AWS authenticates the customer and decides the content to send; (4) AWS selects an OCA to deliver the content based on availability/routability and generates the URL; (5) the customer player uses this URL to retrieve and play the content in the device.

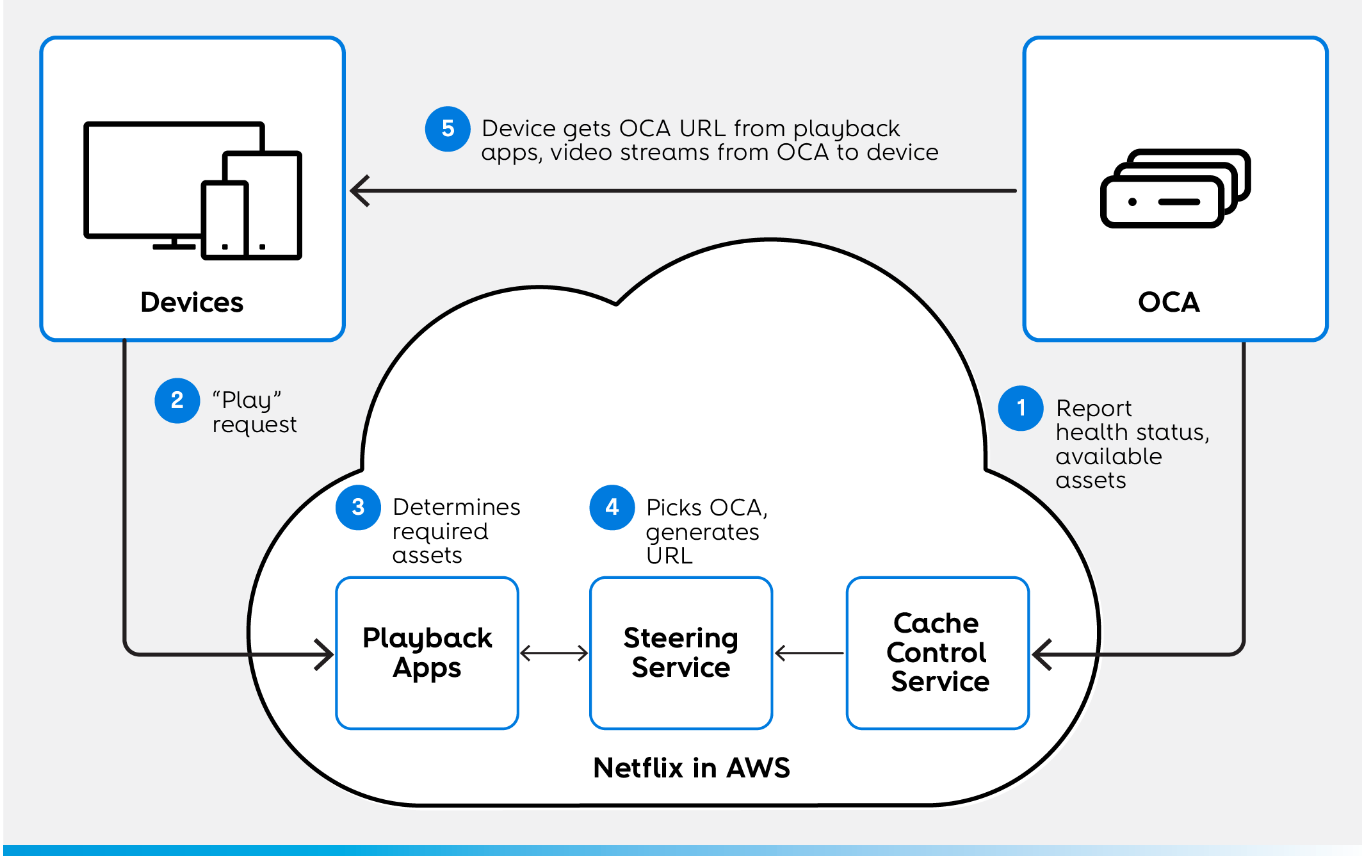


FIGURE 2 – Netflix delivery flow diagram

Proposed strategy

This section describes a proposed strategy from Altice Labs for the Altice Group, considering the different planes: control plane and data plane in two- and five-year timeframes.

Overview

The TV/multimedia of the future will evolve towards a multi-targeting paradigm. Solutions will be: (a) multi-device, supporting different devices and screens; (b) multi-DRM, supporting multiple DRM solutions for different platforms (e.g. Android, iOS); (c) multi-network, using a unified solution to deliver contents to multiple network types (e.g. mobile, Wi-Fi); and also (d) multi-operator, capable of supporting multiple operations on top of the same solution, an important asset for a telco group.

The multi-operator ability relies on the cloud technology, aiming to abstract resources and its location. It allows the service composition (chaining) by using distributed functions around different geographical locations in a simple and powerful way. The cloud principles also support the agility to run different small functions (micro-services) and scale them to adapt the resources to the load. The orchestration of the entire environment with all those functions, resources and location is what we call **software-defined TV** (see **Figure 3**).

Control plane

The CP includes a large set of features: DRM enveloping, stream control, recording management, recommendations, metadata management, etc. It is also responsible for managing subscribers, subscriptions, offers, reporting, charging, and analytics, among others.

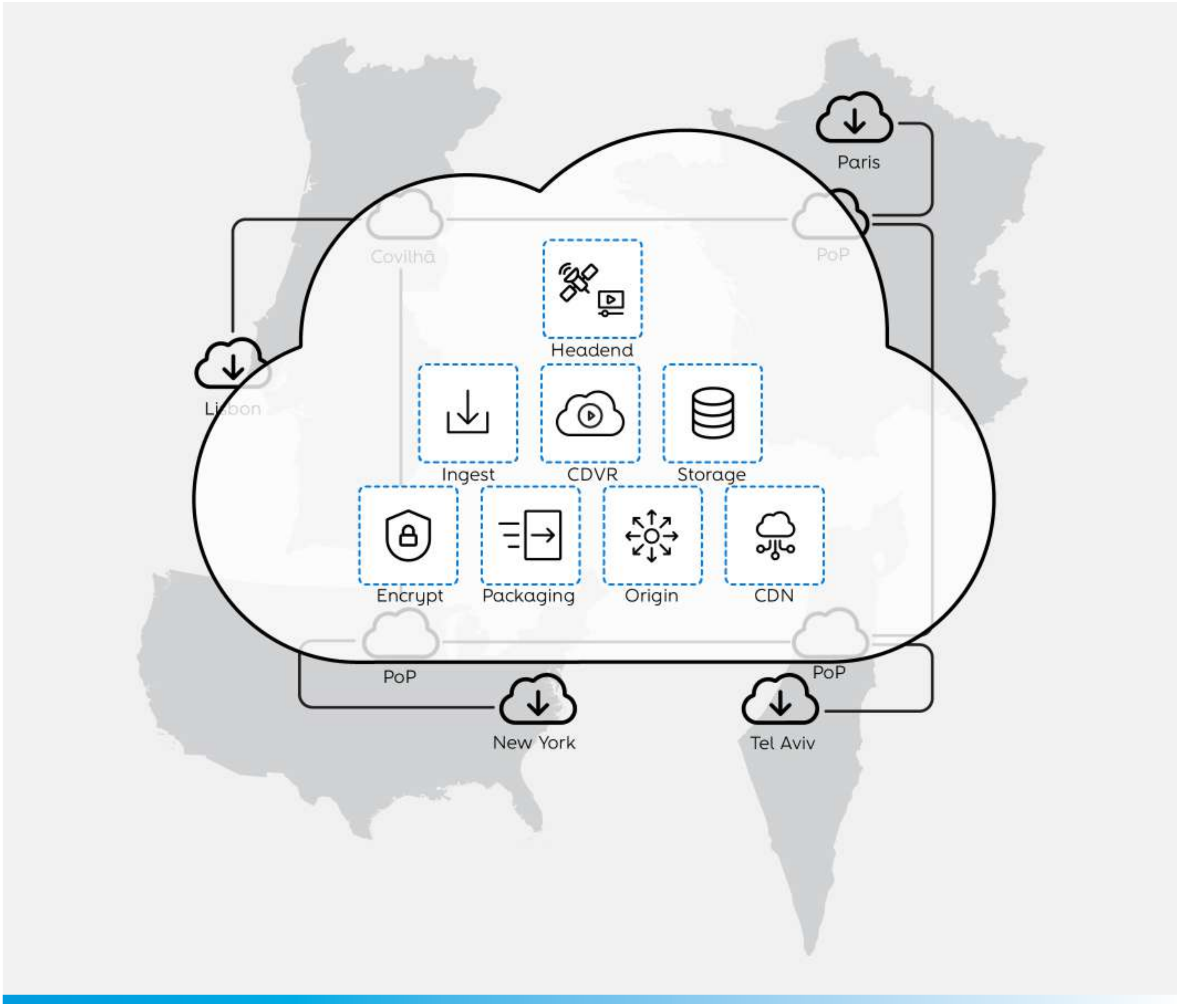


FIGURE 3 – Software-defined TV concept

Today

Today, CP features are highly dependent on the network, the devices and the Media Plane (MP), leading telcos to get locked-in to vendors. Each telco has a different set of systems and uses API to extend them with additional custom logic for its particular operation. This creates a high diversity of telcos' ecosystems, making it difficult for multi-operator groups like Altice to benefit from synergies and take advantage of scale.

Two-year timeframe

In a two-year timeframe, the Altice Group should move towards the control API unification, decoupling and isolating them from internal systems. That means not using the vendor API directly (e.g. mediaroom), rather creating a wrapper around them to perform the desired isolation and remove the dependency. In this sense, all Altice operations should agree on a common API (see **Figure 4**): the **Control-Plane Management Interface** (CPMI). By sharing API, reusability of software pieces will become a reality and synergies will rapidly start to emerge.

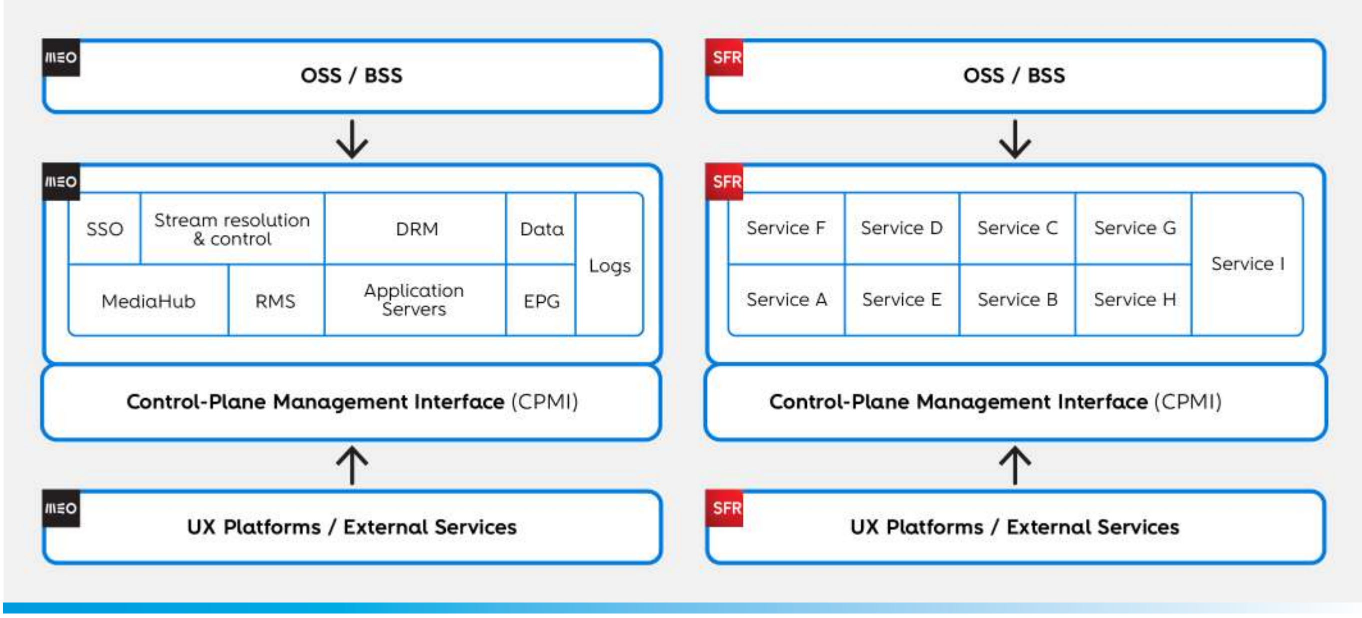


FIGURE 4 – Proposed strategy: Control Plane – two-year timeframe

Five-year timeframe

In a five-year timeframe, the Altice Group should start using the unified API, extending them in order to support multi-operator environments. Gradually, all applications and systems must start using this new API, migrating applications from the existing API in the different operators to the unified API.

Having a common CPMI, Altice can have a global

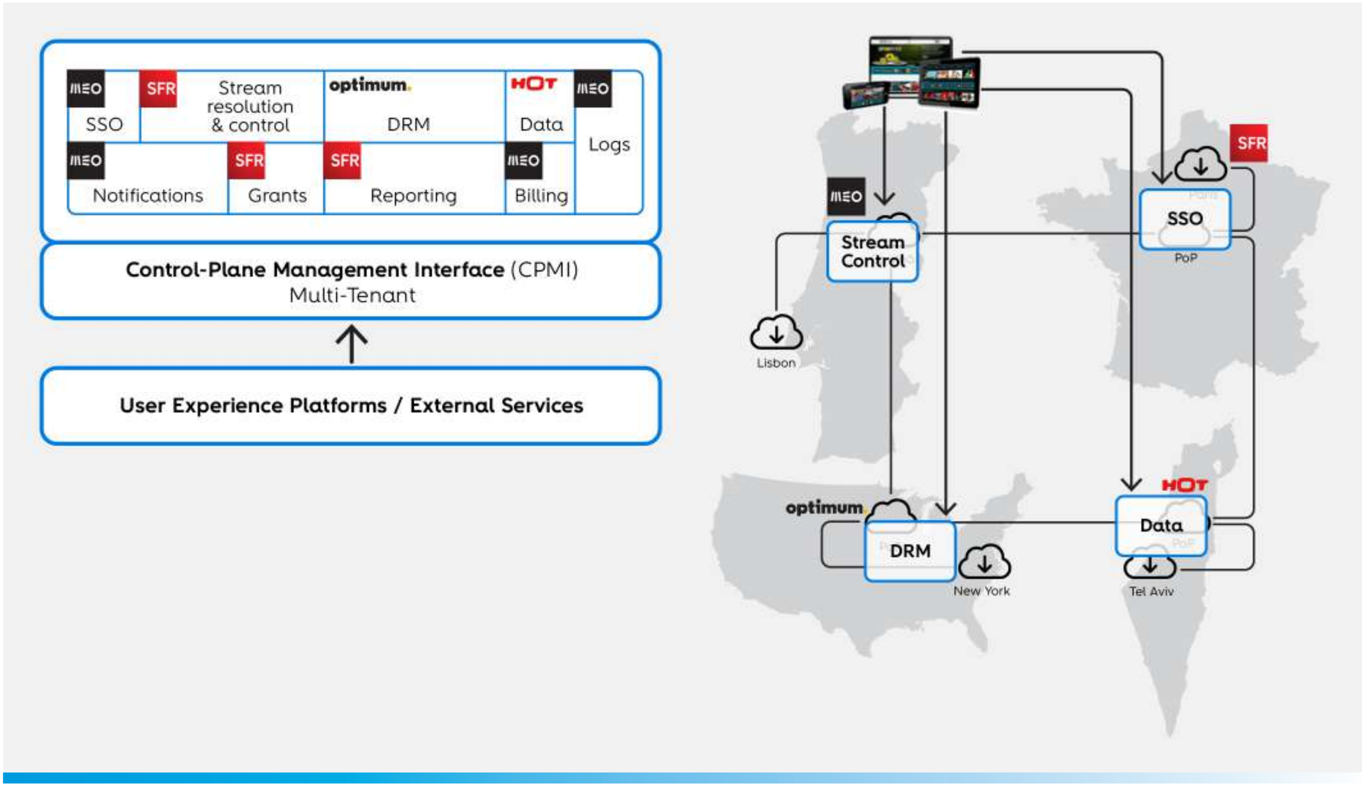


FIGURE 5 – Proposed strategy: Control Plane – five-year timeframe

view of the multiple operations, deploying different services in certain locations to serve multiple operators. For example, a Single-Sign-On (SSO) function can be deployed in location A (e.g. centralized authentication), and all Altice operators perform authentication at this point. The same can happen with DRM or other functions in locations A, B or C. The **Software-Defined TV** model is in charge of abstracting the functions location to the operators, acting as if they were together in the usual manner. **Figure 5** depicts this model.

Media plane

The MP includes a large set of functions (sequenced to acquire, prepare and deliver contents to customers): headend, ingest, storage, DRM, packaging, origin and CDN.

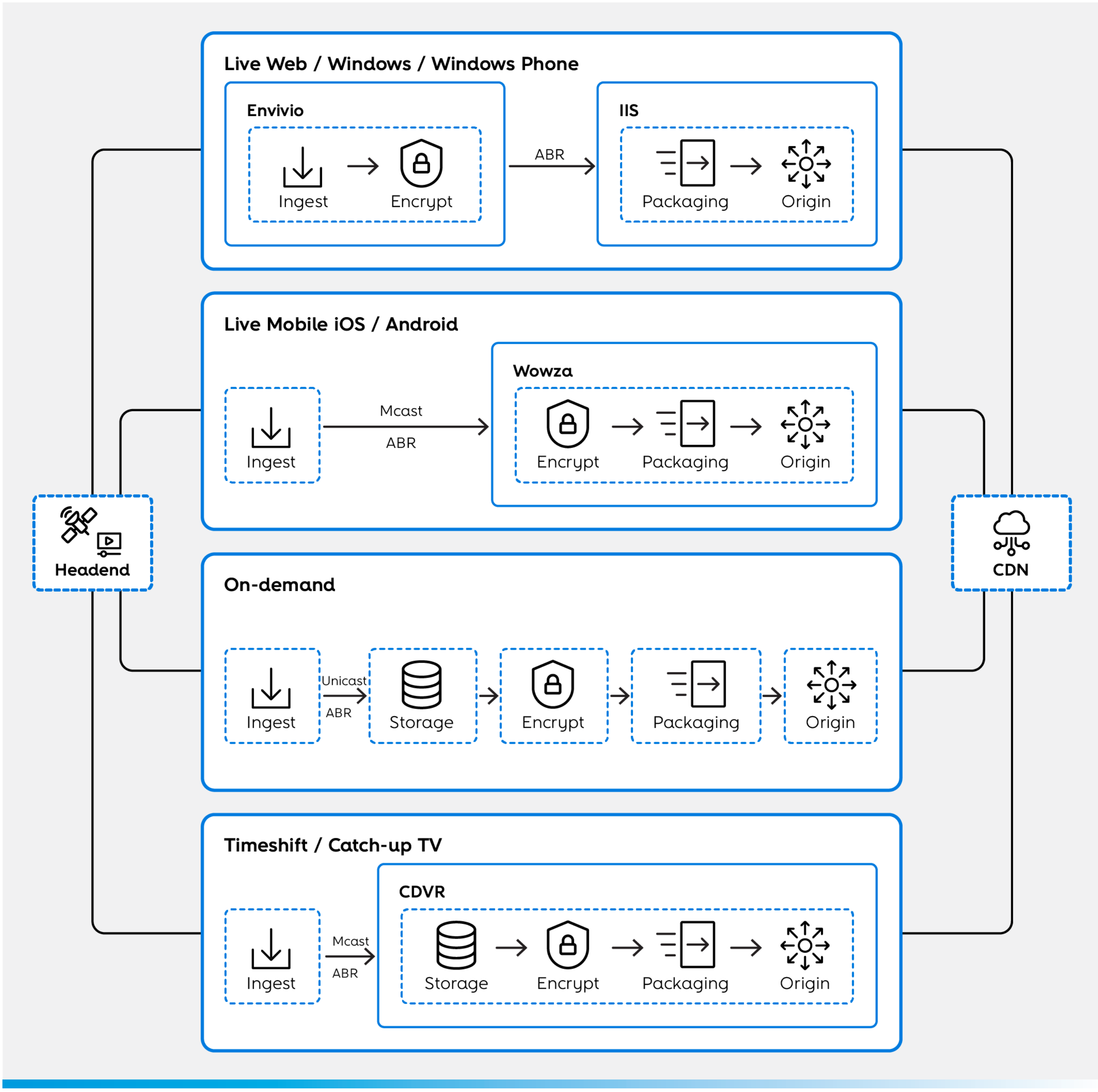


FIGURE 6 – Proposed strategy: Media Plane – today

Today

Today, telcos use multiple monolithic vendor solutions which comprise a mix of functions. The same contents are many times provided by separated systems (mobile, VOD, Android, iOS), leading to vertical silos, where the isolation of functions is difficult. This prevents the reuse of functions, making an inefficient use of resources, licensing, etc. **Figure 6** depicts the example of a typical operator ecosystem, where horizontal chains represent the functional silos used to deliver contents (headend to CDN) for different network/services/devices.

Two-year timeframe

In a two-year timeframe, the Altice Group should segregate and standardize functions, in order to easy reutilization of different services, for various networks and devices. To allow future function reutilization, it must be agreed on a common MP API among operators, the **Media-Plane Management Interface** (MPMI). As a result, the media plane ecosystem of the **Figure 6** could be simplified to the example depicted in **Figure 7**.

Five-year timeframe

In a five-year timeframe, the Altice Group should start using the standard functions, as building blocks, to create TV/multimedia services. Using these functional pieces, a manager can build the desired service, by linking different functions (chain) that can be abstracted and deployed in different locations of the Altice Group. The **Software-Defined TV** model is in charge of abstracting functions location to the operators, acting as if they were together. **Figure 8** depicts this model.

Live multicast vs unicast

This section describes the results obtained in a study regarding the comparison between the delivery of live contents either using traditional IP multicast or unicast (OTT like approach).

Motivation

This chapter attempts to respond to the questions: Is IP multicast still required? Can OTT (unicast) support a telco IPTV delivery? Those are important questions to answer, especially in the current context, where some Altice operations are moving from traditional cable to fibre

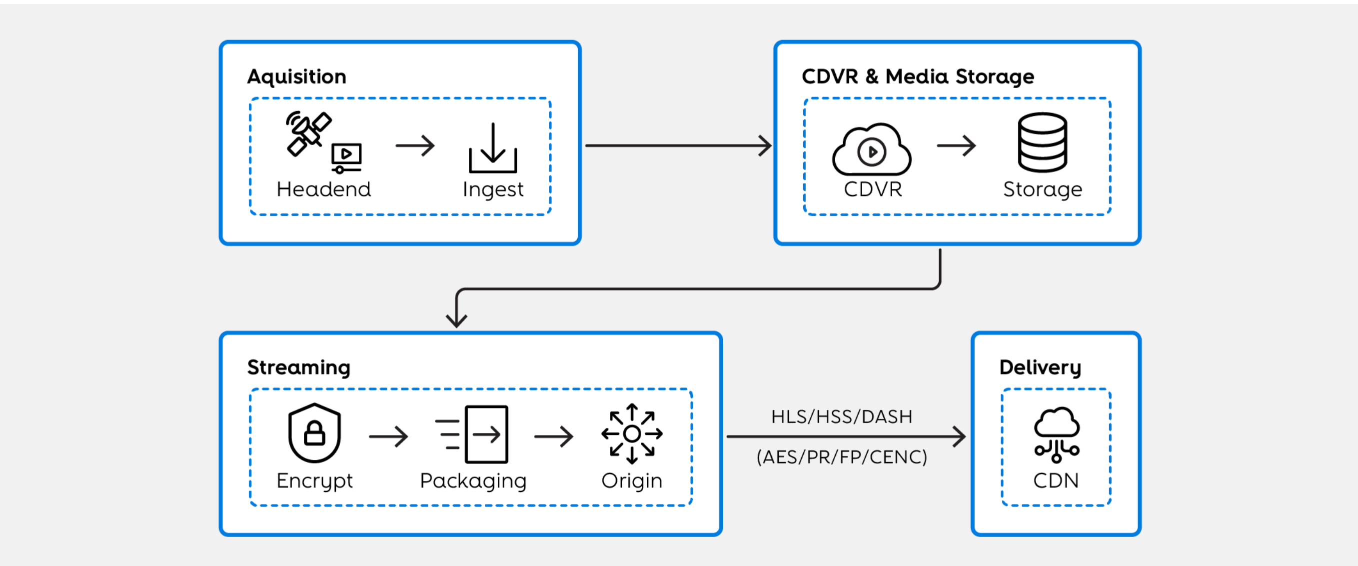


FIGURE 7 – Proposed strategy: Media Plane – two-year timeframe

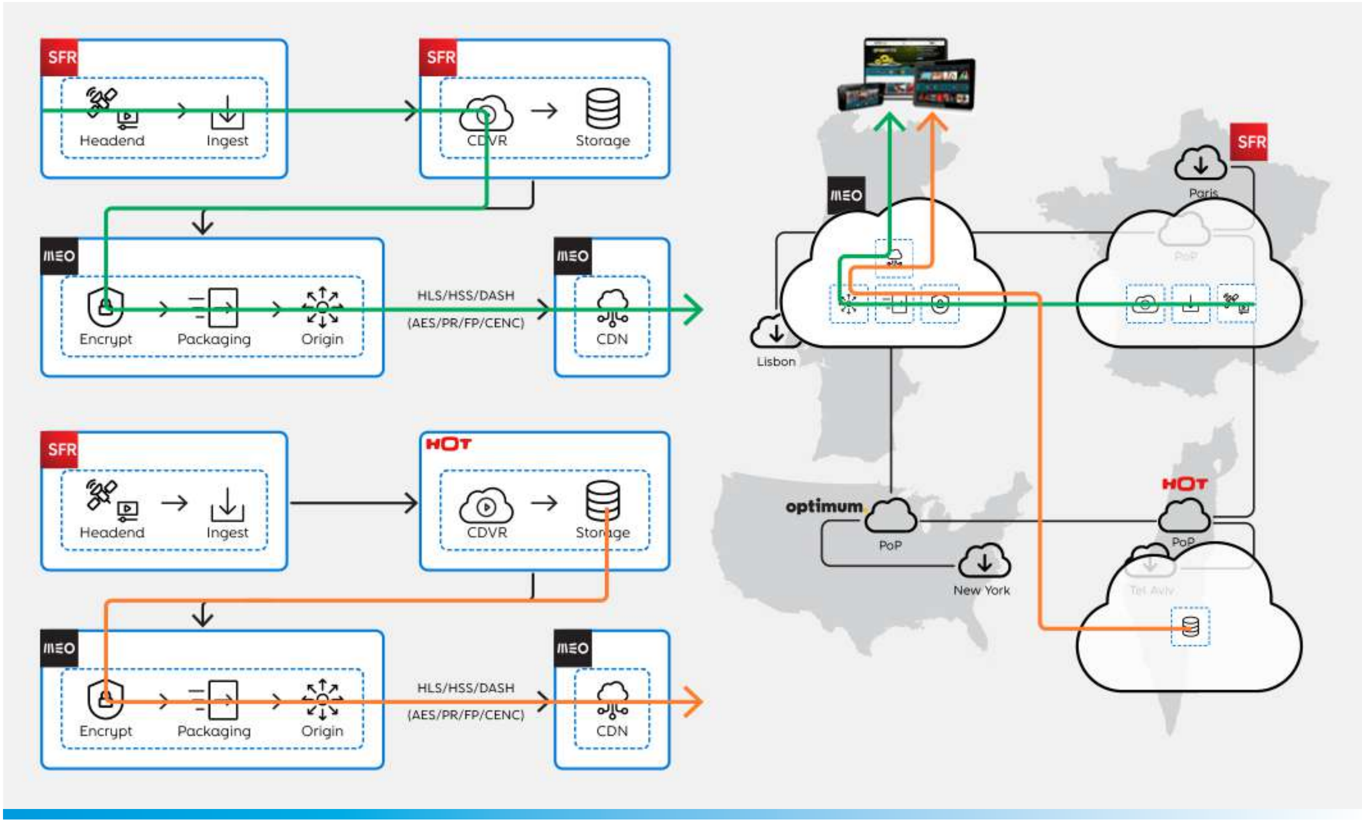


FIGURE 8 – Proposed strategy: Media Plane – five-year timeframe

technologies, and need to decide whether they build a multicast network or use only unicast.

The success of OTT players, like Netflix, delivering contents using OTT technologies, raises doubts on traditional telcos about whether they can do the same (although OTT players mainly deliver VOD content, which is much less challenging than live).

Today, telcos already deliver some contents using unicast. Actually, they use a technology mix: multicast for live and unicast for VOD, timeshift, Fast Channel Change (FCC) and mobile. The unification to unicast would simplify service delivery, reducing overall costs. The gradual reduction of live viewing and the dramatic growing of timeshift and VOD services increases the feeling that multicast is not that important anymore.

Pros and cons

The main multicast advantage is the ability to deliver live contents to millions of customers using a single stream. By definition, multicast creates a delivery tree, forwarding the traffic and forking flows only when required, unlike unicast which requires one flow per customer. This dramatically reduces the amount of traffic crossing the network, smoothing the traffic profile and avoiding peaks at top viewing hours.

However, multicast is not enough to provide all services (e.g. timeshift, VOD, FCC). In the timeshift case, unicast and multicast must be synchronized when the customers push back to the past, increasing the complexity. Furthermore, multicast does not work well in Wi-Fi and requires specific solutions in mobile environments, like Multimedia Broadcast/Multicast Service (MBMS).

The biggest unicast advantage is simplicity. It can be used to provide all services, from live, VOD or timeshift, and can be used in any kind of IP network, from mobile to Wi-Fi or fibre.

The biggest unicast disadvantage is the inefficiency to deliver live contents, which makes mandatory the use of distributed caching systems (CDN). Even though, during peak hours, the traffic will increase dramatically. But the big challenge for unicast, is the support of big events (e.g. the Final FIFA World Cup), when millions of customers change channel almost simultaneously (e.g. at the beginning, half break, etc.), pushing the control plane to the limit.

Reference scenarios

First of all it is important to note that for the sake of simplicity of the simulation, a flat number of clients and TV channels were considered in order to isolate and highlight the potential impact of the traffic mix evolution (the number of clients over multicast and the increase in high quality content streams). However, using the developed simulation tool, all parameters may be changed to reflect the market evolution.

	Portuguese Inspired *						Future USA **					
Parameters	**2017**			**2021**			**2018**			**2021**		
# Clients	1 200 000			1 200 000			1 800 000			1 800 000		
# Set-Top-Boxes	1 600 000			1 600 000			5 400 000			5 400 000		
Peak # Streams	900 000			900 000			1 800 000			1 800 000		
# TV Channels	195			195			500			500		
Channels Bitrate (SD/HD/4K)	2	6	12	2	6	12	2	6	12	2	6	12
Headend Channels Mix (SD/HD/4K)	72%	72%	72%	40%	50%	10%	60%	40%	0%	30%	50%	20%
Channels Bitrate (SD/HD/4K)	80%	72%	72%	55%	40%	5%	70%	30%	0%	25%	60%	15%
Peak # Different TV Channels	161			161			250			250		
Clients Avg. Peak Other IPTV Unicast (Mbps)	0.375			0,750			0,375			0,750		
Clients Avg. Peak Other Internet (Mbps)	0,416			0,832			0,416			0,832		
# Clients per OLT	2 000			2 000			2 000			2 000		
# OLT per Service Router	10			10			10			10		
Computed Parameters												
Headend Multicast Traffic (Mbps)	*610*			*975*			*1 800*			*3 000*		
OLT Uplink Multicast Traffic (Mbps)	*451*			*660*			*800*			*1 475*		
Total # Service Routers	*60*			*60*			*90*			*90*		
Total # OLT	*600*			*600*			*900*			*900*		

* Estimated numbers
** End of 2018 projections

Note: a flat number of clients and TV channels were considered for OTT traffic calculation simplicity.

TABLE 1 – Reference scenario numbers and assumptions

Considering what was said before, this study intends to evaluate how a unicast solution would behave for particular reference scenarios. The purpose is to understand whether it would be possible for the current networks to support an OTT operation, calculating how much the traffic will increase in multiple network sections. The two reference scenarios considered are based on the Portuguese operation (Altice PT) and the USA projections for the end of 2018. Forecasts for a five-year timeframe for each scenario were also considered. The **Table 1** depicts those values and the assumptions considered for the simulation.

Today

Portuguese case inspired

Figure 9 depicts the reference network aggregation.

Clients: Nothing changes since the traffic is basically the same as using multicast. Live IPTV traffic is around 2.1Mbit/s and total (including Internet/VOD/timeshift) around 2.9Mbit/s.

OLT: Significant changes in uplink traffic (downlink nothing changes). Live TV uplink traffic increases from 450Mbit/s to 4.2Gbit/s and the total traffic increases to 5.8Gbit/s. Using 10 Gigabit Ethernet (GbE) interfaces, it would be feasible with current equipment.

Service Routers (SR): Significant changes, in uplink and downlink. Live TV uplink traffic increases from 610Mbit/s to 42Gbit/s and downlink from 4.5Gbit/s to 42Gbit/s. Total traffic raises from 24Gbit/s to 80Gbit/s. Using 100GbE interfaces it is feasible, but not all SR may have this capacity.

This would be a good location to deploy CDN. In this case, the downlink traffic would be the same but the uplink traffic will reduce to a similar value as for the multicast case.

Headend: If CDN are not deployed at SR locations, this section of the network would dramatically increase traffic, raising from 610Mbit/s to 2.5Tbit/s. Definitively, this is not possible, meaning that CDN must be deployed at the SR. In this case (on ideal caching), the traffic would reduce to 37Gbit/s, which is reasonable.

Summary: This simulation exercise shows that an OTT (unicast) operation is really possible using current hardware on the network without changes. The CDN deployment is mandatory and SRs are the perfect locations for that.

USA case inspired

Figure 10 depicts the reference network aggregation. The main differences to the Portuguese inspired case are the number of STB per customer and the SD/HD mix.

Clients: Nothing changes, since the traffic is basically the same as using multicast. Live IPTV traffic is around 3.2Mbit/s and total (including Internet/VOD/timeshift) around 4.0Mbit/s.

OLT: Significant changes in uplink traffic (at downlink nothing changes). Live TV uplink traffic increases from 800Mbit/s to 6.4Gbit/s and the total traffic increases to 8.0Gbit/s. Using 10GbE interfaces, it would be too close to the interface limits. Solutions can be the plug-in of additional 10GbE cards or the reduction of the number of customers per OLT.

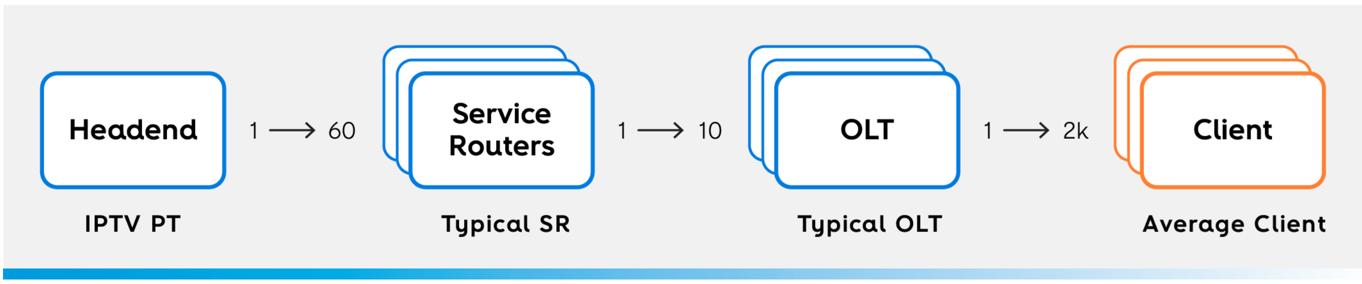


FIGURE 9 – Portuguese inspired network aggregation (estimated numbers)

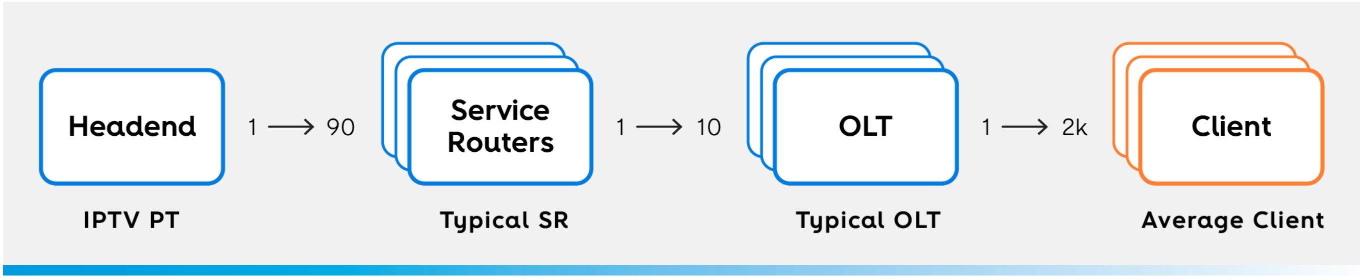


FIGURE 10 – Future USA network aggregation (estimated numbers)

SR: Significant changes, in uplink and downlink. Live TV uplink traffic increases from 1.8Gbit/s to 64Gbit/s and downlink from 8.0Gbit/s to 68Gbit/s. Total traffic raises from 24Gbit/s to 80Gbit/s. Using 100GbE interfaces it would be too close to the interface limits. This is a good location to deploy CDN; in this case, the uplink traffic would reduce to a similar value as for the multicast case.

Headend: If CDN are not deployed at SR locations, the headend would dramatically increase traffic, raising from 1.8Gbit/s in the multicast case to 5.8Tbit/s. Definitively, this is not possible, meaning that CDN must be deployed at SR. In this case (on ideal caching), the traffic would reduce to 162Gbit/s, which is reasonable, but may require additional interfaces.

Summary: This simulation exercise shows that an OTT (unicast) operation is possible, but may require additional interfaces at several points of the network (e.g. OLT, SR, headend). The introduction of CDN is mandatory and the SR are the perfect locations.

Forecasts 2021

After the analysis of the current numbers for the Portuguese and USA inspired scenarios, it is important to understand how this would evolve in a five-year timeframe, where it is expected more Internet traffic, more timeshifted services and higher quality contents (e.g. 4K). Thus, the following assumptions were considered (already shown in the reference scenario in **Table 1**).

Portuguese inspired scenario:

- **Traffic Mix:** Increase from 80% SD / 20% HD / 0% 4K to 55% SD / 40% HD / 5% 4K
- **Internet + Timeshift + VOD:** Increase of 100%

USA inspired scenario:

- **Traffic Mix:** Increase from 70% SD / 30% HD / 0% 4K to 25% SD / 60% HD / 15% 4K
- **Internet + Timeshift + VOD:** Increase of 100%

Portuguese case inspired

Clients: Nothing changes since the traffic is basically the same as using multicast. Total traffic will increase from 2.9Mbit/s in 2017 to 4.7Mbit/s 2021.

OLT: Significant changes in uplink traffic (at downlink nothing changes). Live TV uplink traffic increases from 660Mbit/s to 6.2Gbit/s, and the total traffic (including Internet/VOD/timeshift) increases from 3.8Gbit/s to 9Gbit/s. Using 10GbE interfaces, it would be too close to the limits, requiring additional 10GbE interfaces or the reduction in the number of customers per OLT.

SR: Significant changes, in uplink and downlink. Live TV uplink traffic increases from 975Mbit/s to 61.5Gbit/s and downlink from 6.6Gbit/s to 61.5Gbit/s. Total traffic raises to 93Gbit/s both for uplink/downlink. Using 100GbE interfaces it would be too close the interface limits. This is a good location to deploy CDN; in this case, the downlink traffic would be the same, but the uplink traffic will reduce to a similar value as for the multicast case.

Headend: If CDN are not deployed at SR locations, this section of the network would dramatically increase traffic, raising from 1.8Gbit/s in the multicast case to 3.7Tbit/s. Definitively, this is not possible, meaning that CDN must be deployed at SR. In this case (on ideal caching), the traffic would reduce to 59Gbit/s, which is more reasonable, but anyway may require the addition of interfaces.

Summary: This simulation exercise shows that by 2021, the OTT based operation is possible, but would require additional interfaces at several points of the network (e.g. OLT, SR, headend). The introduction of CDN is mandatory and the SR are the perfect locations.

USA case inspired

Clients: Nothing changes since the traffic is basically the same as using multicast. Total traffic will increase from 4.0Mbit/s in 2017 to 7.5Mbit/s 2021.

OLT: Significant changes in uplink traffic (at downlink nothing changes). Live TV uplink traffic increases from 1.5Gbit/s to 11.8Gbit/s, and the total traffic (including Internet/VOD/timeshift) increases from 4.6Gbit/s to 15Gbit/s. Using one 10GbE interface is not enough, and other interfaces need to be added. The reduction of the number of customers per OLT is also possible, but probably more difficult to do.

SR: Significant changes, in uplink and downlink. Live TV uplink traffic increases from 3Gbit/s to 118Gbit/s and downlink traffic from 15Gbit/s to 118Gbit/s. Total traffic raises to 93Gbit/s both for uplink/downlink. Using 100GbE interface would be too close the interface limits. However, this is a good location to deploy CDN; in this case, the downlink traffic would be the same (118Gbit/s), but the uplink traffic would reduce to a value similar as for the multicast case.

Headend: If CDN are not deployed at SR locations, this section of the network would dramatically increase traffic, raising from 3Gbit/s in the multicast case to 11Tbit/s. Definitively, this is not possible, meaning that caching systems must be deployed at SR. In this case (on ideal caching), the traffic would reduce to 270Gbit/s, which is more reasonable, but anyway may require the addition of interfaces.

Summary: This simulation exercise shows that by 2021, the OTT based operation is possible, but would require additional interfaces at several points of the network (e.g. OLT, SR, headend). The introduction of CDN is mandatory and the SR are the perfect locations.

Final remarks

Overall, these simulations show that the OTT mode of operation is viable, but may require upgrades to the current networks. Although today only minor upgrades are required, in a five-year timeframe the level of upgrades should increase.

In most of the cases, the addition of new interfaces (e.g. 10GbE, 100GbE) is enough. However, in other cases, it may need heavier technology replacement (e.g. 10GbE interfaces by 100GbE). This may force upgrades in the equipment itself (OLT, SR) in order to support such interfaces or to cope with the respective increase in traffic.

The utilization of CDN is mandatory to distribute the content delivery load along the network. Without this, headends will be flooded with Tbit/s of traffic. The best location for the deployment of CDN units is co-located with the SR.

Conclusions

This article presented the results of a study developed by Altice Labs regarding the proposal of a TV/multimedia content delivery strategy for the entire Altice Group, considering two- and five-year timeframes.

The proposed strategy was divided into control and media planes. For both, the focus was put on the utilization of cloud and software-defined technologies in order to leverage the synergies among the different operations from the Altice Group and benefit from scale.

In the control plane, the unification of API and interfaces is the first step to take, allowing the reutilization of control functions by different operators. This unification allows the centralization of services in different locations (e.g. SSO, EPG) to serve all Altice operators in a transparent manner, the so-called Software-Defined TV.

In the media plane, the isolation of standard functions is the priority, making possible the reutilization of functions by multiple services, devices and networks. This allows the Altice Group to use common functions (deployed in certain locations) to serve multiple operations, allowing managers to create flows of function (chains) to also build the service on top of Software-Defined TV.

Finally, a study was conducted to understand whether multicast is still a mandatory enabler for live content distribution, or OTT (unicast) technologies can now replace the traditional model. As a result, the OTT approach seems possible; however, the current network in place may not support the resulting traffic as it is, rather requiring some upgrades, which could go from the simple addition of interfaces to the replacement of equipment. The deployment of CDN is mandatory, ideally co-located with SR.


PAGE

13

From TV Flow to TV Concierge: a new TV experience

Bernardo Cardoso, Altice Labs
bernardo@alticelabs.com

Jorge Ferraz de Abreu, University of Aveiro
jfa@ua.pt

With more and more content available and pay TV platforms UIs getting aesthetically closer to each other, the user is left immersed in an ocean of content and with undifferentiated strategies to deal with it. Recommendation systems have been proposed as a possible tool to lessen these issues when applied to VOD, but with limited success in linear TV. This article proposes two different new TV watching experiences based on recent enhancements in recommendation algorithms, both presenting the case for personalization as a key concept for future TV platforms.

Keywords

TV; UX; Personalization; Recommendations

Introduction

Pay TV offerings are getting more and more alike. Initially, similarities begun with the provided services, nowadays far beyond live TV, with most of the operators delivering pay-per-view (PPV), video on demand (VOD), subscription video on demand (SVOD), catch-up TV, interactive applications and an assortment of tie-ins with mobile and PC OTT services [1]. Recently, there is likewise a tendency for the user interfaces (UI) and the user experiences (UX) to also portrait very similar aesthetics. For instance, the freshly released YouView UI from TalkTalk, a UK operator, has an uncanny resemblance with the current MEO interface, which in the past had some inspiration from Comcast's X1 UI. However, this trend for similar UIs is not limited to operators: some of the biggest interactive TV (iTV) solutions suppliers have propositions that are not that different, for instance, the new Ericsson MediaFirst UI also resemblances MEO UI and to a lesser extent even Cisco's Infinite Video premium UI share the same concepts in its core menu – **Figure 1** illustrates these similarities.

FIGURE 1 – UIs similarities of different operators and vendors

Aiming to design a user experience that goes in a different direction, less glued to current looks and services and more focused in a completely personalized experience, a consortium between Altice Labs, University of Aveiro and Instituto de Telecomunicações was created in order to materialize, under the scope of the Portugal 2020 program [2], the UltraTV project where we plan to achieve this new UX approach by, in one hand, seeking inspiration in the trending concepts in social media and, on the other hand, relying on machine learning and analytics to create an experience in line with modern OTT propositions that savvy users expect.

The UltraTV project has a broad range of uses cases, including aspects of low-level OTT streaming and development technologies but, in this article, we will present only two use cases that are closely related to each other and directly concern the personalization of the TV user experience.

TV Flow

Being one of the objectives of the UltraTV project to propose user experiences that go above and beyond the current offers, the consortium first started to understand what was available in the market. The first step in order to do so was a broad survey of the current UIs and the diversified features proposed from both traditional operators and OTT players. This phase allowed to get a baseline and to understand some patterns already in the field. From that point on, the team started to brainstorm around the concepts of crossing the basics of a TV service with the kind of functionalities users enjoy today in social media, something that may be achieved by creating a kind of news feed similar to the one from Facebook, but with a mix of different TV and OTT sources, creating a continuous flow of cards. These cards will have information about a content item and will allow an easy navigation to full-screen consumption: some cards will offer linear channels, some catch-up programs, some will have clips proposed by the viewer's friends, others will present Netflix recommendations, some will be about targeted advertising and so on. An algorithm will automatically generate this feed, with a strong focus on personalization to keep the user engaged with the service and the contents paving the way for this UltraTV main use case, the TV Flow concept.

The name TV Flow and some of the associated concepts got an inspiration from Mihaly Csikszentmihalyi's book "Flow: The Psychology of Optimal Experience" [3] where he encourages individuals to get into a state of "flow" as a way to achieve happiness. Some of the attributes for that state are also shared with our proposition, for instance, while the viewer goes around in his TV consumption journey we want him to be deep and effortless involved in his activity, by removing the need to hunt for content to watch. We also added some nice touches like the video clips starting to play automatically when the correspondent item is selected. We want the user to have immediate feedback, and in this respect, the UI will be fully responsive and always assure the user is on the right track, giving him the sense of control over its own actions. Ideally, we expect that our viewer will experience an alteration of time, that "hours felt like minutes" while using our solution.

To achieve so, the interaction model designed by the team repurposes a staple from TV watching, the channel surfing. The viewer will zap between items on the feed like he does today between channels, but instead of changing channels the zapping will change to the next/previous item in the feed. In terms of UI, the user will still be able to get to his channel list or access his usual list of catch-up programs alongside the actual Facebook or Youtube feed, whereas in the centre there will be a personalized constant flow of content, as shown in **Figure 2**.

Nevertheless, a complex intertwine of technologies must come together for this vision to materialize, starting with the needed analytics and machine learning algorithms to select the appropriate linear TV programs at the correct time that will be a key ingredient for the TV Flow mix.

Some particularities of linear content add greater complexity in creating an effective personalization system, namely the catalogue in constant change and the short time the TV content is available [4]. A recommendation system that only has access to the programming of the linear channels, can only recommend, at any certain moment, the programs that these channels are broadcasting or programs that will start in the upcoming minutes. Even systems with access to a catch-up TV catalogue need to deal with the fact that fresh content is entering the collection all the time, since the system is constantly recording new programs and, similarly, removing older ones.

Another characteristic of linear TV consumption is that it normally follows a very regular pattern [4]. Contrary to a VOD system, where the viewer usually wants to find a new movie to watch, when watching TV the consumer has habits associated with certain times of the day and follows specific recurring programs on a small number of available channels [5]. This regularity together with other contextual aspects of TV watching was used as the basis for proposing new approaches in recommending linear television content [6] [7] [8].

Consumption patterns on TV are strongly conditioned by time context and channel preferences [4]. In this sense, one way to go beyond the state of the art on the current recommendation systems for linear content is to explore and integrate the attributes "time of the day" and "day of the week" in the user modelling. A significant improvement in the quality of the

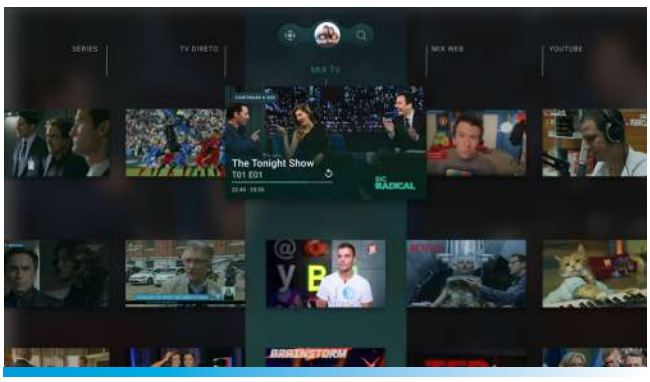


FIGURE 2 – Prototype of the main TV Flow screen

recommendations will be achieved when this time context is taken into account. The quality of the suggestions can also be improved further with the addition of implicit feedback, taken from consumer data analysis, through analytics processing, and taking into account not only the linear TV programs but also the available catalogue of catch-up TV. Actually, in the context of UltraTV, this approach has been shown to improve accuracy while maintaining good levels of diversity and serendipity [7]. In addition to the temporal context, we can also use a concept of sequential context [6], which takes into account the last viewed program at the time of the recommendation to influence the next program to propose in a viewing session. Once again, UltraTV is proving that this concept has been shown to provide considerable precision gains to the personalization.

Another important property of TV that must be highlighted and that the UltraTV project addresses is that different users, whose tastes can vary widely, often share its usage. Typical recommendation systems do not handle this situation very well since visualization data is typically collected and modelled at the device level, aggregating all users and blurring their individual tastes. To help on that, the UltraTV project will focus on multiple ways to address user profiles. In the TV Flow proposition the user profile will be used not only for the personalization and recommendation system but also to link the personalized feed with the OTT content, so that the TV Flow can get personal recommendations from Youtube, Netflix, Facebook, etc., to add to the mix (a blend of different contents). This means that, when there is a profile change in TV Flow, the whole feed changes by not only getting new linear content recommendations but also automatically changing the linked OTT accounts and receiving new items from those systems.

This linkage of OTT accounts with user profiles in TV Flow will allow for users to recommend content to each other content, using the already established network of friends provided by those platforms. For example, if a TV Flow user "likes" a program, it can appear on his friends feed on Facebook. If some of his friends are also TV Flow users, they will also get that recommendation on their TV Flow mix column on TV. A more direct recommendation could also be done in a peer-to-peer way but leveraging this recommendation through "liking" will provide a more organic integration with the current "flow" of the viewers.

For the user profile change to be really easy for the viewer, it was placed at a position where it's really accessible by just pressing the up key on the remote and then selecting the correct user from the menu (shaped as a peaberry) – see **Figure 3**.

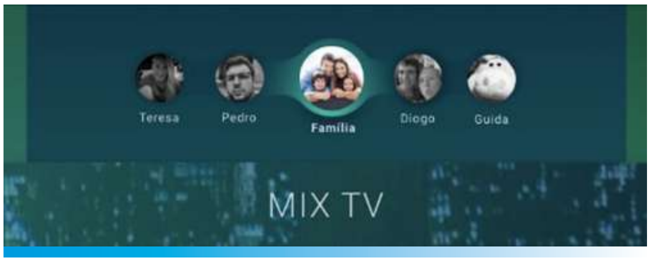


FIGURE 3 – User profile switching interface

Yet the manual profile switching is not practical for all users. Sometimes the viewer forgets to change to the correct profile, other times it could be a child. For these and other situations, in the UltraTV project it is intended to develop an automatic profile change mechanism. To implement it the team plans to resort to a wide range of inference algorithms, for instance detecting the user's smartphone presence near the STB through bluetooth sensing; or real-time evaluation of the programs being watched and recognizing that the viewing pattern is not a good match for the currently selected user profile and proposing a user change to the current viewer.

All this personalization may have a big impact on the user expectation if it assumes that is dealing with a normal pay TV interface. For instance, the Linear TV column in TV Flow will be ordered by channel usage, instead of the traditional channel ordering provided by the operator. That way the channel alignment can change from time to time. Some of our customers will not be very fond of that and will also want a more traditional way to interact with the platform. For that, we also foresee a simple way to change the view to a more conventional and non-personalized content presentation. The UI will still be based on columns like TV Flow, but instead of a personalized list, the user will get the fixed sorting that the operator already proposes with a standard channel alignment, recordings, VOD, etc.

The following section presents the approach that can be leveraged on this more straightforward UI.

TV Concierge

In the last decades, the quantity and quality of the content available in a typical living room have increased considerably. All this growth on content and functionality can add up really quick, resulting in a much wider choice for the end users. However, this abundance of content has some drawbacks, for example, the user may get lost while seeking for a suitable program to watch. In "The Paradox of Choice: Why More Is Less" [9], the psychologist Barry Schwartz explains that a lot of choices can be actually detrimental to people's psychological and emotional well-being. He explains how a culture that grows with an infinite availability of constantly evolving options can also foster deep dissatisfaction, which can lead to paralysis in decision making and, in some cases, depression.

Along with this, there is also a progressive complexity on the interaction modes. The amount of content is so great that the viewer has difficulty in selecting a program to watch since the tools that could help him in this task are of limited practical use. For instance, remote controls did not evolve to an adequate set of features and, in most cases, the STB do not have enough capabilities to cope with all the new interactivity and content.

Recommendation systems have been presented as a way to address both these issues, by promoting content more aligned with the tastes and expectations of the user and by being readily available, making life easier for the user. However, traditional recommendation and personalization systems are often very passive. Usually, when the user is in a screen with program or VOD information he also gets some sort of recommendation, typically with a phrase like "Viewers who watched this movie also watched these other titles…", but it's up to the user to act upon the recommendation and sometimes he doesn't even notice it.

To ease some of the anxiety in the user and at the same time minimize the burden of having to navigate complex interfaces, just to get to the content he wants to watch, we devised a kind of butler for the TV consumption that we aptly named TV Concierge. It will use the same personalization algorithms proposed in TV Flow, that will take into account diversified contextual aspects of TV consumption more focused on linear content problems.

Our design took an approach that tries to minimize the number of interactions the user needs to perform. Simultaneously we aim to bring the recommendations to the upfront of the platform – in this respect, the system will not just wait for the user to ask for some recommendation but instead will suggest them beforehand. This will start from the very beginning, i.e., when the user turns on his STB he will be presented with the program that makes sense to play at the time the box is being turned on, rather than with the last channel tuned in the previous night. In addition, making use of the time-shifting capabilities of modern TV platforms, the suggested program will start from the beginning. For instance, if a potential viewer, who usually watches the TV newscast when he gets home about 8:00 pm, turns on the STB around this time, the recommendation system will automatically start playing (from the beginning) the newscast that the viewer usually watches and not a spurious channel kept from the last time he used the TV. This use case is illustrated in **Figure 4**.

But the automation system has more to offer since there is also the intention that when a program

FIGURE 4 – Automatic program suggestion

FIGURE 5 – Automatic next program suggestion

FIGURE 6 – TV Concierge allowing the user to select the following suggestion

finishes and whenever the system has a high degree of confidence (still to be determined with real life data) TV Concierge will automatically start playing the next suggestion, without any intervention from the viewer. This situation is illustrated in **Figure 5** – at the end of the newscast the system suggested an episode from a series and started its reproduction automatically, from the beginning.

When the system does not have a sufficient degree of confidence, it will select a reduced number of proposals for the viewer. In this case, the playback will still start automatically, but the viewer will get, for a few seconds, the opportunity to choose another program to watch, a concept that is visible in **Figure 6**, mimicking a typical binge-watching scenario, which is somewhat the brand mark of the current OTT VOD systems.

It is also possible that the system has no suggestion to propose at a certain time, for example when no consumption pattern has been identified. In this case, the system needs to propose a new program, i.e., to really make a recommendation. Our approach for this use case is that the system will suggest a new release (e.g. a new season of a TV series) between a limited set of the most watched channels on that STB. The idea behind this methodology is that, usually, if a series that a viewer normally watches has ended, a new one will be released to take its timeslot, and the system will offer that. If that does not happen, some other channel that the viewer also regularly watches may have something new to promote that may also be used for the system to suggest (if the user likes the channel there is a higher probability that he also likes a new show from that channel). This is the purpose of the interface shown in **Figure 7** – in this case, we also opted for the binge-watching interface for the user to have an opportunity to actually select the new program.

Although the TV Concierge interactive application will start automatically when the STB is turned on and will keep providing suggestions and playing content in an automated way, this doesn't mean that the user relinquishes all the control

FIGURE 7 – TV Concierge proposing new releases in a binge-watching UI

of his STB. The user can, at any moment, use his interactive TV system as usual and TV Concierge will disappear. It can be summoned again by the standard means of the interactive platform, for instance with a menu item or from a dedicated button on the remote control. The system will also be able to re-engage and restart offering suggestions automatically if it detects that a program the viewer is watching just ended, even better it will do it as soon as the credits for a movie or a series are identified. That way it will be able to make a suggestion in the moment the user is more receptive to it since a program just ended and the viewer will be open to the propositions from the system.

The system will also reappear automatically with suggestions upon detection of what we call a "mindless zapping", that is when the user starts channel surfing in a pattern that appears to be just hunting for something to watch. Again, this is an use case where the user would be very approachable with a recommendation since he is not actually watching anything [10]. These two occasions are the moments we imagine a well-behaved human butler would step in to propose something, and of course, TV Concierge will do the same.

Conclusions

In this article we presented two fresh proposals for novel TV watching experiences. Both are based on the same personalization concepts and algorithms, but presenting the viewer with radically different ways to interact with them or them with the user. The TV Flow concept takes its inspiration from the current social media trends. It is not, however, a kind of social TV, essentially it brings to the TV world some of the concepts that we believe make social media engaging and appealing and we tried to repurpose and aggregate them to a unique UI approach. In what may be seen as a kind of antithesis, the TV Concierge proposal uses some of the concepts popularized by video OTT services but goes a step further by making features like binge-watching available in a traditional pay TV platform and by that removing most of the UI.

Both proposals are right now in the prototyping phase and as soon as some components are made available they will be assessed in a set of evaluation steps, initially in the lab and afterwards in field tests with real users. Since we believe that personalization is key for the future of user interaction in pay TV platforms, these two proposals put forward some concepts, taken from social media and OTT, which we think will play a relevant role in this scope.

Special thanks

The authors would like to thank all UltraTV teams involved in the ideation process in order to successfully reach a set of innovative use cases and, in special, to design team at the University of Aveiro for all the UI mock-ups that helped us a lot in the visualization process.

References

[1] J. Abreu, J. Nogueira, V. Becker, and B. Cardoso, "Survey of Catch-up TV and other time-shift services: a comprehensive analysis and taxonomy of linear and nonlinear television", *Telecommun. Syst.*, vol. 64, no. 1, pp. 57–74, 2017.

[2] Acordo Portugal 2020, "Portugal 2020", 2014. [Online]. Available: https://www.portugal2020.pt/Portal2020.

[3] M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience". New York: Harper Perennial, 1991.

[4] R. Turrin, A. Condorelli, P. Cremonesi, and R. Pagano, "Time-based TV programs prediction", in *1st Workshop on Recommender Systems for Television and Online Video at ACM RecSys.*, 2014.

[5] Digitalsmiths, "Q4 2015 Video Trends Report - Consumer Behavior Across Pay-TV, VOD, PPV, OTT, TVE, Connected Devices, and Content Discovery", 2015.

[6] M. Aharon, E. Hillel, A. Kagian, R. Lempel, H. Makabee, and R. Nissim, "Watch-It-Next: A Contextual TV Recommendation System", in *Proceedings, Part III, of the European Conference on Machine Learning and Knowledge Discovery in Databases - Volume 9286*, 2015, pp. 180–195.

[7] D. Gonçalves, M. Costa, and F. M. Couto, "A Flexible Recommendation System for Cable TV", in *3rd Workshop on Recommendation Systems for Television and Online Video (RecSysTV 2016)*, 2016.

[8] J. Vanattenhoven and D. Geerts, "Contextual aspects of typical viewing situations: a new perspective for recommending television and video content", *Pers. Ubiquitous Comput.*, vol. 19, no. 5, pp. 761–779, 2015.

[9] B. Schwartz, "The paradox of choice: Why more is less", 2004.

[10] J. Abreu, P. Almeida, and B. Teles, "TV Discovery & Enjoy: a New Approach to Help Users Finding the Right TV Program to Watch", in *TVX 2014 - Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video,* 2014.

14

Demystifying truths and myths of 5G technology

Francisco Fontes, Altice Labs
fontes@alticelabs.com

5G, the next generation of (wireless) communications, is full of expectations. This is exacerbated by the fact that many try to fit under 5G all the different evolution vectors of the telecommunications' ecosystem. It seems that there is an installed fear that the ones not protected under the 5G umbrella, will be demoted to run in a secondary telecommunications league and predestined to fail.

Keywords

5G; IMT2020; 3GPP; ITU

Introduction

Every ten years a new "generation" of terrestrial mobile wireless network emerged. It happened with 1G in the 1980s, 2G in the 1990s, 3G in the 2000s, 4G in 2012, and now the transition to 5G will be no exception with 5G being expected to have its first commercial deployments by 2020.

3GPP [1], who bears the ultimate responsibility for the standardization of 5G, felt the need to coin a "generation" logo for 5G, centring the 5G standardization role in this organization, thus clarifying what 5G is and what it is not. This never happened in the former generations, which were an aggregation of technical standards under promising, descriptive names, last one being Long Term Evolution (LTE) Advanced Pro.

Even if other "generations" were also embraced in expectation, this feeling is higher with 5G, as the technology is likely to have a much significant presence and impact in society, industry and economy. This is based on the increasing dependency the modern way of living has in wireless communications and mobile terminals, and 5G will foster it. Within that scope, and as IHS analysis states, 5G has the potential to become a General Purpose Technology (GPT): *"GPTs lead to deep and sustained impacts across a broad range of industries that often redefine economic competitiveness and transform societies"* [2]. That is also reflected on how the International Telecommunication Union Radio-communication Sector (ITU-R) refers to 5G: *"enabling a seamlessly connected society in the 2020 timeframe and beyond that brings together people along with things, data, applications, transport systems and cities in a smart networked communications environment"* [3].

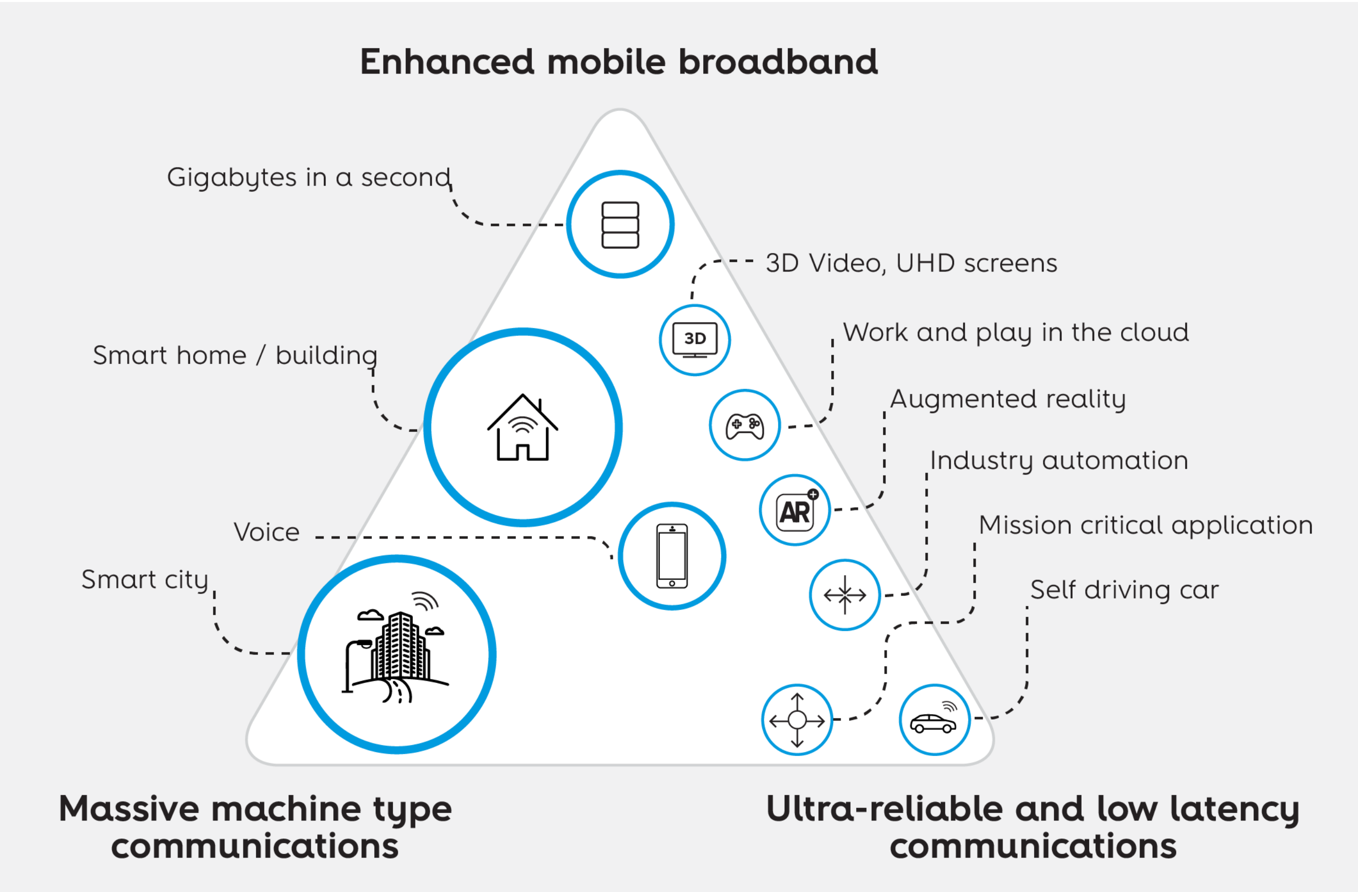


FIGURE 1 – ITU's IMT-2020 use cases organization [6]

ITU-R, in its Resolution ITU-R 65 [4], outlines the essential criteria and principles which will be used in the process of developing the recommendations and reports for International Mobile Telecommunications (IMT) for 2020 [3], including recommendations for the radio interface specification. One of the IMT objectives [5] is to set requirements which shall apply globally. IMT started with IMT-2000, setting the requirements for 3G. IMT-Advanced followed and 4G emerged as a technology targeting compliance with requirements set there. 3GPP aims at having its Release 15 and, especially, Release 16 as candidate technology, fulfilling ITU IMT-2020 requirements.

ITU identified usage scenarios for IMT-2020 (or use cases) and organized those according to three main vectors [6]: enhanced Mobile Broadband (eMBB), massive Machine Type Communications (mMTC) and Ultra Reliable and Low Latency Communications (URLLC), as represented in **Figure 1**.

In the same document and considering the previous usage scenarios, a number of key capabilities enhancements are identified (and shown in **Figure 2**) that set significant

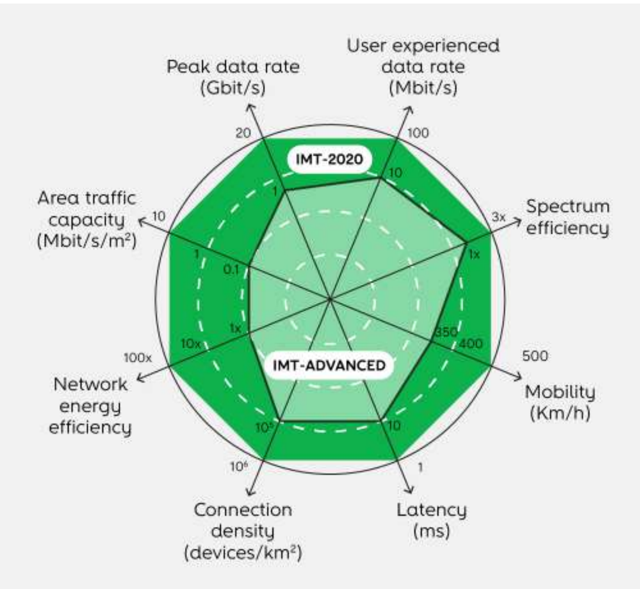


FIGURE 2 – ITU's identified key capabilities enhancements [6]

technological challenges going well beyond higher bandwidth and faster terminals. Energy efficiency and devices density requirements, for instance, push the supporting technology to fulfil Internet of Things (IoT) use cases requirements.

5G, under standardization by 3GPP, is a candidate technology to answer IMT-2020 requirements by addressing the identified usage scenarios.

5G standardization

As a natural continuation of previous wireless communications generations, the heart of 5G standardization work is being conducted by 3GPP. Naturally, there are other Standards Development Organizations (SDO) and *Fora* defining and standardising complementary technologies which will contribute to 5G feasibility, efficiency, reliability and fulfilment of defined IMT-2020 requirements, like Multi-access Edge Computing (MEC) and Network Functions Virtualization (NFV) from the European Telecommunications Standards Institute (ETSI) [7], complemented by Software Defined Networks (SDN) from the Open Networking Foundation (ONF) [8].

Besides IMT-2020 identified enhancements, 3GPP itself identified capabilities and extracted requirements for 5G from identified use cases [9] [10], as presented in **Table 1**.

Based on the previous requirements and capabilities, 3GPP provides a stage 2 definition (the functional capabilities and the information flows [11]) for a System Architecture for the 5G System [12], aiming at fulfilling IMT-2020 requirements. That definition is complemented by the identification of a set of "general concepts" that are the basis for that architecture definition:

1. Separate the User Plane (UP) functions from the Control Plane (CP) functions;

2. Modularize the function design, e.g. to enable flexible and efficient network slicing;

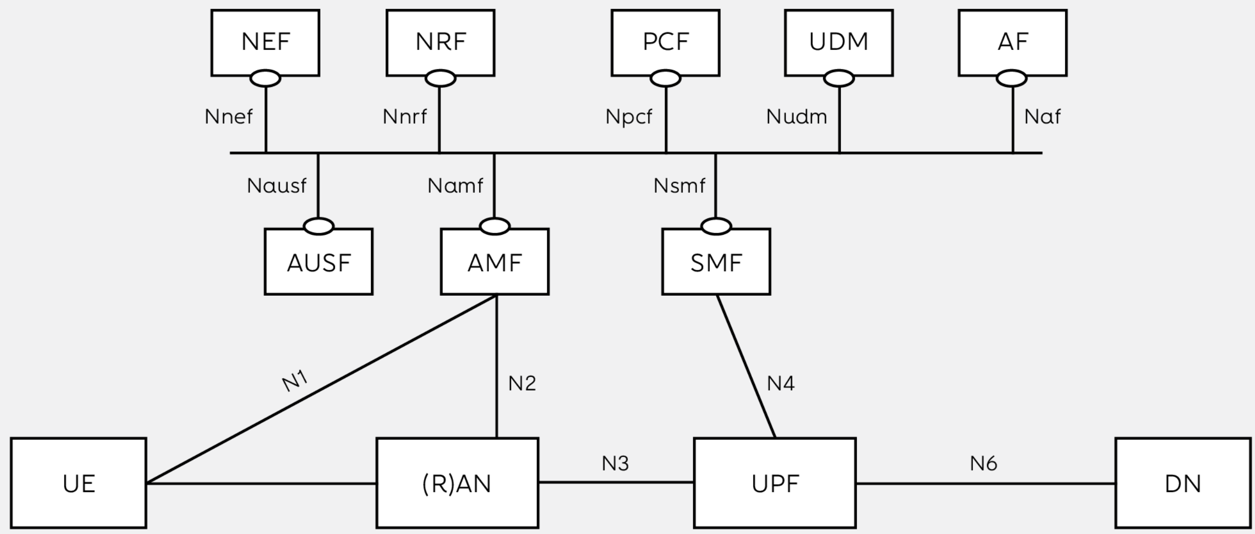
Basic Capabilities	
Network slicing	Self backhaul
Diverse mobility management	Flexible broadcast/multicast service
Multiple access technologies	Subscription aspects
Resource efficiency	Energy efficiency
Efficient user plane	Markets requiring minimal service levels
Efficient content delivery	Extreme long range coverage in low density areas
Priority, QoS, and policy control	Multi-network connectivity and service delivery across operators
Dynamic policy control	3GPP access network selection
Connectivity models	eV2X aspects
Network capability exposure	5G-RAN Sharing
Context aware network	
Performance Requirements	
High data rates and traffic densities	Higher-accuracy positioning
Low latency and high reliability	
Security Features	
Authentication	Regulatory
Authorization	Fraud protection
Identity management	Resource efficiency

TABLE 1 – 3GPP capabilities and requirements for 5G

3. Define procedures (i.e. the set of interactions between network functions) as services;
4. Enable each Network Function (NF) to interact with other NF directly if required (direct interaction);
5. Minimize dependencies between the Access Network (AN) and the Core Network (CN);
6. Support a unified authentication framework;
7. Support "stateless" NFs, where the "compute" resource is decoupled from the "storage" resource;
8. Support capability exposure;
9. Support concurrent access to local and centralized services;
10. Support roaming with both home routed traffic as well as local breakout traffic in the visited Public Land Mobile Network (PLMN).

In this architecture document, a 5G System is described as being a *"3GPP system consisting of 5G Access Network (AN), 5G Core Network and UE"* [12]. Complementarily, a 5G AN is defined as *"an access network comprising a NG-RAN (New Radio, NR) and/or non-3GPP AN connecting to a 5G Core Network"* [12]. **Figure 3**, taken from that same Technical Specification (TS), shows the 5G system architecture (refer to the TS for details, where a more traditional, reference points based system representation is available).

With the exception of the User Equipment (UE), (Radio) Access Network ((R)AN) and Data Network (DN), all other functional blocks belong to the 5G



Control Plane:
- AMF Access and Mobility Management Function
- AUSF Authentication Server Function
- NEF Network Exposure Function
- NRF NF Repository Function
- PCF Policy Control Function
- SMF Session Management Function
- UDM Unified Data Management

User Plane
- UPF User Plane Function
- UE User Equipment
- (R)AN (Radio) Access Network

Not represented:
- UDSF: Unstructured Data Storage function (used by any NF)
- SDSF: Structured Data Storage function (used by NEF)

Other (outside 5G scope):
- AF Application Function
- DN Data Network

FIGURE 3 – 5G system architecture [12]

Core Network and almost all are there because of the wireless and mobile features to be supported by the 5G system. Some of its general concepts are clearly reflected, like user and control planes separation. A service oriented control plane, suitable for a cloud deployment, is also reflected.

3GPP work on the radio interface definition is also progressing, from physical to upper layers. Recently, 3GPP agreed on anticipating the production of RAN related 5G standards, able to allow the usage of the 5G NR data plane, to provide increased performance for eMBB scenarios. This will be part of 3GPP Release 15.

The overall 3GPP 5G related standardization calendar is conditioned by ITU-R calendar (shown in **Figure 4**) since 3GPP is committed to submit a candidate technology to the IMT-2020 process. For that, *"3GPP has decided to submit the final specifications at the ITU-R WP5D meeting in February 2020, based on functionally frozen specs available in December 2019"* [13]. This will allow enough time for 3GPP specifications transposition for submission to IMT 2020 process by October 2020.

In this calendar, two relevant World Radio-communication Conferences (WRC), WRC-15 and WRC-19, in 5G evolution, can also be seen.

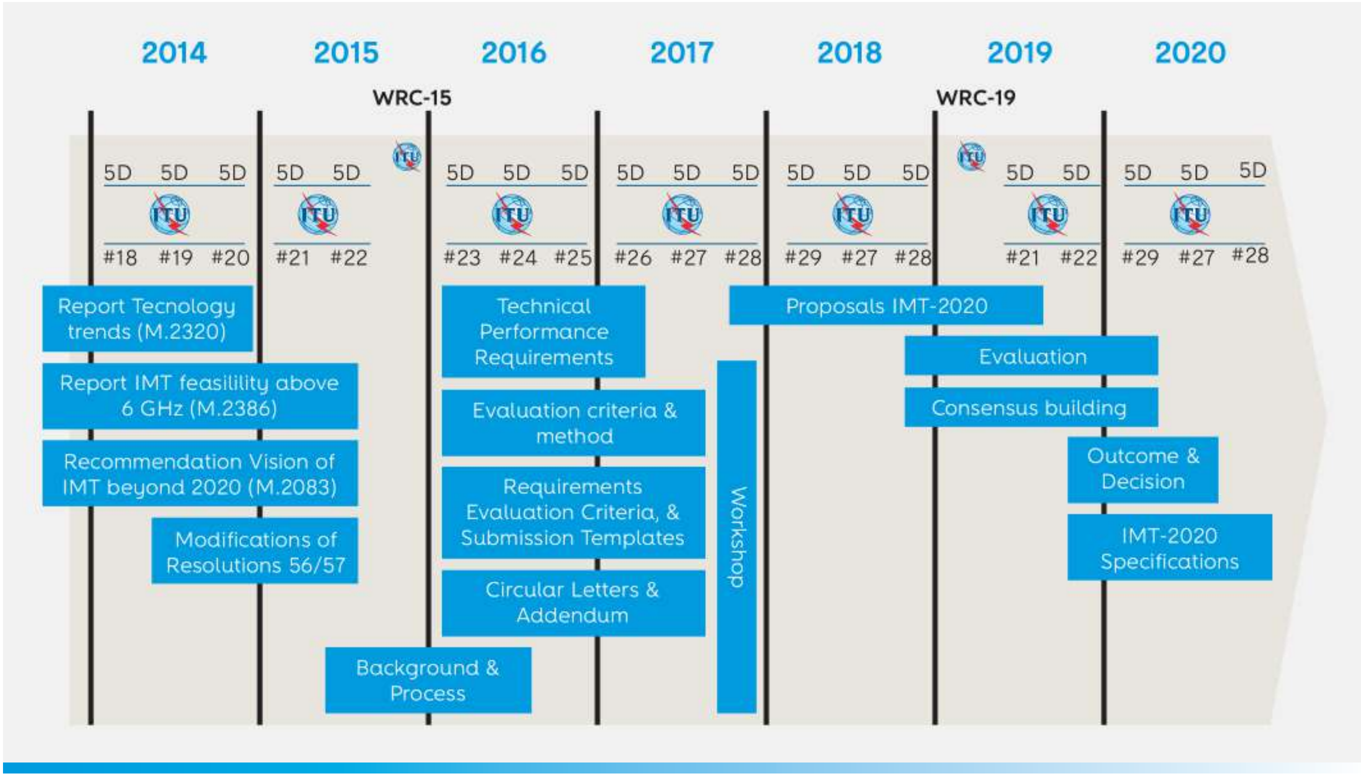


FIGURE 4 – Detailed timeline & process for IMT-2020 in ITU-R [3]

Questions and answers to demystify 5G

A lot of expectation exists around the 5G technology, sometimes inflated by the entities that want and need to make money out of the required investments associated with the technology deployment. The following answers to a set of selected questions, intend to demystify 5G, providing an overview, with some details, on this emerging technology, allowing each reader to make the right judgment and ascertain their own expectations on the technology.

How revolutionary is 5G relatively to 4G? Is it truly a revolution or just a natural evolution?

Technologically, 5G is an evolution from its predecessors but it targets a revolution, expanding significantly the supported use cases. And the difference between 4G and 5G do depend on what is being compared. If considering the successive 3GPP standardization Releases, 4G is Release 10 (the release which complies with requirements set by IMT-Advanced) and 5G starts on Release 15. In between, the different releases introduce many improvements to 4G (and to 3G) which will also appear in 5G, resulting in what is, for 4G, called Long Term Evolution (LTE), LTE-Advanced and LTE-Advanced Pro (**Figure 5**).

4G is significantly different from 3G, with the circuit-switching components absence (all services

FIGURE 5 – 3GPP logos regarding 4G evolution and 5G

are IP based) and providing higher bandwidth and lower latency. Now, in its essence, 5G adopts, while improving, many of the 4G principles, in areas like mobility, QoS control, and packet (IP) switching mechanisms, just to mention a few. However, it targets substantial improvements, set by ITU in IMT-2020, with a more significant impact on the new radio interface. **Figure 2**, taken from ITU-R M.2083 [6], summarizes in eight performance indicators a comparison of the two technologies.

Thus, the evolution from 4G to 5G will be gradual, with some 3GPP standard releases in between. As stated above, 5G Phase 1 will be standardized by 3GPP Release 15, with full 5G being achieved in Release 16 (full IMT-2020 compliance).

At the radio level, with the definition of a new radio interface, fundamental for 5G wireless target use cases, differences will be significant. One major new feature is the possibility to use multiple numerologies, by the adoption of a flexible frame structure. 3GPP provides a study [14] on the physical aspects of the "New Radio Access Technology", which will help 3GPP RAN Working Group (WG) to define and describe the potential physical layer evolution under consideration.

Is 5G more than just another (new) radio interface?

As seen before, a "5G Access Network", based on an NR interface, will be standardized as an integral part of the 3GPP "5G System", intended to have a universal applicability (in the eMBB, mMTC and URLLC use cases), and be connected to a new core, the 5G Core (5GC).

The definition of an NR interface is of primary relevance for 5G success. In order to be a universal wireless interface, it must provide significant performance improvements, reducing latency, providing higher bit rates, with more efficiency and with slicing capabilities being extended to it. Many of the imposed key capabilities (e.g. latency and bandwidth) depend on the new technological solutions this radio interface will present.

The 3GPP 5GC will handle other types of access networks, able to connect to interfaces N2 and N3 (see **Figure 3**), even if requiring interworking modules (N3IWF for non-3GPP accesses). Thus, 5G encompasses an innovative radio access, which is complemented by a new core, able to provide its services also to legacy and other future fixed and wireless access networks.

But most of the 5GC complexity, reflected in the identified functional block and capabilities, exists because it targets, in first hand, the support of wireless accesses and mobile terminals. Adding the support of other access technologies allows for richer use cases, where wired and wirelessly connected devices of different natures take part.

The adoption by the operators of a new common 5G core impacts overall network deployments. This shall be in the roadmap for operators that want to have a truly converged fixed-mobile network, exploiting technologies like NFV to address diverse networking and evolving requirements.

Why an access agnostic 5G core network?

Mobile networks, be it any of the 3GPP RAN, Wireless Local Area Networks (WLAN) or any other wireless access network, depend significantly on fixed infrastructure (in fact only terminals are mobile). But fixed networks do not need many of the complex mechanisms the former ones require. Network attachment processes, quality of service and mobility support do not exist or are much simpler for wired accesses. Thus, why integrate "fixed networks" in a common core and make them more complex than what they need to be? Or why incorporate in a common core the characteristics that make it agnostic to access types? First, having this allows independent core and access evolutions since none is developed thinking about specific characteristics of the other. Next, having an integrated network control, e.g. where IP addresses are assigned and managed by the same entities, simplifies inter-technology handovers and, overall, allows for a more rational control plane. Users' authentication methods and credentials can also be unique, normally with stronger ones adopted by "mobile networks". Applicability scenarios involving diverse wired and wireless access networks can also be targeted.

Additionally, future networks will be more programmable, with information and functions being exposed, via open interfaces, to external entities. Having a common core allows having all that in the same place, with a holistic network view. For instance, network slicing "on demand", a hot topic associated to 5G, could in that way easily address wireless and wired access networks together.

Can 5G exist without SDN and NFV? Which are 5G supporting technologies?

5G NR and the CN, both part of a 5G System, are being defined according to a set of new technological trends. Modularity, programmability, virtualization and user/control planes separation are some of those. NFV and SDN are technologies emerging in that context but they are orthogonal to 5G. While SDN contributes (at least with the associated concepts) and may even impose changes in networks architecture (e.g., Control and User Plane Separation, CUPS), NFV targets the mechanisms for managing the lifecycle of the network (virtualized) functions. NFV does not define what they do and on how they relate, thus it does not define the technology. However, technologies may be defined in such a way that they make better use of NFV (and SDN), without being dependent on it. It is the responsibility of 3GPP to guarantee 5G work goes in that direction. Some features, like slicing, for instance, one of the 5G selling factors and an important aspect when it comes to providing support to the different vertical markets over a single physical infrastructure, will benefit significantly if it is implemented with NFV mechanisms.

SDN and NFV will define and provide support to other network functions deployment, well beyond 5G scope, as the 5G definition is focused in the specification of functions and interactions, which may or not be realized with such technologies. Current technologies like 3G, 4G, GPON, HFC, WLAN and other networking technologies will and can already benefit from SDN, NFV, MEC (see below) and optical technologies. In fact, it is already happening with, e.g., C-RAN, vEPC and vIMS implementations and deployments being "virtualized" (NFV) and adopting CUPS principles.

MEC is another technology which, not being 5G specific, will contribute to its feasibility. One of the key capabilities 5G has to tackle is the required extremely low latency (<1ms) and MEC may play an important role here, by bringing computational infrastructure closer to the network edge. But even with MEC, if no changes are made to the radio interface and access nodes, latency values in the order of a few milliseconds will never be achieved.

Thus, the path to 5G must incorporate its own, specific, technological evolutions, and it is possible that some very specific scenarios are implemented without NFV. Nevertheless, SDN and NFV will certainly have an important role in future networks, where 5G will have its own share and role.

Is 5G feasible without biting another portion of radio spectrum?

Spectrum is an essential aspect of wireless communications. Both the location and size of available spectrum chunks matter. Lower frequencies are good to reach farther, while higher frequencies are essential to be fast. The amount of spectrum impacts the speed directly (400MHz are foreseen for the more demanding 5G scenarios). However, higher frequencies mean smaller cells and the signal not crossing all materials. Thus, its applicability may be limited, e.g., to some hotspots with high commercial value, like indoor hotspots in shopping malls. Thus, in order to address all the planned use cases, it is important to have spectrum available in a large range of frequencies.

mMTC and URLLC scenarios may be foreseen at lower frequencies, while higher frequencies will be appropriate for eMBB. Several frequency bands are targeted for 5G:

- Sub-1GHz: for extended 5G coverage, out of the big urban centres and deep inside buildings;
- 1 to 6GHz: for a good compromise between coverage and speed;
- Above 6GHz: to provide very high-speed mobile broadband.

In that same context, the European Radio Spectrum Policy Group (RSPG) [15] released a document [16] regarding the radio spectrum for 5G:

- <1GHz (e.g. 700MHz) to *"enable nationwide and indoor 5G coverage"*;
- 3400-3800MHz, >100MHz (400MHz) of continuous spectrum, to *"put Europe at the forefront of the 5G deployment"*;
- 24.25-27.5GHz, a *"pioneer band for earlier implementation in Europe"*;
- 31.8-33.4GHz *"looks a promising band which could be made available"*;
- 40.5-43.5GHz *"is a viable option for 5G in the longer term"*.

The appointed higher frequency band will be key for the fastest 5G services and to support the expected mobile traffic growth. The technology will be 5G only when key capabilities enhancements set by IMT-2020 are fulfilled. Besides others, bandwidth performance enhancements to be achieved are challenging:

- Area traffic capacity: 10Mbit/s/m2;
- User experienced data rate: 100Mbit/s;
- Peak data rate: 20Gbit/s.

To achieve this performance level, especially the last one, and even if using short distances with high-level constellations (256QAM) providing a high spectral efficiency, still higher frequencies are required. Thus, the new spectrum claim is a performance requirement, not a functional one.

In that same line, the radio technologies envisaged for 5G shall work at any frequency. Thus, spectrum selection for 5G is also not a technological question. It is a strategic decision, to be taken by politicians and operators. The World Radio-communication Conference 2019 (WRC-19), *"where deliberations on additional spectrum are taking place in support of the future growth of IMT"* [3], is an important milestone in 5G spectrum definition. Its agenda item 1.13 states *"to consider identification of frequency bands for the future development of International Mobile Telecommunications (IMT), including possible additional allocations to the mobile service on a primary basis, in accordance with Resolution 238 (WRC-15)"* [17].

Complementarily, even if licensed spectrum will be the focus for 5G, this can be complemented with unlicensed spectrum. While licensed spectrum is essential to protect operators investment and assure service quality to customers, the simultaneous usage of both will help increase user experience, especially in situations of congestion.

Technology-neutral spectrum is another relevant topic for 5G, allowing operators to refarm frequency bands currently used for other technologies, simplifying the upgrade of their networks and promoting a more efficient spectrum usage.

Is there a hype in 5G foreseen use cases?

Almost all organizations addressing 5G have a document on use cases. For some, this was useful for requirements extraction. For others, it is used to demonstrate 5G potential and expected impact.

While previous mobile generations also announced innovative use cases support, those were focused in Mobile Broad Band (MBB) and 5G has the ambition to go beyond that. Considering ITU usage scenarios (**Figure 1**), MBB will be extended in eMBB use cases, but mMTC and URLCC clearly intend to take 5G to areas where other wireless but especially wired technologies were dominant. However, some of the appointed use cases may go beyond a reasonable boundary.

Remote surgery is one of them. Fixed networks, most likely fibre based, most reliable and immune to uncontrollable phenomena like atmospheric effects, are much more likely to support such scenarios where human life may be at risk. But, are wired access technologies part of 5G? See previous answers on that subject.

The same applies to autonomous cars. These may be assisted by the network (V2x scenarios) but no vendor will make its driving algorithms, which will also imply liability, based in components that are out of their control, like mobile communications, even if a global URLLC coverage exist, what will not happen soon.

In addition, a significant part of the foreseen use cases can already be covered by 4G and its near future evolutions. Many IoT-related ones are such a case. NB-IoT and LTE-M, standardized by 3GPP, are there. What may be missing is value chain readiness.

So, 5G will substitute WLAN, right?

3GPP New Radio, as stated by Qualcomm [18], aims at becoming a “unified air interface”, meaning *”it will not only significantly enhance mobile broadband, but will also enable new services such as mission-critical control and massive IoT”*. Thus, at least from the purely technical point of view, 5G may be considered to have the potential to substitute WLAN. However, why shall it be different from the current situation with 4G and the broadband services it provides access to? Why is WLAN preferred over 4G to stream videos and access social networks?

First, WLAN normally comes for free and provides more bandwidth, even if with no quality guarantees. In contrast, often 4G is used even where WLAN is available, as a kind of ubiquitous access network, especially when data plans are more generous. But, still, it cannot compete with WLAN speed, especially when this one is backed by a fast fixed access line (e.g. fiber GPON).

It is worth noting that service guarantees do not come for free. WLAN operates under unlicensed spectrum and the technology does not have strong resources management mechanisms. In opposition, mobile networks operate under licensed spectrum and have strong resources management mechanisms. This will be even more important with 5G, considering the large number of different use cases to be simultaneously supported.

Thus, these and other types of wireless accesses are expected to continue existing. WLAN will continue evolving in parallel to 5G. Scenarios of coordinated simultaneous usage will be possible and exploited with the help of the new 5G, access agnostic, Core Network, taking benefit of the different characteristics and operation modes they provide.

How does 5G relate to other hot topics like smart cities and Industry 4.0?

Smart cities and Industry 4.0 (I4.0) are about use cases. 5G is about technology to fulfil use cases requirements. 5G or, more precisely, IMT-2020 use cases were organized by ITU in the following areas: eMBB, mMTC and URLLC (as explained before). There, ITU clearly situated smart cities in the “machine type communications” (3GPP terminology for IoT) and I4.0 in the URLCC domain.

Some of the less demanding smart cities and I4.0 use cases may be supported by 4G and other existing, specifically tailored, wireless and wired technologies. However, highly demanding use cases require support for a large number of connected devices, with very disparate connectivity and bandwidth needs, long lasting batteries, very low latency and highly reliable communications. Supporting all these, plus the eMBB scenarios, by the same unifying technology and over the same shared physical infrastructure, is what 5G aims for. Having this ”universal tool” is a relevant, differentiator factor. Overall, 5G will certainly power I4.0 and smart cities communications. Having it as an ubiquitous technology is a question of simplification, which means lower costs and higher efficiency.

What is the expected 5G deployment timeline?

The 5G deployment timeline will be strongly determined by the evolution of the euphoria around this new technology, which is fed by the technology success in itself, and by the ecosystem that will develop around it. Operators, network equipment vendors, services and applications vendors, the business verticals that will be served, and the standardization bodies will make the push that will define how fast the 5G calendar will evolve.

The pressure towards 5G standardization recently pushed 3GPP to anticipate some 5G RAN standards, with a focus on features required for eMBB. The 3GPP RAN group agreed in establishing an intermediate milestone for March 2018, targeting the early completion of the Non-Standalone (NSA) 5G NR mode for the eMBB use case. In NSA mode, connections are anchored in LTE while 5G NR carriers are used to boost data-rates and reduce latency. The Standalone (SA) 5G NR mode is set for September 2018. That means first standards compliant equipment, especially terminals, may start to appear by the end of 2018 (NSA) and 4Q19 for full 5G systems, both according to Release 15 standards and targeting eMBB scenarios. mMTC and URLLC will be supported in Release 16 which means 2022 for first commercial use cases.

Some global events will advance the development and demonstration of innovative services, which will demonstrate 5G advantages, even if initially supported by pre-5G standards and by prototypes.

The 2018 Winter Olympics in Pyeongchang, South Korea, will be the first opportunity to unveil innovative use cases, made possible by 5G, to a large audience, supported by the South Korean vendors. Also next year, the 2018 FIFA World Cup will take place in Russia *"with the Government and mobile operators planning to create a pilot 5G zone"* [19].

The 2020 Summer Olympics in Tokyo, Japan, and UEFA Euro 2020 in 13 different European cities will show the first real 5G terminals and networks. Later, the 2022 Winter Olympics in Beijing, China, will see full 5G implementation, with significant Chinese vendors and operators showing all the 5G potential.

Paris Summer Olympics, taking place in 2024, will already benefit from 5G cruising speed, with the focus on exploiting the technology rather than in deploying it.

In the meantime, Europe is committed to having at least one city per member state with 5G by 2020. This is part of the European Commission (EC) strategy for 5G, stating that by 2025 all major cities and transport paths must have 5G coverage. In fact, the EC sees 5G as a technology that can boost European economy. Current work done by the 5G Infrastructure Public Private Partnership (5G-PPP) trials Work Groups [20] identifies current and future private and public 5G trials, and a list of "5G cities", which will be the first to showcase 5G. Thus, 5G activity will accelerate in the coming months and years, till first commercial deployments.

Will 5G promote different value chain(s)?

"Business as usual" will not work for 5G: operators deploying wireless, broadband connectivity are struggling with reduced margins, fostered by a strong competition, while OTT services take advantage of the established, excellent and expensive infrastructure.

First, investments in 5G infrastructure will be higher, in order to exploit all the technology potential. That needs new spectrum, a larger number of cell sites, front/backhaul/core networks capacity expansion (the unseen wired part of mobile networks), training, and, of course, 5G specific functional modules, at access and core, and equipment, e.g. antennas.

Second, 5G allows operators to addresses other markets, where "enhanced" mobile broadband, providing Internet access according to “net neutrality” principles, will be a limited part of the business, expanding to industry and automotive, specifically addressing verticals of interest. This may lead to business models where the entities exploiting those verticals will be called to share their profit, willing to pay for a differentiated service. The foreseen creation of network slices (a hot 5G topic), providing a tailored network behaviour with services *a la carte*, will be a good way for operators to finance the required 5G capex and opex.

In this context and now fostered by 5G required investments for the next years, the mentioned “net neutrality” principles are again a topic under discussion. Some see its maintenance as a blocker to 5G deployment. Thus, even for pure mobile Internet accesses, things may change.

In that context, 5G business models may well be more complex and of a different kind, if not for another reason, at least because network operators won´t be able to survive under the current models.

Will 6G appear by 2030?

References to 6G and even 7G can already be found. However, most likely, 5G will be the last generation of “mobile communications”. In fact, mobile and fixed are merging and 5G already reflects that, defining an access agnostic core network. With the 5G System Architecture, the Core Network will evolve per se, and access networks, wired and wireless, will, in independent processes, emerge, evolve and connect to it, with very disparate release calendars.

Direct interactions between Core Network functional blocks will exist and be service based, easing their replacement and the addition of new ones. Thus, evolutions may be expected to happen, independently and in different areas, without the need for the definition of a complete new 6G system architecture.

In that context, a significant next milestone in digital communications may occur when those evolve to a non-packet-based one, more efficient and flexible, with all the impact that will have. ETSI Next Generation Protocols (NGP) [21], in a first instance, and Quantum processing and communication, later, are in that evolutionary path. Till then, we will assist to a continuous increment in bitrates, without any further ”big step for technology” [22]. Hence, if a sixth generation is to exist, it will probably be due to marketing reasons.

5G opportunities and impacts

5G Systems will be more complex than 3G/4G ones, providing support to a larger number of use cases. Thus, while it threatens some, it also opens a large spectrum of opportunities for new developments. The following areas are here identified.

5G specific technology and components

As described, 5G Systems are composed of an (R)AN and Core. A New Radio will exist as a new 3GPP RAN and the new Core is composed of many new and redefined functional blocks. A careful look into that is required, identifying direct and indirect opportunities for new developments and the impacts it may have in existing solutions, with the required adaptations and enhancements to be made.

Supporting technologies

5G incorporates many recent concepts and its deployment will most likely be based on some emerging technologies. The move towards network “softwarization” and “cloudification” is unstoppable. SDN and NFV represent this trend, for 5G as well as for other networks. Mastering these technologies (and an overall approach increasingly closer to IT) is key for future networks, creating the basis for a good network planning and operation.

MEC is emerging but still involved in uncertainty. It is ETSI contribution to answer low latency requirements and create an IT environment where only operators can provide it.

Fixed/Mobile Convergence (FMC) will be seen in other areas, besides in the 5G Core Network, which will be agnostic to access network technologies. FMC will include the transport of mobile and fixed data and control signals, sharing the same access and transport infrastructure. Evolution of fronthaul and backhaul will result in more efficient and flexible packet based solutions.

Supporting systems

OSS/BSS must evolve, following future networks trends. Network management is expanding to other areas, resulting from the flexibility and agility that SDN and NFV provide, namely to autonomic behaviour, where big data is used to provide data to AI-based analytics, which trigger the orchestration of actions on the network.

Along with flexibility, network softwarization (NFV, SDN, cloud) - not only for 5G but as a major trend - is expected to bring additional complexity to operations. Processes like software upgrades, which are already complex today with only a few providers and one or two updates per year, might become completely unmanageable. A whole new approach is needed, not only in terms of technology but also to processes across the organization. DevOps represent a new approach to tackle this challenge.

New services & applications

Considering the target key performance enhancements set by ITU-R for IMT-2020, its fulfilment allows 5G to enlarge its scope of applicability. A large number of use cases have been devised and many verticals have been appointed which may take benefit from 5G. For sure others will exist but it is certain that a new generation of applications and services will appear.

Conclusions

This article intends to clarify what the 5G technology is, what are the realistic expectations to have, where are the main challenges to face, what is its current status, focusing in standardization, how does it reflect and impact future networks, ending, not exhaustively, with a brief identification of development and operational areas open by 5G.

ITU, via IMT-2020, set the requirements for 5G technologies. Answering to that, 3GPP defined a "5G System" and is specifying its main components: a new radio interface and an access agnostic new core network. This happens in a process of continuous evolution, where 5G appears as 3GPP Releases 15 and 16 milestones. IMT-2020 requirements push the 5G technology to target use cases that go well beyond better mobile broadband, positioning itself to provide support to, e.g., IoT and I4.0.

Emerging technologies like SDN, NFV and MEC deployments will ease 5G deployments and its operation, representing a new way of building communication networks, where flexibility and fast response times will be key.

It is also explained that the technology per se does not require additional spectrum, but the required performance is only achieved if higher frequencies and bandwidths become available.

5G targets its first commercial deployments around 2020, with first trials and demonstrations appearing before, especially in the scope of significant international events, as stated.

References

[1] 3GPP, "3GPP," 2017. [Online]. Available: www.3gpp.org.

[2] "The 5G economy: How 5G technology will contribute to the global economy," IHS Markit, London, UK, 2017.

[3] ITU-R, "ITU towards "IMT for 2020 and beyond"," ITU, 2017. [Online]. Available: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx.

[4] ITU-R, *Resolution ITU-R 65, Principles for the process of future development of IMT for 2020 and beyond, 2015*.

[5] ITU-R, *FAQ on International Mobile Telecommunications (IMT)*, 2013.

[6] ITU-R, *Recommendation ITU-R M.2083-0, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond"*, September: ITU, 2015.

[7] ETSI, "European Telecommunications Standards Institute," 2017. [Online]. Available: http://www.etsi.org/.

[8] ONF, "Open Networking Foundation," 2017. [Online]. Available: https://www.opennetworking.org/.

[9] 3GPP, TS 22.261, *"Service requirements for next generation new services and markets", Release 15*, 2017.

[10] 3GPP, TR 38.913, *"Study on scenarios and requirements for next generation access technologies", Release 14*, 3GPP, 2016.

[11] ITU-T, *Recommendation I.130 : Method for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN*, ITU, 1988.

[12] 3GPP, TS 23.501, "System Architecture for the 5G System", Release 15, 3GPP, 2016.

[13] 3GPP, "Tentative 3GPP timeline for 5G," 3GPP, 17 March 2015. [Online]. Available: http://www.3gpp.org/news-events/3gpp-news/1674-timeline_5g.

[14] 3GPP, TR 38.802, *"Study on new radio access technology Physical layer aspects"*, Release 14, 3GPP, 2017.

[15] RSPG, "Radio Spectrum Policy Group," [Online]. Available: http://rspg-spectrum.eu/.

[16] Radio Spectrum Policy Group, *Strategic Roadmap Towards 5G for Europe, Opinion on spectrum related aspects for next-generation wireless systems (5G)*, 2016.

[17] ITU-R, "ITU-R Preparatory Studies for WRC-19," 2017. [Online]. Available: http://www.itu.int/en/ITU-R/study-groups/rcpm/Pages/wrc-19-studies.aspx.

[18] Qualcomm, *Making 5G NR a reality*, San Diego, CA, USA: Qualcomm Technologies, Inc., 2016.

[19] TASS, Russian News Agency, "Russia connected. 5G for Russian digital economy," 29 May 2017. [Online]. Available: http://tass.com/sp/948333.

[20] 5G-PPP, "5G-PPP Work Groups," 5G-PPP, [Online]. Available: https://5g-ppp.eu/5g-ppp-work-groups/.

[21] ETSI, "Next Generation Protocols (NGP)," 2017. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/next-generation-protocols.

[22] F. Fontes, "'5G': a big step for technology, one giant leap for mankind," *InnovAction*, no. 1, pp. 80-93, 2016.


5G

15

Multi-access edge computing a 5G technology

Carlos Parada, Altice Labs
carlos-f-parada@alticelabs.com

Cristina Leitão, Altice PT
cristina-i-leitao@telecom.pt

Francisco Fontes, Altice Labs
fontes@alticelabs.com

Vitor Cunha, Instituto de Telecomunicações
vitorcunha@av.it.pt

One of the most challenging KPI defined by the ITU IMT-2020 for 5G is "latency below 1ms". This is a key requirement for an emerging era of applications, such as virtual and augmented reality, video analytics, or Industry 4.0. Those applications require extremely low latency to enable the so-called tactile Internet. However, relying only on large centralized data centres as we do today, the latency requirements defy the laws of physics.

Keywords

5G; Edge Computing; MEC; PoC; Evaluation

Introduction

An International Telecommunication Union - Radiocommunication Sector (ITU-R) 5G Report draft [1] approved by November 2017, includes the 13 *"minimum requirements related to technical performance for IMT-2020 radio interface"*. Those performance indicators are based on 8 key capabilities, also identified by ITU-R, which are defined in [2]. The same approach was already adopted for 3G and 4G, with ITU setting up the requirements for the International Mobile Telecommunications (IMT) 2000 and IMT-Advanced, respectively. These requirements are depicted in **Figure 1**, which compares 5G requirements (IMT-2020) with 4G requirements (IMT-advanced).

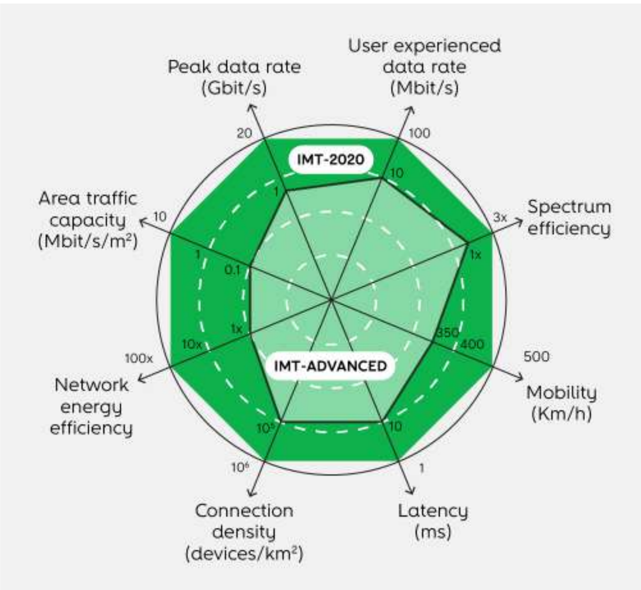


FIGURE 1 – IMT-2020, 5G key capabilities enhancements [2]

One of the most challenging 5G targets is maximum latency of 1 ms. This is a key requirement for Ultra Reliable and Low Latency Communications (URLLC) applications, where applications like Industry 4.0, augmented reality, virtual reality, code offloading, or new gaming experiences, just to name a few, are included.

Today, many services are provided from centralized data centres (Clouds), where massive amounts of computing, storage and network resources are located. Although most services can be efficiently delivered (at low cost) in this way, in order to achieve millisecond-level latencies traditional architectures have to be re-engineered. The solution is to move some computing capacity to edge clouds and deliver certain services from edge data centres, smaller in size, but closer to customers' devices.

This paradigm shift, generically designated as *edge computing*, is becoming a very popular network architecture trend and an essential pillar for 5G. Approaching in services and customers is the easier solution to provide the low latencies required for the 5th generation. The European Telecommunications Standards Institute (ETSI) created a group devoted to this topic and some results have already been released.

This article starts with an introduction to the edge computing concept and its contribution to 5G. The section "ETSI MEC" describes the edge computing architecture, according to the ETSI MEC (Mobile/ Multi-access Edge Computing), which is currently developing the technology. "Prototype" section presents the Altice Labs prototype, explaining the currently available features, the development strategy, and the next development steps. "Proof-of-Concept" section elaborates on the PoC launched in collaboration with Altice PT. Finally, the section "Evaluation" shows the preliminary evaluation results of this PoC and section "Conclusions and Future Work" concludes summarising the key points of this article.

ETSI MEC

In December 2014, the ETSI board decided to create the Industry Specification Group (ISG) MEC. Initially, in phase 1, MEC was a short for Mobile Edge Computing, targeting only mobile networks. Later, in phase 2, MEC became the multi-access edge computing, reflecting a new and broader perspective of the group, aiming to extend the edge computing concept to access networks other than mobile, like wired (copper, fibre) or Wi-Fi.

The ETSI ISG MEC intends to standardize a framework, able to run 3rd party non-telco applications at the edge, providing a dynamic environment on top of cloud technologies. In addition, 3rd party applications can benefit from locally provided services (e.g. location, radio network information, time of day), leveraging the ability to provide new and attractive applications to end customers. This framework also envisions powerful management and orchestration abilities, taking full advantage of cloud principles, such as flexibility, agility and scalability. With these features, it can be decided whether Mobile/Multi-access Edge Applications (ME Apps) should be instantiated, scaled, moved or terminated in a particular edge cloud.

Generally speaking, the MEC system environment is characterized by: *ultra-low latency*, which ensures the extremely low latencies required by some applications; *bandwidth efficiency*, allowing for the reduction of traffic flowing between the edge and the core, thus saving bandwidth; and *local services accessibility*, providing locally available services to leverage applications.

The ETSI ISG MEC has standardized the MEC architecture [3], as depicted in **Figure 2**. Readers familiar with the ETSI Network Function Virtualisation (NFV) architecture may find this similar as it was inspired by it.

The architecture can be split horizontally by the dashed line into edge level (one per edge), and the system level, with only one centralized system level. The edge level includes all the components required to be at the edge, while the system level

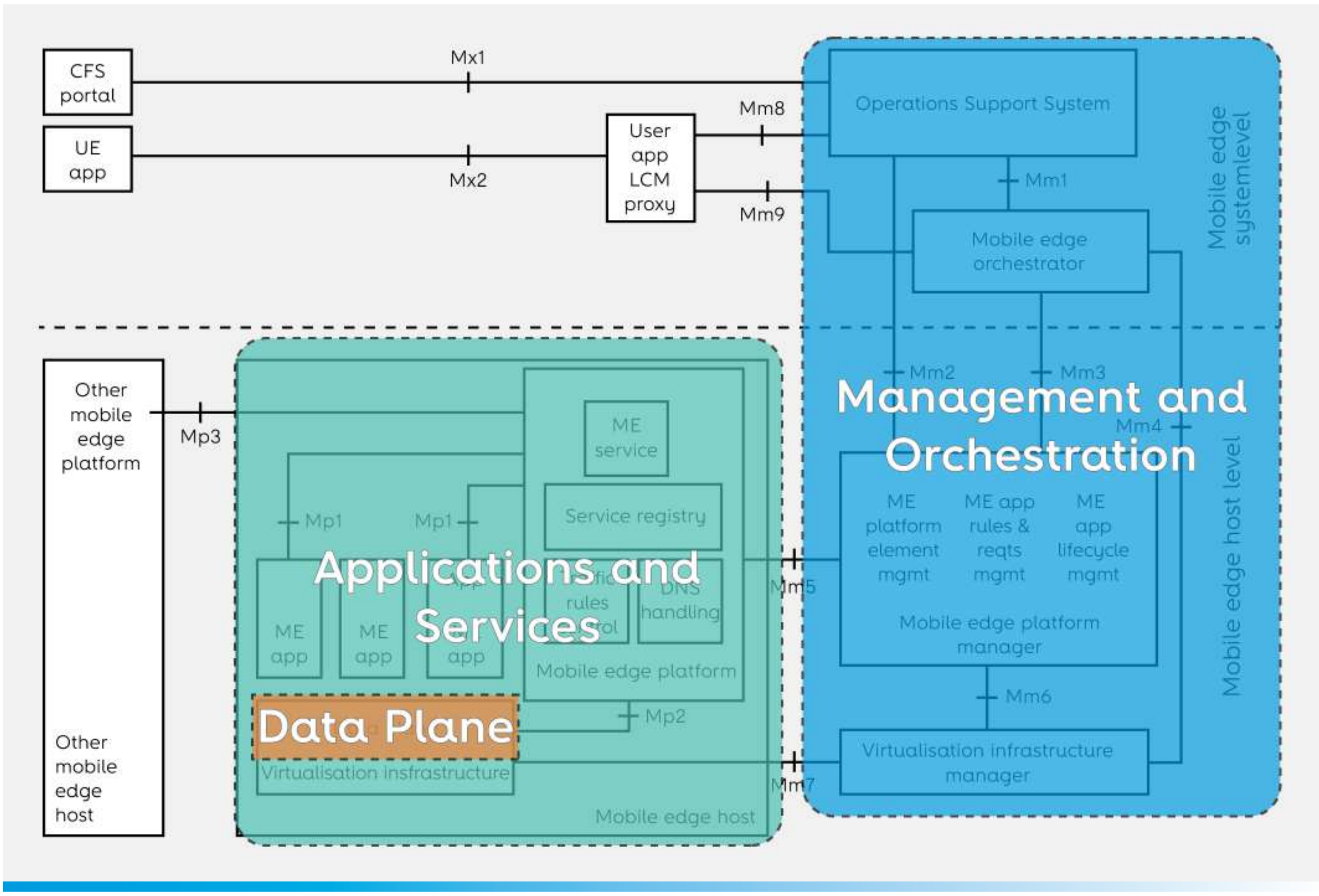


FIGURE 2 – ETSI MEC architecture [3]

comprises the edges, together with the centralized management and orchestration components.

The architecture can also be split vertically, by considering the execution environment, represented on the left, and the management and orchestration environment, on the right. The execution environment is composed of the edge infrastructure, running ME Apps and the ME Platform. The MANagement and Orchestration (MANO) is composed of the edge and centralized management components, which are responsible for making decisions about the execution environment.

Functionally, three main macro-blocks can be identified, overlaying on **Figure 2**: Data Plane, Applications and Services, and Management and Orchestration.

The Data Plane macro-block is located at the user data plane. Although the ETSI MEC does not define a particular network location, it should be somewhere between the Radio Access Network (RAN) and the Core (S1-U interface), usually physically deployed in traditional Central Offices (CO), typically the same deployment location as the future Cloud-RAN (C-RAN) environment. However, it can be located somewhere else, as ETSI ISG MEC is agnostic about the edge location. Its main responsibility is to inspect the entire user traffic and divert (offload) some packets towards a particular ME App instantiated at the edge, according to a set of rules. This action is expected to be transparent both to the end user and to the network.

The Applications and Services macro-block comprises two main components: the Cloud Infrastructure, a Network Function Virtualisation Infrastructure (NFVI), the same component as defined in the ETSI NFV architecture, supporting the instantiation and execution of ME Apps; and the ME Platform, which provides local services exposed via API (e.g. user context information, network conditions, user location) to be consumed by ME Apps. ME Apps can also provide services and expose API to be consumed by other ME Apps.

The Management and Orchestration macro-block manages the ME Apps lifecycle at the edge (e.g. instantiate, move, scale), and the ME Platform and Services, allowing the discovery and authorization of Services. Furthermore, it orchestrates the entire MEC system, deciding where - edge - and when ME Apps are instantiated, scaled, moved or disposed of.

The MEC technology may benefit from the NFV adoption, in particular, the virtualization of the access network, e.g. C-RAN for mobile, virtual OLT (vOLT) for Fiber To The Home (FTTH). Access virtualization requires operators to bring IT resources to edge data centres, leveraging e.g. the traditional CO as a convergent location of access technologies. This is also a good candidate location to host edge computing nodes.

In such NFV edge environment, MEC can take advantage of already deployed resources, reducing technology adoption costs. Initially, MEC can start by using NFV-unused IT resources, increasing the utilization as edge computing becomes popular and more service providers get interested in deploying their applications at the edge. In a second phase, additional resources can be provisioned specifically for MEC purposes.

Prototype

Today, Altice Labs is actively contributing to ETSI MEC standardization. In parallel, it is prototyping a solution to fully implement all components of the architecture. The prototype follows the MEC architecture, splitting the development into 3 major building blocks: Traffic Offloading Function (TOF), responsible for offloading the User Equipment (UE) traffic to the ME Apps; Mobile Edge Host (ME Host), responsible for hosting ME Apps at the edge data centre and offer services APIs; and MANO, responsible for managing and orchestrating the entire system. **Figure 3** depicts the software architecture of the developed prototype.

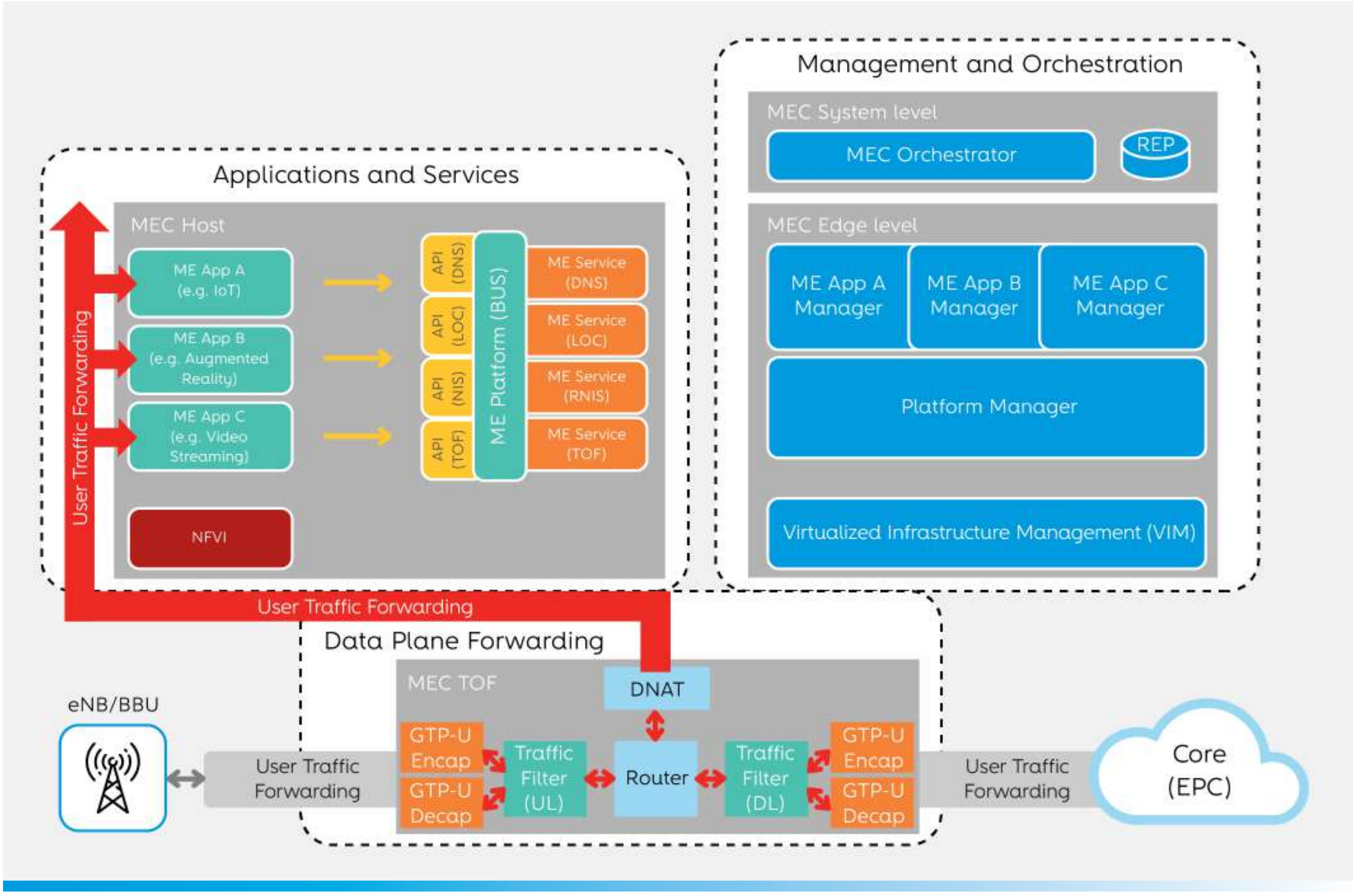


FIGURE 3 – Overall prototype software architecture

ME TOF

The TOF is responsible for diverting (offloading) specific UE traffic towards particular ME Apps instantiated on the edge, or to forward it, as usual, towards the core. The TOF development relies on a Linux platform and uses well-known tools to implement the required capabilities. Examples of used technologies are Open Virtual Switch (OVS), network namespaces, or netfilter/iptables, among others.

The TOF function interrupts the S1-U interface, acting as an L2 switch and connecting to the RAN in one leg and to the CORE in the other. It should be introduced transparently without any integration work on the mobile network. There is no need to change network parameters (e.g. IP adresses) allocated to the enhanced Node B (eNB) on the RAN side, the Serving GateWay (S-GW) on the Core side, or any other element. The only requirement is that the TOF can receive an IP address for address resolution purpose by the Address Resolution Protocol (ARP). Once plugged, it is able to manage everything by learning from the General Packet Radio Service (GPRS) Tunnelling Protocol (GTP) headers.

By default, the TOF has four distinct interfaces (virtual or physical): the eNB interface, towards the RAN and where the UE connects; the S-GW interface, towards the S-GW in the Core, Internet direction; the Apps interface, towards the edge cloud, where ME Apps are instantiated; and the API interface devoted to management, exposing APIs to provision TOF offloading rules.

The TOF is designed as a pipeline of small functional blocks, which are chained together to build the expected behaviour, following the Software-Defined Networks (SDN) pattern. The approach brings significant advantages: new functional blocks can be added and removed as requirements change, these are easier to update and replace, and can be distributed along multiple physical machines, easing the load balancing and scaling of the system.

Following a multi-access approach, the TOF is composed of the main body, common to any access and where the UE plain non-encapsulated IP traffic is handled and controlled via Openflow, and a set of plug-ins, responsible for the encapsulation/decapsulation of specific access networks. Currently, only the GTP plug-in is available, but others will be made available.

Figure 4 depicts the TOF flow diagram for traffic coming from the UE. GTP traffic is received from the S1-U interface and forwarded to the first data plane function, the GTP Filter. This function determines whether a particular GTP packet needs to be sent to the TOF Controller Agent in order to make a control decision, or can proceed in the data plane. If a control decision is required (e.g. discovery of new UEs), the TOF Controller Agent analyses the packet and enforces control actions on the TOF Service Function Forwarder (SFF). Then it can proceed to the next function, the Encapsulator/ Decapsulator, where the GTP header is removed.

After GTP decapsulation, the plain IP packet proceeds to the Classifier function, which inspects the traffic and determines the chain the packet must go next: either be offloaded or be forwarded to the core (as usual). In the future, other options may exist, e.g. ME Apps chaining.

Finally, if the packet is to be forwarded to the core, it is sent through the S-GW interface. If it is to be offloaded, it goes through a Network Address Translation (NAT) function, translating the destination IP to the ME App edge IP, and forwarding it to the edge cloud (Apps interface). The use of IP NAT can impact some applications that expect unchanged/original packets. For this reasons, in the future, the offloading may use tunnelling protocols, like Generic Routing Encapsulation (GRE) or Virtual eXtensible Local Area Network (VxLAN), instead of NAT, to accommodate ME Apps that require unmodified traffic, e.g. traffic analysers.

Pinger is a helper function crucial to the TOF design and operation. As the TOF learns GTP tunnels automatically from the traffic, it relies on the existence of some traffic between the Core (S-GW) and the RAN (eNB). If no traffic ever comes from the core, there is no way to discover

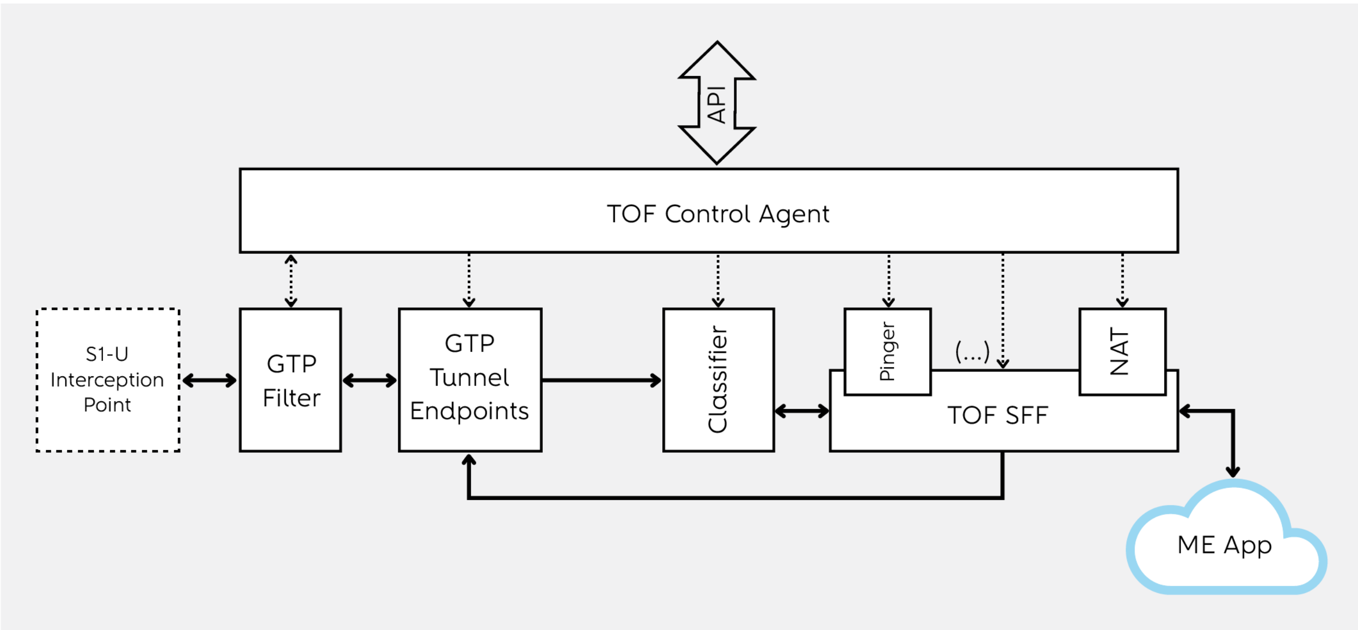


FIGURE 4 – ME TOF pipelining functions

the downlink Tunnel Endpoint ID (TEID), which identifies UE. A simple approach is to generate ICMP requests, or pings, on the UE's behalf, forcing some GTP traffic to be exchanged.

Finally, the TOF SFF is responsible to perform packet switching according to a flow-table. This table is dynamically modified by the TOF Control Agent to implement the required chains, for example, to decide whether a packet must be forwarded or offloaded. This function is aligned with the SFF entity of the Internet Engineering Task Force (IETF) Service Function Chaining (SFC) Working Group [4].

ME Host

The ME Host is composed of two main components: the Edge Virtualization Infrastructure and the Mobile Edge Platform (MEP). The Edge Virtualization Infrastructure is an Openstack-based cloud which provides computational, storage, and networking resources to run ME Apps at the edge. The MEP comprises a collection of ME Services that can be consumed and/or provided in the edge environment, such as Location or Radio Network Information. Both components are managed and orchestrated by the ME MANO. So far, there is a basic implementation of the Edge Virtualization Infrastructure and its connectivity to the TOF and MEP. The MEP implementation is an ongoing work, although not yet available.

Considering that ME Apps may belong to different 3rd parties, it is mandatory to design a solution with appropriate isolation, avoiding direct connectivity among the different parts. To achieve this, ME Apps are deployed in different cloud tenants, taking advantage of the Openstack isolation capabilities. This way, when the TOF sends traffic to an ME App it must send it through the Apps interface to an Internal Network/Subnet (see **Figure 5**), which will forward the traffic to the ME App's tenant private networks. To allow for this connectivity, every ME App needs a floating IP reachable from external networks. This is the only visible IP for the ME App, and the one targeted when the NAT is performed by the TOF or, in the future, a tunnel endpoint.

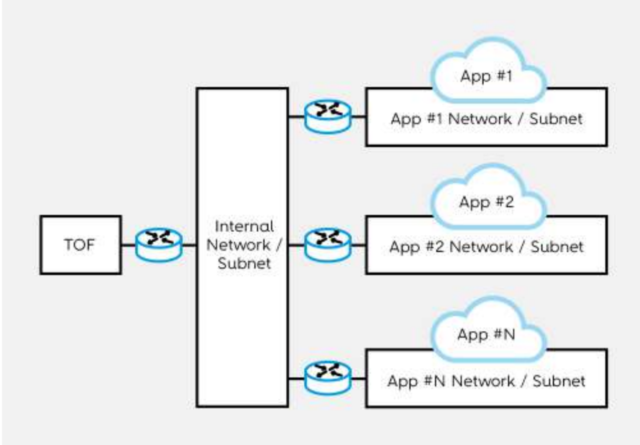


FIGURE 5 – ME Host and TOF communication strategy

A similar design is used to access the ETSI MEC MP1 interface, which connects ME Apps with the ME Platform, providing access to Service APIs (see **Figure 6**).

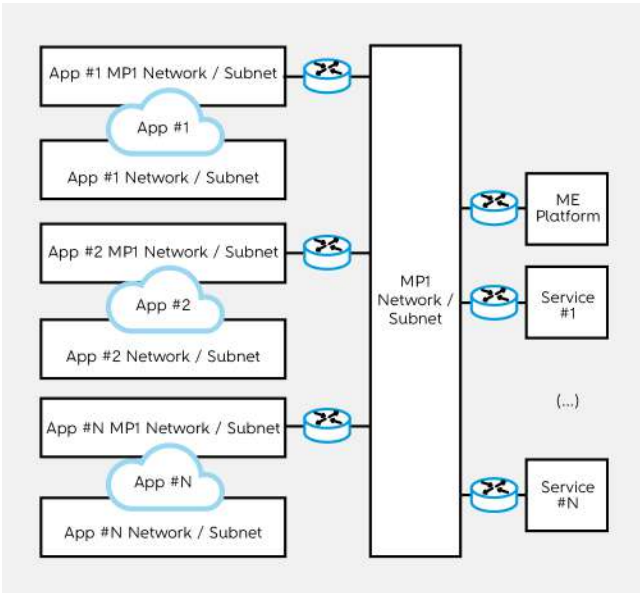


FIGURE 6 – ME Host and ME Platform communication strategy

ME MANO

The ME MANO is a complex component whose major asset is flexibility. To decide on the support of these functions, an exhaustive evaluation work was conducted, to select the most suitable open source tool to be used. According to these results, the Open Source MANO (OSM) tool [5] was chosen. Although the OSM has been created to implement NFV MANO functions, the MEC components are very similar. Nevertheless, an adaptation work still has to be done.

Most OSM adaptations required to comply with the MEC framework are related to the data models, using TOSCA/YAML to describe ME Apps, which are slightly different from the ones used for Virtual Network Functions (VNF). Also, there are some differences in the management and orchestration mechanisms for, by definition, MEC environments are highly distributed, with tens or hundreds of small edge data centres spread all over the operator's access network, and ME Apps may need to be deployed on any of them, according to the associated policies.

Another relevant difference is in the entity layers. While in the NFV world there are two main entities, VNF and Network Services (NS), in MEC there is only a single entity: ME Apps. For this reason, we implemented ME Apps as VNFs and created an additional NS with a single VNF inside. This is mandatory as OSM does not allow the deployment of VNF, but only NS. So, to onboard an ME App we need to create two different descriptors: a VNF Descriptor (VNFD) and an NS Descriptor (NSD).

The OSM is built by the integration of three main components: Riftware [6], a Network Service Orchestrator (NSO), responsible to orchestrate NSs in the NFV context, OpenMANO [7], an open source Resource Orchestrator (RO) led by Telefónica, and Juju [8], a configuration manager devoted to performing ME App lifecycle management.

Proof-of-concept

To validate the developed MEC prototype in a Lab environment, a PoC was built in the Altice PT lab, in Lisbon. For this PoC, Altice PT mobile lab elements were used, replicating the equipment and network structure deployed in production all over the country.

A Huawei eNB and a Cisco 4G Core – Evolved Packet Core (EPC) were used. Connected to the 4G mobile was a Windows laptop tethered via USB to a Samsung phone acting as a router.

The MEC solution was deployed in an all-in-a-box model, meaning that the TOF and the ME Host were deployed on the same physical machine (see **Figure 7**). The ME MANO was not available at this stage. The hardware is COTS, with the characteristics identified in **Figure 7**.

The ME App used was the Unified Origin (UO), a video content delivery application capable of performing video transmuxing functions, performing multiple video encapsulations to accommodate different devices and platforms. The advantage of deploying this at the edge is the capability to receive a single raw video content stream and make the adaptation closer to the end user, saving bandwidth and storage capacity.

As **Figure 7** depicts, the MEC box interrupts the S1-U interface in the user data plane (S1-U) acting as an L2 switch. For the sake of simplification, the link between the eNB and the S-GW is represented as a line, but in fact, there are several Local Area Networks (LAN) and routers involved. There is also a centralized App deployed at Altice Labs premises, here representing an ME App deployed on the Internet. To evaluate the PoC performance, a ME App was instantiated on-demand.

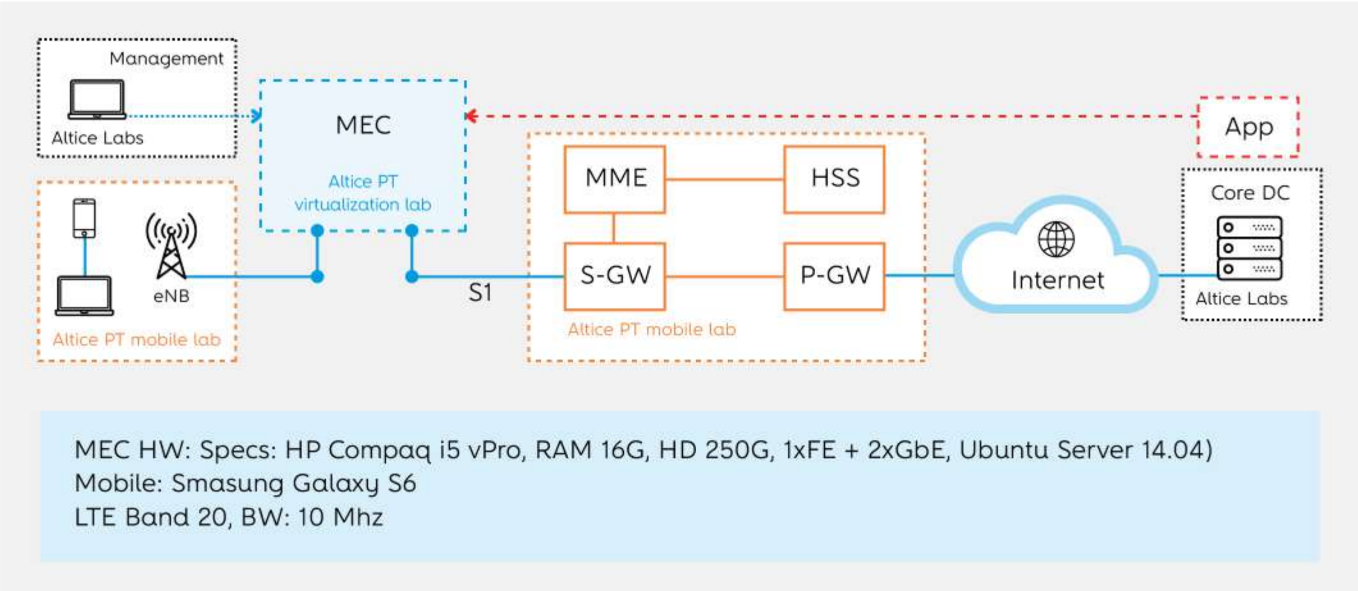


FIGURE 7 – PoC architecture at Altice PT mobile lab

Evaluation

This section describes the set of six tests performed in order to evaluate the behaviour and performance of the MEC prototype. This evaluation was performed in the Altice PT PoC environment described in the section above. **Table 1** to **Table 6**, describe the test performed as well as the results obtained.

Handover time test	
Description	The goal of this test is to measure the service handover time between core and edge. <u>Handover time edge to core:</u> Assuming the service is provided from the edge, it measures the time elapsed to move the service to the core. That includes the TOF provisioning to offload that particular traffic and the disposal of the server at the edge cloud. <u>Handover time core to edge:</u> Assuming the service is provided from the core, it measures the time elapsed to move the service to the edge. That includes the instantiation of the server at the edge cloud and the TOF provisioning to offload that particular traffic.
Methodology	This test is performed by measuring the time elapsed between the handover request and the moment when the service is completely instantiated and delivering content.
Measurement tools	No specific measurement tool is used.
Metrics and results	edge_to_core_handover = **1m 13.3s** core_to_edge_handover = **0m 29.0s**

TABLE 1 – Handover time test definition and results

Throughput test	
Description	The goal of this test is to measure the maximum throughput allowed on the TOF, both in the case when all traffic is offloaded to an edge application and in the case when all traffic is forwarded towards the core. This test also intends to measure the percentage of CPU utilization in both situations. TOF offloading throughput: Throughput when all traffic is offloaded to an edge application (UDP and TCP). CPU utilization: CPU (%) in both cases above. TOF forwarding throughput: Throughput when all traffic is forwarded towards the core (UDP and TCP). CPU utilization: CPU (%) in both cases above.
Methodology	This test is performed by generating and measuring the throughput in both cases (offloading and forwarding) and analysing the CPU utilization.
Measurement tools	The tools used were *iperf* (traffic generation/measure) and *top* (CPU%).
Metrics and results	tof_offloading_throughput = **300Mbit/s (UDP), 400Mbit/s (TCP), 20%** tof_forwarding_throughput = **300Mbit/s (UDP), 400Mbit/s (TCP), 15%** ** This result was limited by the traffic generation. It seems the TOF can do a lot more.*

TABLE 2 – Throughput test definition and results

Round-Trip latency test	
Description	The goal of this test is to measure the Round Trip (RT) latency from the UE to the UO server, both when the service is delivered from the core (Internet) and when it is delivered from the edge (MEC). Core RT latency: RT latency between the UE and the server when the service is delivered from the core (Internet). Edge RT latency: RT latency between the UE and the server when the service is delivered from the edge (MEC).
Methodology	This test is performed by measuring the RT delay using a TCP ping tool towards the server and calculating the average of N consecutive requests.
Measurement tools	The tool used was the PsPing tool (Windows platform).
Metrics and results	core_rt_latency = **30.7 milliseconds** edge_rt_latency = **17.1 milliseconds** ** Core results will greatly depend on where the Internet is (how close is the data centre).*

TABLE 3 – RT latency test definition and results

MEC Round-Trip latency impact test	
Description	The goal of this test is to measure the RT extra latency introduced by the TOF on the traffic, in the situation where packets are forwarded towards the core of the network (we could not measure towards the edge due to the lack of an appropriate tool). RT Extra Latency TOF Forwarding: RT extra latency introduced by the TOF when the traffic is forwarded towards the core, simply because this component is in the network.
Methodology	This test is performed by measuring the difference between RT latency for traffic forwarded to the core using MEC and a passive UTP cable connector, i.e.: *core_rt_latency (using MEC) - core_rt_latency (using switch)*
Measurement tools	The tool used was the PsPing tool (Windows platform).
Metrics and results	extra_core_rt_latency = **0.3 milliseconds** (30.7s – 30.4s)

TABLE 4 – MEC Round-Trip latency test definition and results

Backhaul bandwidth savings test	
Description	The goal of this test is to calculate the backhaul bandwidth savings when using the MEC technology. This compares the bandwidth that should be used when the service is provided from the core versus when the service is provided from the edge.
Methodology	This test is performed by making calculations for different scenarios about the amount and percentage of backhaul bandwidth saved for a different number of requests of the same video, assuming video streams of 3Mbit/s, i.e., for N video streaming requests: *saved_bandwidth_mbits (N) = (N – 1) * 3Mbit/s* *saved_bandwidth_percent (N) = (N – 1) / N*
Measurement tools	No tool is needed, only calculations.
Metrics and results	saved_bandwidth (2) = **3Mbit/s, 50%** saved_bandwidth (10) = **27Mbit/s, 90%** saved_bandwidth (50) = **147Mbit/s, 98%**

TABLE 5 – RT latency test definition and results

Handover service loss test	
Description	The goal of this test is to measure the service loss caused by the handover, both from the core to the edge and from the edge to the core. Note that this test only considers whether the service is affected, even though some traffic may be lost. Core to Edge Service Loss: Service loss caused by a handover from the core to the edge. Edge to Core Service Loss: Service loss caused by a handover from the edge to the core.
Methodology	This test is performed qualitatively by checking if any glitch can be observed in the video (more objectively, it can also be checked if the buffer reaches the size of 0).
Measurement tools	No tool is used, only observation.
Metrics and results	edge_to_core_handover_loss = **No service loss observed** core_to_edge_handover_loss = **No service loss observed**

TABLE 6 – Handover service loss test definition and results

Conclusions and future work

This article presents the edge computing technology as a key to 5G requirements fulfilment since it can reduce latency dramatically and save uplink bandwidth. This benefits emerging applications for which latency is critical and promotes a more efficient use of network resources.

To address this topic, ETSI has created the ISG MEC. Altice Labs is actively involved in ETSI ISG MEC standardization, contributing to this group since the very beginning.

In parallel, Altice Labs has been developing a MEC-compliant prototype. The scope of this work covers the entire ETSI MEC architecture, including the TOF, ME Apps Hosting, ME Platform and MANO features.

At the time of writing, Altice Labs has already built a basic prototype, able to demonstrate the MEC concept by implementing the TOF and a simplified ME Host. The ME Platform and MANO components are in the queue for further developments.

Using the current version of the Altice Labs MEC prototype, a PoC has been performed in collaboration with Altice PT, using the Mobile Lab with network equipment for validation. It was possible to perform an easy integration thanks to the L2 model designed for TOF, which is a strong asset to deploy the MEC technology on existing networks.

The PoC evaluation demonstrated the advantages of the MEC technology, reducing latency and increasing bandwidth efficiency. The prototype performance was very promising, allowing the deployment of edge ME Apps in around one minute, reducing the network latency roughly by half, and showing high forwarding and offloading capabilities of at least 300Mbit/s, using COTS hardware.

In conclusion, this PoC demonstrated the potential of the MEC technology and the flexibility of the adopted design options. It allowed for an easy integration with existing mobile networks, and has shown to be capable of providing high-performance solutions, even when using low-end segment hardware.

The work produced until now will continue for the next months, extending the scope of the development to the ME Platform and the MANO layer, to automate the MEC environment.

Special thanks

The authors would also like to thank the following persons for their contributions, either direct or indirect, to the projects that made this article possible:

- Carlos Marques, Altice Labs (carlos-martins-marques@alticelabs.com)
- Daniel Corujo, Instituto de Telecomunicações (dcorujo@av.it.pt)
- Isabel Borges, Altice Labs (isabel@alticelabs.com)
- João Paulo Barraca, Instituto de Telecomunicações (jpbarraca@av.it.pt)
- José Filipe Pereira, Altice PT (jose-g-pereira@telecom.pt)

References

[1] ITU-R, *Draft new Report ITU-R M.[IMT-2020.TECH PERF REQ] - Minimum requirements related to technical performance for IMT-2020 radio interface(s),* International Telecommunication Union, 2017.

[2] ITU-R, *Recommendation ITU-R M.2083-0, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond"*, September: ITU, 2015.

[3] Mobile Edge Computing (MEC) ETSI Industry Specification Group, "ETSI GS MEC 003 V1.1.1, Mobile Edge Computing (MEC); Framework and Reference Architecture," ETSI, Sophia Antipolis, France, 2016.

[4] IETF, "IETF SFC Working Group," [Online]. Available: https://datatracker.ietf.org/wg/sfc/.

[5] ETSI OSM, "Open Source MANO," [Online]. Available: https://osm.etsi.org/.

[6] RIFT.io Inc., "RIFT.ware," [Online]. Available: https://riftio.com/.

[7] Telefónica, "OpenMANO," [Online]. Available: http://www.tid.es/long-term-innovation/network-innovation/telefonica-nfv-reference-lab/openmano.

[8] Canonical, Ltd., "Juju," [Online]. Available: https://jujucharms.com/.

Acronyms

1

1G First generation mobile networks

2

2FA Two Factor Authentication
2G Second generation mobile networks

3

3D Three-dimensional
3G Third generation mobile networks
3GPP 3rd Generation Partnership Project

4

4G Fourth generation mobile networks
4K 3840 x 2160 pixels resolution

5

5G Fifth generation mobile networks
5GC 5G Core
5G-PPP 5G Infrastructure Public Private Partnership

6

6G Sixth generation mobile networks
6LoWPAN IPv6 over Low-Power Wireless Personal Area Networks

7

7G Seventh generation mobile networks

8

802.1x IEEE Standard for port-based Network Access Control
8K 7680 x 4320 pixels resolution

A

AAA Authentication, Authorization and Accounting
ABR Adaptative Bit Rate
ACM Active Campaing Manager
ADSL Asymmetric Digital Subscriber Line
AES Advanced Encryption Standard
AGI Artificial General Intelligence
AI Artificial Intelligence
AN Access Network
API Application Program Interface
App Application
AR Augmented Reality
ARP Address Resolution Protocol
ARPU Average Revenue Per User/Unit
ASN Autonomous System Numbers
AWS Amazon Web Services

B

B2B Business-to-business
B2C Business-to-consumer
BGP Border Gateway Protocol
BIRD Bird Internet Routing Daemon
BM Business Model
BPI Baseline Privacy Interface
BPI+ BPI Plus
BSS Business Support System

C

C.2020 Consumer 2020
CA Certification Authority
CAD Computer Assisted Design
CAE Computer-Aided Engineering
CAM Computer-aided Manufacturing
capex Capital Expenditures
CASD Controlling Authority Security Domain
CDN Content Delivery Networks
CDVR Cloud Digital Video Recorder
CEIIA Centro de Excelência para a Inovação da Indústria Automóvel
CENC Common Encryption
CF Collaborative Filtering
CFD Computational Fluid Dynamics
CI Certificate Issuer
CIA Central Intelligence Agency
C-Level Chief-Level
CM Cable Modem
CML Câmara Municipal de Lisboa
CMTS Cable Modem Termination System
CN Core Network
CNN Cable News Network
CO Central Office
COO Country Of Origin
COTS Commercial Off-The-Shelf
CP Control Plane

CPE	Customer Premise Equipment
CPMI	Control-Plane Management Interface
CPS	Cyber-Physical Systems
CPU	Central Processing Unit
CRA	Certificate Requesting Agent
C-RAN	Cloud RAN
CSP	Communications Service Providers
CTO	Chief Technology Officer
CUPS	Control and User Plane Separation
CVC	Code Verification Certificates

D

DApp	Decentralized Application
DASH	Dynamic Adaptive Streaming over HTTP
DDoS	Distributed Deny of Service
DFKI	German Research Centre for Artificial Intelligence
DHCP	Dynamic Host Configuration Protocol
DI-Labs	Digital Innovation Labs
DLT	Distributed Ledger technologies
DN	Data Network
DOCSIS	Data Over Cable System Interface Specification
DPA	Data Protection Authorities
DPIA	Data Protection Impact Assessments
DPO	Data Protection Officer
DPoE	DOCSIS Provisioning of EPON
DRM	Digital Rights Management
DSP	Digital Service Provider
DSR	Demand-Side Response
DT	Deutsche Telekom
DTSI	Deutsche Telekom Strategic Investments
DVR	Digital Video Recorder

E

EC2	Elastic Compute Cloud
EAE	Early Authentication and Encryption
EAP	Extensible Authentication Protocol
EAP-AKA	EAP Authentication and Key Agreement
EAP-AKA'	EAP Authentication and Key Agreement Prime
EAP-SIM	EAP Subscriber Identity Module
EAP-TLS	EAP Transport Layer Security
EAP-TTLS	EAP Tunneled Transport Layer Security
EC	European Commission
ECASD	eUICC CASD
ECOMP	Enhanced Control, Orchestration, Management and Policy
EDI	Electronic Data Interchange
EDP	Energias de Portugal
EID	eUICC-ID
eMBB	enhanced Mobile Broad Band
EMEL	Empresa Municipal de Mobilidade e Estacionamento de Lisboa, E.M. S.A.
eNB	enhanced Node B
EPC	Evolved Packet Core
EPG	Electronic Programming Guide
EPON	Ethernet PON
ERP	Enterprise Resource Planning
eSIM	embedded SIM
ETSI	European Telecommunications Standards Institute
EU	European Union
EU-28	European Union of 28 Member States
eUICC	embedded Universal Integrated Circuit Card
EV	Electrical Vehicles
eV2X	enhanced Vehicle to everything

F

FCC	Fast Channel Change
FEA	Finite Element Analysis
FIFA	Féderation Internationale de Football Association
FI-PPP	Future Internet Private-Public Partnership
FITMAN	Future Internet Technologies in the Manufacturing
FIWARE	Future Internet (like software and hardware)
FMC	Fixed Mobile Convergence
FP	FairPlay
FP7	European Union Framework Programme Number 7
FTTH	Fiber To The Home
FreeBSD	Free Berkeley Software Distribution

G	GAFA	Google, Amazon, Facebook and Apple (group of companies made up of)
	GBA	Generic Bootstrapping Architecture
	GbE	Gigabit Ethernet
	GDPR	General Data Protection Regulation
	GLA	Greater London Authority
	GMT	Greenwich Mean Time
	GPON	Gigabit Passive Optical Network
	GPRS	General Packet Radio Service
	GPS	Global Position System
	GPT	General Purpose Technology
	GRE	Generic Routing Encapsulation
	GSMA	Global System for Mobile Communications Association
	GTP	GPRS Tunnelling Protocol

H	H2020	Horizon 2020
	HD	High Definition
	HDFS	Hadoop Distributed File System
	HFC	Hybrid Fiber Coaxial
	HGW	Home Gateway
	HLS	HTTP Live Streaming
	HOT	Israeli telecommunications company from Altice Group
	HSS	HTTP Smooth Streaming
	HTML5	Hyper-Text Markup Language (version 5)
	HTTP	HyperText Transfer Protocol

I	I4.0	Industry 4.0
	IBM	International Business Machines
	ICMP	Internet Control Message Protocol
	ICO	Initial Coin Offering
	ICT	Information and Communications Technology
	IDaaS	Identity as a Service
	ID	Identification/Identifier
	IDC	International Data Corporation
	IETF	Internet Engineering Task Force
	IIoT	Industrial Internet of Things
	IIS	Internet Information Services
	iOS	Mobile operating system created and developed by Apple Inc.
	IMDb	Internet Movies Database
	IMS	IP Multimedia Subsystem
	IMSI	International Mobile Subscriber Identity
	IMT	International Mobile Telecommunications
	INESC TEC	Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
	IoE	Internet-of-Everything
	IoT	Internet of Things
	IP	Internet Protocol
	IPDR	Internet Protocol Detail Record
	IPO	Initial Public Offering
	IPTV	Internet Protocol Television
	IQ	Intelligence Quotient
	ISD-P	Issuer Security Domain - Profile
	ISD-R	Issuer Security Domain - Root
	ISG	Industry Specification Group
	ISIM	IP Multimedia Services Identity Module
	ISO	International Organization for Standardization
	ISO 27001	Information Security Management System standard by ISO and International Electrotechnical Commission
	ISP	Internet Service Providers
	IT	Information Technology
	ITU-R/T	International Telecommunication Union, Radiocommunication Sector/Telecommunication Standardization Sector
	iTV	Interactive television
	IVP	Infinite Video Platform
	IX	Internet eXchanges

K	KNN	K-Nearest Neighbours
	KPI	Key Perform Indicators

L	L2	Layer 2
	Lab/Labs	Laboratory/Laboratories
	LAG	Link Aggregation Group
	LAN	Local Area Network
	LED	Light Emitting Diode
	LISA	Local Integrated Software

		Architecture (Apple LISA)
	Lisboa E-Nova	Agência de Energia e Ambiente de Lisboa
	LMA	Logging, Monitoring and Alarms
	LPA	Local Profile Assistant
	LPAd	LPA device
	LPAe	LPA in the eUICC
	LTE	Long Term Evolution
	LTE-cat0	LTE Category-0
	LTE-M	LTE category M1

M	**M2M**	Machine-to-Machine
	MAC	Media Access Control
	MANO	MANagement and Orchestration
	MBB	Mobile Broad Band
	MBMS	Multimedia Broadcast/Multicast Service
	Mcast	Multicast
	ME	Mobile/Multi-access Edge
	ME Apps	Mobile/Multi-access Edge Applications
	MEC	Multi-access Edge Computing
	MEP	Mobile Edge Platform
	MFA	Multi-Factor Authentication
	MFG	Manufacturer
	Mgmt	Management
	ML	Machine Learning
	mMTC	massive Machine Type Communications
	MNO	Mobile Network Operator
	MNO-SD	MNO Security Domain
	MOCAP	Motion Capture
	MP	Media Plane
	MPMI	Media-Plane Management Interface
	MR	Mixed Reality
	MS	Microsoft
	MSO	Multiple System Operator
	MVP	Minimal Viable Product

	NAA	Network Access Application
N	**NAT**	Network Address Translation
	NB-IoT	Narrow Band IoT
	NF	Network Function
	NFV	Network Function Virtualization
	NFVI	Network Function Virtualisation Infrastructure
	NGH	Next Generation Hotspot
	NGINX	Engine X
	NGP	Next Generation Protocols
	NG-RAN	Next Generation Radio Access Network
	NLG	Natural Language Generation
	NLP	Natural Language Processing
	NMS	Network Management System
	NPS	Net Promoter Score
	NR	New Radio
	NS	Network Services
	NSA	Non-Standalone
	NSD	NS Descriptor
	NSO	Network Service Orchestrator

O	**OCA**	Open Connect Appliance
	OCF	Open Connectivity Foundation
	OEM	Original Equipment Manufacturer
	OLT	Optical Line Termination
	ONAP	Open Network Automation Platform
	ONF	Open Networking Foundation
	OPC	Open Platform Communications Foundation
	OPC-UA	OPC-Unified Architecture
	opex	Operational Expenditures
	OR	Object Recognition
	OSM	Open Source MANO
	OSS	Operation Support System
	OSU	Online Sign-Up
	OTA	Over-the-Air
	OTT	Over-the-Top
	OVS	Open Virtual Switch
	OWASP	Open Web Application Security Project

P	**P2P**	Peer-to-Peer
	PC	Personal Computer
	PDM	Product Data Management
	PHY	Physical Layer
	PIN	Personal Identification Number
	PKI	Public Key Infrastructure
	PL	Public Lighting
	PLC	Programmable Logic Controller
	PLM	Product Lifecycle Management
	PLMN	Public Land Mobile Network

	PNI	Private Network Interconnection
	PoC	Proof-of-Concept
	PON	Passive Optical Network
	PoP	Point-of-Presence
	PPV	Pay-per-view
	PR	PlayReady
	PSP	PlaySation Portable
	PV	Photovoltaic Production
Q	**QAM**	Quadrature Amplitude Modulation
	QoE	Quality of Experience
	QoS	Quality of Service
	QR	Quick Response
	QREN	Quadro de Referência Estratégico Nacional
R	**R&D**	Research and Development
	R&D+I	Research, Development and Innovation
	RAN	Radio Access Network
	RDK	Reference Design Kit
	RFID	Radio-Frequency Identification
	RMS	Rich Media Server
	RO	Resource Orchestrator
	RS-232	standard for serial communication transmission of data
	RS-485	standard for serial communications systems
	RSA	Rivest, Shamir and Adleman
	RSP	Remote SIM Provision
	RSPG	Radio Spectrum Policy Group
	RT	Round Trip
S	**S3**	Simple Storage Service
	SA	Standalone
	SaaS	Software as a Service
	SD	Standard Definition
	SDN	Software Defined Networks
	SDO	Standards Development Organization
	SEMS	Sustainable Energy Management System
	SEPS	Sustainable Energy Planning System
	SFC	Service Function Chaining
	SFF	Service Function Forwarder
	SFR	French telecommunications company from Altice Group
	SFI	Settlement-Free Interconnection
	S-GW	Serving GateWay
	SHA-256	Secure Hash Algorithm with 32-bit words
	SIM	Subscriber Identity Module
	SLA	Service Level Agreement
	SM-DP	Subscription Manager – Data Preparation
	SM-DP+	Subscription Manager Data Preparation - enhanced compared to the SM-DP
	SM-DS	Subscription Manager - Discovery Service
	SME	Small and Medium Enterprises
	SM-SR	Subscription Manager – Secure Routing
	SNMP	Simple Network Management Protocol
	SoC	Separation of Concerns
	SON	Self Organizing Networks
	SR	Service Router
	SSD	Supplementary Secure Domain
	SSO	Single-Sign On
	STB	Set-Top Box
	STEEP	Social-Technological-Economical-Environmental-Political
	STT	speech to text
	SVOD	Subscription Video on Demand
T	**T-Labs**	Telekom Innovation Laboratories from Deutsche Telekom
	T-Systems	German global IT services and consulting company from Deutsche Telekom
	TB	TeraBytes
	TCP	Transmission Control Protocol
	TEID	Tunnel Endpoint ID
	TEK	Traffic Encryption Key
	TELCO/ TELCOS	Telecommunication Operators
	TELECOM	Telecommunication
	TFTP	Trivial File Transfer Protocol
	TGE	Token Generation Events

	Thread	IPv6-based networking protocol for IoT
	TIP	The Innovation Pipeline
	TOF	Traffic Offloading Function
	TOSCA/ YAML	Topology and Orchestration Specification for Cloud Applications based on YAML
	TS	Technical Specification
	TSP	Telecommunications Service Providers
	TV	Television
U	**UDP**	User Datagram Protocol
	UE	User Equipment
	UEFA	Union of European Football Associations
	UHD	Ultra High Definition
	UI	User Interface
	UICC	Universal Integrated Circuit Card
	UK	United Kingdom
	UM	User Model
	UO	Unified Origin
	UP	User Plane
	URL	Uniform Resource Locator
	URLLC	Ultra Reliable and Low Latency Communications
	US	United States
	USA	United States of America
	USB	Universal Serial Bus
	USD	United States Dollar
	USIM	Universal Subscriber Identity Module
	USP	Urban Sharing Platform
	UTAD	University of Trás-os-Montes and Alto Douro
	UTP	Unshielded Twisted Pair
	UX	User Experience
V	**V&V**	Verification and Validation
	V2x	Vehicle to everything
	VA	Virtual Assistants
	VC	Venture Capital
	vEPC	Virtualized EPC
	VIM	Virtual Infrastructure Manager
	vIMS	Virtualized IMS
	VM	Virtual Machine
	VNF	Virtual Network Functions
	VNFD	VNF Descriptor
	VOD	Video On Demand
	vOLT	virtual OLT
	VR	Virtual Reality
	VSPP	Video Storage and Processing Platform
	VxLAN	Virtual eXtensible LAN
W	**WAN**	Wide Area Network
	WBA	Wireless Broadband Alliance
	WFA	Wi-Fi Alliance
	WG	Working Group
	Wi-Fi	IEEE 802.11x - Wireless Network (Wi-Fi Alliance)
	WLAN	Wireless LAN
	Wowza	unified streaming media server software by Wowza Media Systems
	WP5D	Working Party 5D
	WRC	World Radiocommunication Conference
X	**X.509**	ITU-T standard for public key certificates format
Y	**YAML**	YAML Ain't Markup Language
Z	**ZigBee**	IEEE 802.15.4-based open global standard for wireless technology for M2M networks
	Z-Wave	wireless communications protocol for home automation

www.alticelabs.com