# **Wireless Video**

## Guest Editors: Fulvio Babich, David R. Bull, and Jianfei Cai



**Wireless Video** 

## **Wireless Video**

Guest Editors: Fulvio Babich, David R. Bull, and Jianfei Cai

Copyright  $\circledcirc$  2008 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in volume 2008 of "EURASIP Journal on Advances in Signal Processing." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **Editor-in-Chief**

Phillip Regalia, Institut National des Télécommunications, France

### **Associate Editors**

A. Enis Çetin, Turkey Kenneth E. Barner, USA Richard J. Barton, USA Kostas Berberidis, Greece J. C. M. Bermudez, Brazil Jonathon Chambers, UK Liang-Gee Chen, Taiwan Huaiyu Dai, USA Satya Dharanipragada, USA Florent Dupont, France Frank Ehlers, Italy S. Gannot, Israel Fulvio Gini, Italy M. Greco, Italy Irene Y. H. Gu, Sweden Fredrik Gustafsson, Sweden Ulrich Heute, Germany Jiri Jan, Czech Republic Magnus Jansson, Sweden Sudharman K. Jayaweera, USA

Søren Holdt Jensen, Denmark Mark Kahrs, USA Moon Gi Kang, South Korea W. Kellermann, Germany Joerg Kliewer, USA Lisimachos P. Kondi, Greece Alex Chichung Kot, Singapore C.-C. Jay Kuo, USA Tan Lee, China Geert Leus, The Netherlands T.-H. Li, USA Mark Liao, Taiwan Y.-P. Lin, Taiwan S. Makino, Japan Stephen Marshall, UK C. F. Mecklenbräuker, Austria Gloria Menegaz, Italy Ricardo Merched, Brazil Marc Moonen, Belgium Vitor Heloiz Nascimento, Brazil Sven Erik Nordholm, Australia Antonio Ortega, USA D. O'Shaughnessy, Canada Bjorn Ottersten, Sweden Wilfried Philips, Belgium Aggelos Pikrakis, Greece Ioannis Psaromiligkos, Canada Markus Rupp, Austria William Allan Sandham, UK B. Sankur, Turkey Dirk Slock, France Y.-P. Tan, Singapore George S. Tombras, Greece Dimitrios Tzovaras, Greece Jacques G. Verly, Belgium Bernhard Wess, Austria Jar-Ferr Kevin Yang, Taiwan Azzedine Zerguine, Saudi Arabia A. M. Zoubir, Australia

### Contents

**Wireless Video**, Fulvio Babich, David R. Bull, and Jianfei Cai Volume 2008, Article ID 749159, 2 pages

**Distortion-Based Link Adaptation for Wireless Video Transmission**, Pierre Ferré, James Chung-How, David Bull, and Andrew Nix Volume 2008, Article ID 253706, 17 pages

A Cross-Layer Approach for Maximizing Visual Entropy Using Closed-Loop Downlink MIMO, Hyungkeuk Lee, Sungho Jeon, and Sanghoon Lee Volume 2008, Article ID 864606, 14 pages

Joint Video Summarization and Transmission Adaptation for Energy-Efficient Wireless Video Streaming, Zhu Li, Fan Zhai, and Aggelos K. Katsaggelos Volume 2008, Article ID 657032, 11 pages

**Optimal JPWL Forward Error Correction Rate Allocation for Robust JPEG 2000 Images and Video Streaming over Mobile Ad Hoc Networks**, Max Agueh, Jean-François Diouris, Magaye Diop, François-Olivier Devaux, Christophe De Vleeschouwer, and Benoit Macq Volume 2008, Article ID 192984, 13 pages

Scalable and Media Aware Adaptive Video Streaming over Wireless Networks, Nicolas Tizon and Béatrice Pesquet-Popescu Volume 2008, Article ID 218046, 11 pages

**Fast and Accurate Video PQoS Estimation over Wireless Networks**, Pasquale Pace and Emanuele Viterbo Volume 2008, Article ID 548741, 10 pages

**Objectives for New Error Criteria for Mobile Broadcasting of Streaming Audiovisual Services**, Heidi Himmanen, Miska M. Hannuksela, Teppo Kurki, and Jouni Isoaho Volume 2008, Article ID 518219, 12 pages

A Method to Estimate the Horizontal Handover Decision Effect on Indoor Wireless Conversational Video Quality, Alfonso Fernandez Duran, Raquel Perez Leal, and Jose I. Alonso Volume 2008, Article ID 370524, 15 pages

**Protection of Video Packets over a Wireless Rayleigh Fading Link: FEC versus ARQ**, Julie Neckebroek, Frederik Vanhaverbeke, Danny De Vleeschauwer, and Marc Moeneclaey Volume 2008, Article ID 852697, 15 pages

A Transparent Loss Recovery Scheme Using Packet Redirection for Wireless Video Transmissions, Chi-Huang Shih, Ce-Kuen Shieh, and Wen-Shyang Hwang Volume 2008, Article ID 437128, 15 pages

**Power-Constrained Fuzzy Logic Control of Video Streaming over a Wireless Interconnect**, Rouzbeh Razavi, Martin Fleury, and Mohammed Ghanbari Volume 2008, Article ID 560749, 14 pages An Adaptive Fair-Distributed Scheduling Algorithm to Guarantee QoS for Both VBR and CBR Video Traffics on IEEE 802.11e WLANs, Saeid Montazeri, Mahmood Fathy, and Reza Berangi Volume 2008, Article ID 264790, 12 pages

## Editorial Wireless Video

#### Fulvio Babich,<sup>1</sup> David R. Bull,<sup>2</sup> and Jianfei Cai<sup>3</sup>

<sup>1</sup>Dipartimento di Elettrotecnica, Elettronica ed Informatica, Università degli Studi di Trieste, 34127 Trieste, Italy

<sup>2</sup> Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, UK

<sup>3</sup> School of Computer Engineering, Nanyang Technological University, Singapore 639798

Correspondence should be addressed to Fulvio Babich, babich@units.it

Received 1 July 2008; Accepted 1 July 2008

Copyright © 2008 Fulvio Babich et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile communications, the Internet, and other emerging consumer technologies continue to have a major impact on our personal and business lives. Increasing numbers of media centric devices now have integrated wireless functionalities, and video-based services are becoming increasingly important. However, longer-term adoption of wireless video will hinge on a number of issues: the business model (does it make money?), the content (do we want to watch it?), the terminal capability (is the display bright enough and the battery life long enough?) and, not least, accessibility and visual quality (is it watchable?). This critical latter area is the topic of this special issue.

Because of its large bandwidth requirements, coupled with increased demand, video will become the dominant and most critical form of traffic in and beyond 3G/4G wireless systems. Reliable digital video transmission over wireless connections is widely acknowledged as challenging, given the hostile communication environment which is characterised by unpredictable connection quality, variable delay, significant error rates, limited available bandwidth, and severe energy constraints. These are compounded by the very nature of video information which, when compressed, is inherently sensitive to bit and packet losses.

When developing a reliable video transmission system for operation over a wireless network, many varied and interacting technical problems must be addressed, some of which may be application or standard specific. Many of these issues are addressed in the 12 papers contained in this special issue. The papers broadly fit into three categories: (i) error-resilient video and cross-layer optimisation, (ii) quality assessment, and (iii) the impact of the wireless channel and network. Due to the nature of the subject, however, several of the papers cut across more than one category.

The first 5 papers deal with error-resilience and crosslayer optimisation. The first paper by Pierre Ferré et al. addresses the issue of cross-layer optimisation for enhanced quality transmission over wireless LANs. A novel link adaptation scheme is presented that improves the quality of service (QoS). Rather than maximising the error-free throughput, this minimises the video distortion of the received sequence enabling the system to select the link speed which offers the lowest distortion and to adapt to varying channel conditions. The second paper by Hyungkeuk Lee et al. presents a crosslayer approach to improve video quality in an MIMO system. This is based on unequal error protection, where the coding strength is dictated by visual importance. The paper by Zhu Li et al. focuses on the issue of video transmission over a severely impaired, bandwidth-limited channel, taking into account battery life. The approach is based on the joint optimisation of video summarisation, coding, modulation, and packetisation, demonstrating substantial advantages under these constraints.

Scalability is an important codec characteristic which can be exploited to define new scheduling and prioritisation algorithms for the efficient delivery of time-sensitive video traffic. The paper by Max Agueh et al. presents an optimised FEC scheme which is JPEG2000 compliant and which provides an important step towards providing QoS guarantees in JPEG2000-based wireless multimedia systems. Following this, the paper by Nicolas Tizon and Béatrice Pesquet-Popescu exploits the properties of H.264/SVC in the context of a wireless network. Their approach combines temporal and SNR scalability features with an intelligent packet scheduling strategy to achieve substantial quality gains over convention approaches.

The next two papers deal with the crucial issue of picture quality. The contribution by Pasquale Pace and Emanuele Viterbo proposes a method for assessing the perceived quality of streaming media taking into account both loss and bandwidth. Based on VQM, the approach correlates well with subjective assessments and can offer real-time quality assessment and adaptation. From a practical viewpoint, the second paper by Heidi Himmanen et al. addresses the assessment of video quality in the context of streaming DVB-H services to mobile handhelds. Their results demonstrate the shortcomings of existing approaches and help to lay the foundations for future objective criteria.

The final five papers focus on specific aspects of the wireless physical layer, MAC, or network and how these impact video transmission. The first contribution by Alfonso Fernandez Duran et al. deals specifically with low-latency video services, addressing the impact of wireless handover on latency. The paper demonstrates that improved performance can be obtained through the use of alternative decision thresholds. Julie Neckebroek et al. consider the merits of FEC an ARQ in the context of a wireless Rayleigh fading link. They analyse the diversity gain offered by FEC and ARQ in terms of fading parameters, latency, and transmission overhead and apply their results to the case of an indoor 60GHz HDTV link. The paper by Chi-Huang Shih et al. introduces a transparent loss recovery scheme based on a transparent end-to-end QoS mechanism and an instantaneous framelevel FEC allocation technique, which provides near optimal performance with low delay. Rouzbeh Razavi et al. address the issue of power constraints in a Bluetooth network. They propose fuzzy logic as a means of controlling ARQ in the context of packet delay deadlines and buffer management. Finally, Saeid Montazeri et al. propose a new distributed QoS (MAC scheduling) scheme for WLANs which is capable of dealing with both CBR and VBR traffics in terms of delay and throughput.

#### ACKNOWLEDGMENTS

The guest editors hope you enjoy this special issue and would like to thank all the authors who have contributed to it. In addition, special thanks go to all the reviewers who have donated their valuable time to inform the decision process and to provide constructive feedback to the authors.

> Fulvio Babich David R. Bull Jianfei Cai

## Research Article Distortion-Based Link Adaptation for Wireless Video Transmission

#### Pierre Ferré,<sup>1</sup> James Chung-How,<sup>2</sup> David Bull,<sup>1</sup> and Andrew Nix<sup>1</sup>

<sup>1</sup> Centre for Communications Research, University of Bristol, Woodland Road, Bristol BS8 1UB, UK <sup>2</sup> ProVision Communication Technologies Limited, 3 Chapel Way, St. Anne's, Bristol BS4 4EU, UK

Correspondence should be addressed to Pierre Ferré, pierre.ferre@bristol.ac.uk

Received 15 October 2007; Accepted 10 March 2008

Recommended by F. Babich

Wireless local area networks (WLANs) such as IEEE 802.11a/g utilise numerous transmission modes, each providing different throughputs and reliability levels. Most link adaptation algorithms proposed in the literature (i) maximise the error-free data throughput, (ii) do not take into account the content of the data stream, and (iii) rely strongly on the use of ARQ. Low-latency applications, such as real-time video transmission, do not permit large numbers of retransmission. In this paper, a novel link adaptation scheme is presented that improves the quality of service (QoS) for video transmission. Rather than maximising the error-free throughput, our scheme minimises the video distortion of the received sequence. With the use of simple and local rate distortion measures and end-to-end distortion models at the video encoder, the proposed scheme estimates the received video distortion at the current transmission rate, as well as on the adjacent lower and higher rates. This allows the system to select the link-speed which offers the lowest distortion and to adapt to the channel conditions. Simulation results are presented using the MPEG-4/AVC H.264 video compression standard over IEEE 802.11g. The results show that the proposed system closely follows the optimum theoretic solution.

Copyright © 2008 Pierre Ferré et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

#### 1. INTRODUCTION

Low-latency video transmission is highly demanding in terms of the performance of all layers in the protocol stack. Over the last decade, research has mainly focused on enhancements to each individual layer without considering cross-layer interactions. Adapting the source coding according to the channel and network conditions (and vice *versa*) [1] via the cross-layer exchange of information has only recently been investigated. In [2, 3], van der Schaar et al. develop a cross-layer optimisation that combines application layer forward error correction (FEC), adaptive medium access control (MAC) retransmission and adaptive packetisation for video transmission over an IEEE 802.11b network. In [4], the authors discuss the challenges and principles of cross-layer optimised multimedia transmission. The choice of optimal modulation using Application/MAC/PHY interactions for video over IEEE 802.11b [5] is discussed as well as the choice of modulation scheme for optimal power consumption. Moreover, the authors stress the fact that an optimal solution for throughput may not be appropriate for

multimedia transmission. In [6], Setton et al. detail the basis of a cross-layer framework where packet size is dynamically adapted for a given link layer and channel condition. For a given packet length, the proposed scheme optimises the link layer parameters, such as the constellation and the symbol rate, in order to optimise the throughput. In [7, 8], the authors develop a hybrid link adaptation mechanism, combining different link adaptation techniques and using a cross-layering signalling system aimed at improving the received video quality. In [9], a cross-layer architecture is developed for MPEG-4/AVC H.264 [10] video over the IEEE 802.11e [11] MAC layer by assigning priority values to network abstraction layer (NAL) units that are then converted into priority accesses, specific to the MAC layer. However, with the exception of [3, 4, 7], adaptive link and MAC layer techniques, involving coding rate and modulation adaptation, are rarely considered in the design of cross-layer systems.

This paper investigates a link adaptation mechanism appropriate for the delivery of low-latency real-time video without relying on retransmission. Distortion models are



FIGURE 1: IEEE 802.11a/g PER performance, ETSI, BRAN Channel A [14], 825 byte packets.

developed and simulations are performed in order to evaluate the proposed scheme. The algorithm presented uses cross-layer exchange of information and is designed to optimise perceptual video quality (by minimising the perceived distortion) at the receiver. The paper is organised as follows. Section 2 presents the principles of link adaptation in IEEE 802.11 WLANs and describes the existing algorithms. The models used for the estimation of the distortion are described and validated in Section 3. Section 4 details the proposed link adaptation algorithms, and results are presented in Section 5. Finally, Section 6 concludes the paper.

#### 2. LINK ADAPTATION IN IEEE 802.11 WLANs

#### 2.1. IEEE 802.11a/g PHY and MAC

The PHY layers of COFDM-based WLANs at 2.4 GHz and 5 GHz, such as IEEE 802.11g [12] and IEEE 802.11a [13], respectively, offer numerous coding rates and modulation schemes, each providing different throughputs and reliability levels. Table 1 summarises the different link-speeds (commonly called operating modes) available for the IEEE 802.11a/g PHY layers. These range from BPSK 1/2 rate (mode 1) which provides a nominal bit rate of 6 Mbps, to 64 QAM 3/4 rate (mode 7), with a nominal bit rate of 54 Mbps. The BPSK 1/2 rate mode provides a more reliable transmission link than the 64 QAM 3/4 rate mode for a given received power level. Figure 1 shows the packet error rate (PER) performance versus power level (carrierto-noise ratio (C/N) for the 7 link-speeds available in IEEE 802.11a/g with a PHY packet length of 825 bytes (selected as a compromise between PHY PER performance and MAC layer throughput). Since the PER performance varies considerably between modes, the choice of operating link-speed is crucial to system performance. It should be noted that operating modes and link-speeds are equivalent and, in the remainder of this paper, both terms are used interchangeably.

Due to the range of operating modes available at the PHY layer, the ability for a system to adapt to the fluctuations of the environment (mobility, interference, and congestion) is vital to optimise overall performance. This ability to change link-speeds is used to control the reliability of the system and provides the radio with the ability to switch to a better configuration to improve the QoS of the transmission. Many parameters can be varied at the MAC and PHY level; examples include the maximum number of MAC level retries (or automatic repeat requests (ARQ)), the packet size, the operating mode (modulation, coding rate, link-speed), and the type and number of antennas. Neither the IEEE 802.11 MAC [15] nor the IEEE 802.11a/g standards specifies an algorithm for dynamic rate switching. The IEEE 802.11 MAC only defines rules for the mode selection of the management frames and declares dynamic rate selection for user data beyond the scope of the specifications [8, 15, 16]. It is therefore left to manufacturers to implement their own switching algorithms and metrics, examples of these include throughput, PER or delay.

#### 2.2. Existing link adaptation algorithms and related work

A simple link adaptation algorithm can be based on statistics about the transmitted data. Such schemes are known as *Statistics-based automatic rate control* algorithms [7, 8, 16]. These aim to provide the highest throughput [17, 18] since the statistics are directly related to user-level throughput. Other techniques use direct measurement of the link conditions, based for example on power levels which are closely related to the PER, and therefore to the throughput [7, 8].

#### 2.2.1. Statistics-based control

- (i) Throughput-based control: in these algorithms, a constant (small) fraction of data (up to 10%) is sent at two adjacent link-speeds (lower and higher than the current rate). At the end of a decision window, the transmitter computes the different throughputs and a switch is made to the rate that provides the highest throughput. In order to have meaningful statistics, the decision window must be sufficiently long (approximately one second [7, 8]).
- (ii) PER-based control: in these algorithms, the PER of the transmitted data is used to select the link-speed. The PER can be determined by counting the ACKs of the IEEE 802.11 MAC frame received at the transmitter during a sliding decision window (a missing ACK means that the corresponding packet has not been received correctly). This approach was not designed for video transmission, and optimises the PER to achieve an improved throughput. It does not take into account the nature of the content and its timebounded requirements.
- (iii) *Retry-based control*: in these algorithms, the decision metric used is the number of failed ARQs. If a transmission is unsuccessful after a certain number of

TABLE 1: Mode-dependent parameters for IEEE 802.11a/g.

Operating mode	Modulation	Coding rate	Link-Speed in Mbps	Bit rate ratio with mode 1
1	BPSK	1/2	6	1
2	BPSK	3/4	9	3/2
3	QPSK	1/2	12	2
4	QPSK	3/4	18	3
5	16 QAM	1/2	24	4
6	16 QAM	3/4	36	6
7	64 QAM	3/4	54	9

retries,  $N_{\text{fail}}$ , the link-speed is downscaled. Similarly, upscaling would occur after a certain number of successful contiguous transmissions,  $N_{\text{success}}$  [19]. This method offers a very short response time to channel changes. Upscaling can also be implemented with a PER-based control scheme using a decision window. This has been developed under the name of AutoRate Fall Back (ARF) [20, 21] and has been designed to optimise the application throughput [19].

#### 2.2.2. SNR-based control

In this method, the carrier-to-noise ratio (C/N), also known as the signal-to-noise ratio (SNR), is used to determine the transmission rate. The value of C/N is directly related to the PER. The throughput at the PHY layer can be expressed as a function of the PER and can be estimated as in [22–24]:

$$\Gamma hroughput = R \times (1 - PER), \tag{1}$$

where *R* is the operating link-speed (or nominal bit rate) (see Table 1). Link adaptation based on SNR/throughput is presented in Figure 2 for a MAC packet length of 825 bytes. The crossing points of the curves define the switching points (in terms of C/N) at which the system should up or downscale. A simple SNR-based algorithm would employ a look-up table (made available at the MAC) to obtain the best throughput for a given C/N [25]. These tables could theoretically be generated off-line for different packet lengths for all modes, C/Ns and different channel conditions. It should be noted that this assumes that ARQ is used for retransmitting packets until the packet is received correctly, or the maximum number of retries is reached (whichever comes first). Data are therefore received error-free but delays are incurred and the nature of the data is not taken into account.

#### 2.2.3. Other rate adaptation algorithms

Several rate adaptation algorithms have been presented in the literature. A selection of these is presented here. A good review of link adaptation design guidelines can be found in [26], where the authors compare the merits of the more common algorithms to derive a mechanism overcoming their disadvantages. In [27], the authors develop



FIGURE 2: Link adaptation based on throughput, IEEE 802.11a/g, 825 byte packets.

the minimum energy transmission strategy (MiSer) scheme, which minimises the communication energy consumption by combining the transport power control with the PHY rate adaptation. In [28], the receiver-based autorate (R-BAR) protocol is presented which optimises the application throughput [19], where the choice of transmission rate is made at the receiver based on its own stored statistics [21]. The information on the chosen rate is then transferred back to the transmitter via the CTS frame of the handshaking RTS/CTS. In [29, 30], the authors develop a hybrid automatic rate controller, combining a throughput-based rate controller with an SNR-based approach. By dynamically adjusting RSSI-look up tables, the algorithm selects the most appropriate rate. This scheme aims at improving throughput as well as reducing delay and PER, but is also able to adjust the transmitted video rate. A hardware solution is discussed in [7], together with video results. In [31], the authors derived an algorithm which allows differentiating packet loss due to channel errors from packet collisions. Using the RTS frame of IEEE 802.11 in an adaptive manner, the proposed system is more likely to make the correct rate adaptation. Variations of the above algorithms can be found in many papers, among which [25, 32-35] are notable.

Almost all the reported link adaptation algorithms have been designed to provide throughput and/or PER performance improvements [18] and/or to reduce the power consumption. They do not take into account the nature of the transmitted data or the low-delay requirements common to real-time video applications. They strongly rely on the use of retransmission and do not consider transmission delays. Moreover, in the case of multimedia transmission, they also do not optimise the perceived video quality [4].

#### 2.3. Motivation

In our previous work [17, 36], we have shown that existing algorithms are generally not suitable for low-latency video applications as (i) they do not take into account the nature of the transmitted data, and (ii) they are primarily designed to provide the highest throughput without regard for delay and retransmission. For video transmission where a strong reliance on ARQ is not desirable, a completely error-free communication is not essential when robust video compression techniques are applied. For example, it is possible to obtain an improved decoded video quality using a higher link-speed but with some degree of error, rather than an error-free video stream at a lower bitrate (using a lower link-speed). This is demonstrated in Figure 3 for the *foreman* sequence (average peak-to-peak signal-to-noise ratio (PSNR) over the whole sequence is shown here) for the case with no ARQ. Each mode can carry one video bit rate and, hence, higher modes support better video quality if the PER is sufficiently low. The overall quality of the received video sequence depends on a tradeoff between video bit-rate and error rate, as shown in Figure 4. For a given C/N of 18 dB, mode 1 provides error-free transmission at low video bit rates (700 kbps with a peak signal-to-noise ratio (PSNR) of 37.07 dB), whereas mode 5 provides a transmission with a PER of  $10^{-2}$  with a higher video bit rate (4235 kbps). However, Figure 4(b) shows better resolution and presents a better PSNR (44.85 dB) than Figure 4(a) (37.07 dB). Impairments due to errors are insignificant and can not be noticed visually.

Whenever the MAC layer adapts its link-speed, the application layer also adapts its encoding rate, based on the following two assumptions:

- (i) the ratios between the bit rates carried on each mode follow the ratios of the link-speeds available at the PHY layer for each mode, as shown in the last column of Table 1. In this way, similar PHY resources are used for each link-speed;
- (ii) the maximum size of the video packet generated at the encoder is not modified. A nonadaptive packetsize assumption is the most realistic case for such a system.

Therefore, if mode 1 is used to stream video at 500 kbps, modes 2, 3, 4, 5, 6, and 7 will carry video encoded at 750, 1000, 1500, 2000, 3000, and 4500 kbps, respectively. As the C/N increases, changing to higher link-speeds with a



FIGURE 3: Video quality-based algorithm, *foreman*, NAL unit max size: 750 bytes.

higher bit rate provides a better PSNR. For example, the best-video quality is obtained with QPSK 1/2 rate (mode 3) with 1000 kbps at a C/N of 17 dB, with some degree of error, whereas BPSK 1/2 rate with 500 kbps is error-free. A natural and empirical switching point would therefore be based on PSNR; effectively selecting the link-speed with the highest PSNR at any time and for any C/N level. However, in a realistic scenario, the decoder cannot derive PSNR because it does not have access to the original video reference. Moreover, PSNR performance depends on the content, the video bit rate, the concealment algorithm, and the packet length (amongst others).

A switching scheme using PER thresholds was presented by the authors in [17]. Comparisons of this approach with existing throughput-based solutions were made. The principle is shown in Figure 5 where it can be seen that switching occurs at lower PHY PERs for the video qualitybased algorithm. In [17], it was shown that parameters such as packet size, video rate, and content had a strong influence on the PER thresholds. A rigorous derivation of the PER thresholds was therefore found difficult to establish, and a practical design could not be proposed.

#### 2.4. Proposed approach

Building on the preliminary work in [17], this paper investigates a rigorous switching scheme based on the received video distortion. The distortion measured here is to the mean square error (MSE) between the received and original pixels. This includes the encoding distortion (due to the coding, transform, and motion compensation operation of the encoder) as well as the end-to-end distortion (due to error propagation and error concealment). The



(a) Mode 1, 700 kb, PER = 0, PSNR = 37.07 dB (b) Mode 5, 4235 kbps, PER = 0.04, PSNR = 44.85 dB

FIGURE 4: Foreman sequence, frame 30, C/N = 18 dB.



FIGURE 5: Switching points comparison, foreman.

same assumptions remain, that is, the ratio between the bit rates carried on each mode follows the ratio of the link-speeds available at the PHY layer for each mode; and the maximum size of the video packet generated at the encoder is not modified. Rather than using PSNR as a switching metric, the new scheme presented in this paper uses an estimate of the video distortion. The decision to switch from one link-speed to another is made upon the distortion experienced on the current mode, as well as the distortion on adjacent modes. For a given channel condition, the mode offering the lowest distortion, that is, the best video quality, is selected, as shown in Figure 6 (the average distortion over the whole sequence is shown here). Clearly, without a reference, the end-to-end distortions can not be computed at the transmitter and need to be estimated. A simple model to estimate the distortion at the current mode and at the two adjacent has been developed and is presented in the next section. The proposed approach operates on a group of pictures (GOP) basis, where distortions are estimated and switching decisions are made for each GOP.

#### 3. VIDEO TRANSMISSION MODEL DESCRIPTION

To enable mode switching based on distortion we need to estimate (i) the distortion of the received sequence transmitted at the current rate, under the given channel conditions, and (ii) the distortions of the received sequence if transmitted at lower and higher rates, under their corresponding channel conditions. To do so, we need to model (i) the rate distortion curve of the sequence; and (ii) an endto-end distortion. The following discussion is based on the H.264 standard [10] which is used throughout the paper.

#### 3.1. Empirical rate distortion model

Several accurate RD models have been presented in the literature [37–39]. However, these require trial encodings in order to determine sequence-dependent parameters (and hence cannot be used for practical systems), or they are aimed at advanced rate control operation [40]. In this section, we develop a simple empirical model aimed at deriving a local estimation of the rate distortion curve in



FIGURE 6: Distortion-based link adaptation, *foreman*, NAL unit max size: 750 bytes.

order to approximate the distortion at lower and higher rates, without relying on multiple encodings, that is, when only one point on the curve is known. The distortion used here is the MSE between the reconstructed and original pixels and is only due to the motion compensation, quantisation and transform operations of the encoder.

We first assume that a GOP has been encoded at the current rate. The actual average coding distortion of the GOP is therefore available, and we estimate the distortion due to coding for the sequence encoded at higher and lower rates. As stated in [41], in H.264, an increase of 6 in the quantisation parameter (QP) approximately halves the bit rate (equivalent to a decrease of 1 in the log<sub>2</sub> bit rate). A simple linear relationship between the QP and the log<sub>2</sub> of the bit rate can be adopted. As stated in [42], the quantisation design of H.264 allows a local and linear relationship between PSNR and the step-size control parameter QP. This can be expressed mathematically as

$$log_2(R) = a \times QP + b,$$
  
PSNR = c × QP + d, (2)

which can be rewritten as

$$PSNR = \frac{c}{a} \times \log_2(R) + \left(d - \frac{bc}{a}\right).$$
(3)

This linear relationship between PSNR and the base-two of the logarithm of the bit rate has been verified by plotting the actual PSNR versus  $log_2(R)$  for all GOPs in the *table* (Figure 7(a)) and *coastguard* (Figure 7(b)) sequences. Similar curves have been obtained with other sequences and we can thus assume that the curves are locally linear, that is, three adjacent points are aligned.

To fully derive the parameters of this linear model, several parallel encodings would be needed, but this is not

practical. From the encoding of the current GOP, the current PSNR<sub>c</sub> (derived from the averaged MSE), the current rate  $R_c$  and the current average QP<sub>c</sub> are known. Using the fact that an increase of 6 in QP halves the bit rate, we derive a = -1/6. Moreover, empirical studies for CIF sequences (a similar constant can be obtained for sequences with others resolutions and formats) have shown that trial encodings with a QP of 6 leads to an almost constant luminance PSNR of 55.68 dB (±0.3 dB) for *akiyo*, *coastguard*, *table*, and *foreman* sequences. We can now calculate the four parameters *a*, *b*, *c*, and *d* as

$$a = -\frac{1}{6},$$
  

$$b = \log_2(R_c) + \frac{QP_c}{6},$$
  

$$c = \frac{PSNR_c - 55.68}{QP_c - 6},$$
  

$$d = \frac{55.68 \times QP_c - 6 \times PSNR_c}{QP_c - 6}.$$
  
(4)

To validate this model, video sequences (akiyo, foreman, table, and coastguard) were encoded at the following rates 500 kbps, 750 kbps, 1000 kbps, 1500 kbps, 2000 kbps, 3000 kbps, and 4500 kbps. Figure 8(a) shows the estimation of PSNR for the GOP number 10 of the table sequences at 1000 and 2000 kbps (the GOP is encoded at 1500 kbps). It can be seen that the model follows a similar trend to the actual curve. However, because the reference point (QP = 6,  $PSNR = 55.68 \, dB$ ) may be distant from the current operating point, a mismatch can appear. We have found empirically that weighting the parameter *c* by a scalar dependent on the average QP improves the accuracy of the model. Figure 8(b) shows similar performance trends with the GOP number 15 of foreman encoded at 3000 kbps when used to estimate the PSNR at 2000 and 4500 kbps. Figure 9 shows a comparison between the actual and estimated MSE at the lower and higher rates for all the GOPs of table encoded at 1500 kbps and foreman encoded at 750 kbps. Tables 2 and 3 provide the mean and standard deviation of the estimation error calculated over the GOPs, between the actual MSE and the estimated MSEs, for each encoding rate of foreman and table, respectively. It can be seen that the mean error is smaller with the model with linear weighting (and it is below 10%). Similarly, the standard deviation of the error is smaller when linear weighting is applied and kept in the range from 1% to 9%. The proposed model employing weighting factors thus offers an acceptable local estimate of encoding distortions for the sequence at lower and higher bit rates.

The procedure to derive the distortion of the current GOP of a sequence as if it was encoded at the lower and higher local (adjacent) rates is summarised as follows.

- (i) Derive rate  $R_c$ , average QP<sub>c</sub>, average MSE<sub>c</sub> and PSNR<sub>c</sub> =  $10 \times \log_{10}(255 \times 255/\text{MSE}_c)$  from the encoding of the current GOP.
- (ii) Derive *a*, *b*, *c*, and *d* using (4).



FIGURE 7: PSNR versus  $\log_2$  (Bit rate) performance for 25 GOPs.

TABLE 2: Mean and standard deviation (calculated over the GOPs) of the estimation error (in percent) between the actual and the estimated MSE, *foreman*.

		Mean of the estimation error		Standard deviation of the estimation error		
		(percentage of difference)		(percentage of difference)		
Current encoding rate	Estimation rate	Linear model	Linear model with weighting	Linear model	Linear model with weighting	
500 kbps	750 kbps	18.2555	7.8208	7.0821	8.1238	
Current encoding rate 500 kbps 750 kbps 1000 kbps 1500 kbps 2000 kbps	500 kbps	25.7355	7.4049	10.7892	6.0400	
	1000 kbps	16.2241	6.3052	6.2538	3.7887	
1000 kbps	750 kbps	21.3207	7.1663	8.8395	4.5493	
1000 KDps	1500 kbps	22.3845	6.8882	5.2796	3.0656	
1500 kbps	1000 kbps	31.8273	8.8351	8.2769	4.1898	
1500 K0ps	2000 kbps	17.0562	5.6035	4.2309	2.5047	
2000 kbps	1500 kbps	21.2502	6.4256	e)         (percentage of difference)           th weighting         Linear model         Linear model with weighting           8         7.0821         8.1238           9         10.7892         6.0400           2         6.2538         3.7887           3         8.8395         4.5493           2         5.2796         3.0656           1         8.2769         4.1898           5         4.2309         2.5047           6         6.0921         2.9674           1         3.5749         2.7910           0         5.1767         3.0556           5         4.0193         3.8371           3         5.4758         3.2906	2.9674	
2000 KUPS	3000 kbps	21.6382	5.0351	3.5749	2.7910	
3000 kbps	2000 kbps	26.2032	4.8640	5.1767	3.0556	
	4500 kbps	14.5347	4.3805	4.0193	3.8371	
4500 kbps	3000 kbps	16.4630	4.0723	5.4758	3.2906	

TABLE 3: Mean and standard deviation (calculated over the GOPs) of the estimation error (in percent) between the actual and the estimated MSE, *table*.

		Mean of the percentage of difference		Standard deviation of the percentage of difference	
Current encoding rate	Estimation rate	Linear model	Linear model with weighting	Linear model	Linear model with weighting
500 kbps	750 kbps	14.4219	12.3402	8.2494	9.0454
750 kbit/s	500 kbps	19.7089	9.4528	12.6270	5.8535
7 50 KDI(75	1000 kbps	11.4824	4.9793	<u>4.9201</u> 6.2735	3.5082
1000 kbps	750 kbps	14.9569	4.1785	6.2735	2.7079
	1500 kbps	14.4776	9.9738	6.5595	7.1777
1500 kbps	1000 kbps	20.4458	6.6005	10.0650	5.1867
1500 корз	2000 kbps	14.6201	5.4923	ghting         Linear model         Linear model with weig           8.2494         9.0454           12.6270         5.8535           4.9201         3.5082           6.2735         2.7079           6.5595         7.1777           10.0650         5.1867           5.6605         3.3561           9.0542         4.4030           9.5719         5.7515           19.3450         8.7635           12.8395         5.0332           17.3489         4.9546	3.3561
2000 kbps	1500 kbps	20.1543	6.7503	9.0542	4.4030
2000 KDp3	3000 kbps	23.3229	10.9368	6.600510.06505.18675.49235.66053.35616.75039.05424.403010.93689.57195.751515.637919.34508.7635	5.7515
3000 kbps	2000 kbps	36.8940	15.6379	19.3450	8.7635
	4500 kbps	21.8986	14.6120	12.8395	5.0332
4500 kbps	3000 kbps	26.7938	13.5277	17.3489	4.9546





(a) *Table* encoded at 1500 kbps, GOP number = 10; estimation of the points for encoding at 1000 kbps and 2000 kbps







FIGURE 9: MSE comparison: actual MSE and estimated adjacent MSE.

- (iii) Derive PSNR<sub>l</sub> and PSNR<sub>h</sub> video quality using (2) with the corresponding lower and higher rates  $R_l$  and  $R_h$ , respectively.
- (iv) Compute MSE<sub>*l*</sub> and MSE<sub>*h*</sub> from PSNR<sub>*l*</sub> and PSNR<sub>*h*</sub>.

#### 3.2. End-to-end and transmission distortion model

To estimate the distortion of the received video, we use the end-to-end distortion model developed in [38, 43]. We limit

the study to only one reference frame; however the model remains valid with a larger number of reference frames. We consider the previous frame copy (PFC) concealment algorithm at the decoder, in which missing pixels due to packet loss during transmission are replaced by the colocated pixels in the previous reconstructed frame. We assume that the probability of a packet loss is  $p_c$  on the current rate. The current end-to-end distortion for pixel *i* of frame *n*, noted Dist<sub>e2e,c</sub>(*n*,*i*) accounts for (a) the error propagation from frame n - 1 to frame n,  $D_{EP}(n, i)$ ; and (b) the PFC error concealment,  $D_{EC}(n, i)$ . We therefore have

$$\operatorname{Dist}_{e^{2e,c}}(n,i) = (1-p_c) \times D_{\operatorname{EP}}(n,i) + p_c \times D_{\operatorname{EC}}(n,i). \quad (5)$$

Readers are referred to [38, 43] for full details on how  $D_{\text{EP}}(n, i)$  and  $D_{\text{EC}}(n, i)$  are derived. Assuming that a pixel *i* of frame *n* has been predicted from pixel *j* in frame n - 1,  $\text{Dist}_{e2e,c}(n, i)$  can be expressed as

$$Dist_{e^{2e,c}}(n,i) = (1 - p_c) \times Dist_{e^{2e,c}}(n - 1, j) + p_c \times (RMSE_c(n - 1, n, i) + Dist_{e^{2e,c}}(n - 1, i)).$$
(6)

 $\text{RMSE}_c(n - 1, n, i)$  is the MSE between reconstructed frames *n* and *n* - 1 at pixel location *i* at the current rate. If the pixel *i* belongs to an intra block, there is no distortion due to error propagation but only due to error concealment; and  $\text{Dist}_{e2e,c}(n, i)$  is rewritten as

$$Dist_{e2e,c}(n,i) = p_c \times (RMSE_c(n-1,n,i) + Dist_{e2e,c}(n-1,i)).$$
(7)

In order to compute the end-to-end distortion of the sequence transmitted at lower and higher adjacent rates,  $\text{Dist}_{e2e,l}(n, i)$  and  $\text{Dist}_{e2e,h}(n, i)$ , respectively, with a packet loss of  $p_l$  and  $p_h$ , respectively, we assume that the motion estimation is similar at all the rates and the difference in quality between the reconstructed sequences is only due to quantisation. Therefore, if pixel *i* in frame *n* is predicted from pixel *j* in frame n - 1 at the current rate, it will also be predicted from the same pixel *j* in frame n - 1 at lower and higher rates. The two distortions at lower and higher rates can then be expressed as

$$Dist_{e2e,l}(n,i) = (1 - p_l) \times Dist_{e2e,l}(n - 1, j) + p_l \\ \times (RMSE_l(n - 1, n, i) + Dist_{e2e,l}(n - 1, i)),$$
  
$$Dist_{e2e,h}(n,i) = (1 - p_h) \times Dist_{e2e,h}(n - 1, j) + p_h \\ \times (RMSE_h(n - 1, n, i) + Dist_{e2e,h}(n - 1, i)).$$
(8)

Dist<sub>e2e,l</sub> and Dist<sub>e2e,h</sub> only differ from Dist<sub>e2e,c</sub> by the packet loss and the impact of the concealment algorithm, that is, by  $\text{RMSE}_l(n - 1, n, i)$  and  $\text{RMSE}_h(n - 1, n, i)$ . If we consider the lower rate,  $\text{RMSE}_l(n - 1, n, i)$  is given by

DI COD (

$$RMSE_{l}(n, n - 1, i) = [i_{rec,l}(n) - i_{rec,l}(n - 1)]^{2} = [i_{rec,l}(n) - i_{rec,c}(n) + i_{rec,c}(n) - i_{rec,l}(n - 1) + i_{rec,c}(n - 1) - i_{rec,c}(n - 1)]^{2} = [(i_{rec,c}(n) - i_{rec,c}(n - 1)) + (i_{rec,l}(n) - i_{rec,c}(n)) - (i_{rec,l}(n - 1) - i_{rec,c}(n - 1))]^{2},$$
(9)

where  $i_{\text{rec},c}(n)$  and  $i_{\text{rec},l}(n)$  are the reconstructed pixels at location *i* from frame *n* at the current and lower rates, respectively. If we assume that the quality difference between the two rates is evenly spread along the frames of a GOP, the differences  $i_{\text{rec},l}(n) - i_{\text{rec},c}(n)$  and  $i_{\text{rec},l}(n-1) - i_{\text{rec},c}(n-1)$  are cancelled. Equation (9) can therefore be rewritten as

$$RMSE_{l}(n, n - 1, i) = \left[ \left( i_{rec,c}(n) - i_{rec,c}(n - 1) \right) \right]^{2}$$
$$= RMSE_{c}(n, n - 1, i)$$
(10)
$$= RMSE_{h}(n, n - 1, i).$$

The error concealment produces a similar contribution to the end-to-end distortion for the current, lower and higher rates. The overall average distortions for each GOP, including the encoding distortion due to quantisation as well as the end-to-end distortion due to error propagation and error concealment, for the lower, current and higher rates, can thus be estimated by

$$Dist_{l} = Dist_{e2e,l} + MSE_{l},$$
  

$$Dist_{c} = Dist_{e2e,c} + MSE_{c},$$
  

$$Dist_{h} = Dist_{e2e,h} + MSE_{h}.$$
  
(11)

The end-to-end distortion model has been fully validated in [38, 43]. Figure 10 confirms this by plotting a comparison between the estimated received distortions and the actual transmissions. Figure 10(a) shows the actual received distortion along the GOPs of *coastguard* encoded at 1500 kbps, with PER of 1%, against the estimated received distortion of coastguard when encoded at 1500 kbps (current rate), as well as with the estimated received distortion of the higher rate when encoded at 1000 kbps (from the lower rate) and of the lower rate when encoded at 2000 kbps (from the higher rate). Similar performance is shown in Figure 10(b) for table encoded at 3000 kbps with a PER of 0.1%. Figure 11 shows the estimated distortions on the current, lower and higher rates compared to the actually received distortions for a C/N of 23 and 22 dB for *coastguard* with the current mode being 5 and 4, respectively. From these figures, it can be seen that the local estimates from our proposed model closely follow the actual received distortion. It should be noted here that the derivation of more complex (and hence accurate) models would effectively provide better performance. However, this is not the primary aim of this paper, and we believe that the proposed models are suitable for our needs.

#### 4. PROPOSAL FOR IMPROVED VIDEO TRANSMISSION

#### 4.1. Algorithm

The proposed link adaptation scheme assumes that the ratios between the bit rates carried on each mode follow the ratios of the link-speeds available at the PHY layer for each mode. Moreover, it requires that the maximum size of the video packet generated at the encoder is not modified, so that a single PER versus C/N lookup table can be used, assuming a single channel type. It is aimed at low-latency video transmission, without reliance on ARQ. The proposed



FIGURE 10: Estimated received distortion along the GOPs with fixed PER.



FIGURE 11: Comparison estimated and actual distortion for different power levels.

algorithm allows dynamic mode switching at each GOP and operates as follows.

- (i) Encode the current GOP at the specified bit rate on the specified link-speed.
- (ii) Extract the average QP, average MSE, then the average PSNR and average rate *R* for the GOP.
- (iii) Extract the PER from lookup tables using the average received signal strength information (RSSI).
- (iv) Derive the estimated distortion at the current, lower and higher modes MSE<sub>c</sub>, MSE<sub>l</sub>, and MSE<sub>h</sub> as described in Section 3.1.

(v) Compare the distortions:

- if  $MSE_c < MSE_l$  and  $MSE_c < MSE_h$ : the distortion estimated on the current mode is the lowest; stay in the current mode;

- if  $MSE_l < MSE_c$  and  $MSE_l < MSE_h$ : the distortion estimated on the lower mode is the lowest; switch to the lower mode, at a lower rate;

- if  $MSE_h < MSE_c$  and  $MSE_h < MSE_l$ : the distortion estimated on the higher mode is the lowest; switch to the higher mode, at a higher rate.



FIGURE 12: Optimum distortion-based link adaptation, *foreman*, GOP number 8, *Set* (a).



FIGURE 13: Optimum distortion-based link adaptation, *coastguard*, GOP number 21, *Set* (b).

- (vi) Update the video bit rate at the application layer, update the link-speed at the link layer.
- (vii) Proceed to the next GOP and go back to (i).

#### 4.2. Design and issues

This algorithm is fully compliant with the IEEE 802.11a/b standard and could be implemented in a real system. Moreover, it could coexist with existing algorithms aimed



FIGURE 14: Optimum distortion-based link adaptation, *table*, GOP number 21, *Set* (c).

at other types of data and could be simply triggered either by a flag or by using access categories, similar to IEEE 802.11e [11] (e.g., packet classifiers are already used to select QoS mechanisms and service flows). The distortion estimation performed at the video encoder does not significantly increase the complexity since it only requires motion compensation, using the already available motion vectors. The change of video bit rate can be achieved either by dynamically changing the target rate in the rate controller (for real-time encoding), or by using a transcoder (for preencoded sequences). Alternatively, a scalable encoder can be employed, dynamically selecting the parts of the bitstream to transmit in order to adjust the bit rate to the bandwidth fluctuations resulting from changes in the link-speed.

The main design issue would be the communication between the application and the link layer. Prior to estimating the distortions, the application layer requires knowledge of the channel conditions from the link layer. Once the switching decision is made, the application layer needs to notify the link layer to update the link-speed accordingly. This exchange of information may be done with a cross-layer communication bus. It should be noted that the frequency of the switching decision can be extended to several GOPs if needed.

With this algorithm, the transmission mode and video bit rate of the current GOP are determined using the channel and video statistics of the previous GOP. A GOP size of 12 frames at 30 frames per second corresponds to a 400millisecond delay. Unless the sequence contains extremely high motion, or scene changes, the motion activity and the sequence content should not be affected by this delay. Moreover, it is reasonable to assume that the overall channel conditions are stable over 400 milliseconds. This value also provides good reactivity to channel changes. However, the



FIGURE 15: Mode and estimated distortion for *coastguard* encoded with C/N = 15 dB, Set (a).



FIGURE 16: Mode and estimated distortion for *foreman* encoded with C/N = 20 dB, Set (b).

estimated distortions are statistical and might therefore differ from the results of a single transmission.

#### 5. RESULTS

#### 5.1. Simulation conditions

A compliant 802.11a/g PHY-layer simulator developed at University of Bristol, meeting the conformance requirements specified in Annex A of [12, 13], has been used to recreate accurate bit and packet error performance [22, 24]. The simulator supports all the standardised operating modes and variable PHY-layer packet lengths. Moreover, it implements all the components of the PHY layer with all parameters configured in alignment to the standard and is capable of producing error performance at any C/N level. The channel model conforms to the ETSI-BRAN channel A specifications (non line-of-sight office environment), with an rms delay spread of 50 nanoseconds. Using our simulator, an accurate derivation of the PER performance curves and lookup tables, for a PHY packet length of 825 bytes, were produced (in order to fully analyse the proposed mechanism without the influence of others schemes, algorithms which

optimally choose packet lengths were not considered, and a simple packetisation with fixed packet length is used in this paper). We assume that packet losses due to collisions are negligible compared to losses due to channel errors.

Four video sequences (akiyo, foreman, table, and coastguard) at CIF resolution are encoded at 30 frames per second (fps) with our modified version of the H.264 reference software [44] (JM version 12.4). Three sets of video rates were considered: (a) from 125 to 1125 kbps, (b) from 250 to 2250 kbps, and (c) from 500 to 4500 kbps. The results presented here are however representative of lower (OCIF, subQCIF) and higher (4 CIF, SD) resolutions as well as lower and higher bit rates; and are used to illustrate the need for cross-layer optimisation and demonstrate the benefits of deploying the proposed system. The RTP format and a fixed maximum NAL unit size of 750 bytes (the 75 remaining bytes account for the RTP/UDP/IP/MAC headers) are chosen. Generated slices are encapsulated into UDP/IP packets. A GOP size of 12, FMO type 2 (dispersed) and one reference frame were used. At the decoder, lost macroblocks (MBs) are simply replaced by the collocated MBs in the previous frame (PFC concealment).



FIGURE 17: Mode and estimated distortion for *table* encoded with C/N = 21 dB, Set (c).



(b) Selected modes with the proposed algorithm

FIGURE 18: Mode selection comparison, *table*, GOP number 15, initial mode = 3, *Set* (a).



(b) Selected modes with the proposed algorithm

FIGURE 19: Mode selection comparison, *foreman*, GOP number 11, initial mode = 3, *Set* (b).



FIGURE 20: Mode selection comparison, *akiyo*, GOP number 9, initial mode = 6, *Set* (c).

#### 5.2. Optimum link adaptation

The sequences were encoded at fixed rates, using the three sets: *Set* (a) 125, 187.5, 250, 375, 500, 750, and 1125 kbps; *Set* (b) 250, 375, 500, 750, 1000, 1500, and 2250 kbps; and *Set* (c) 500, 750, 1000, 1500, 2000, 3000, and 4500 kbps (each of these is transmitted on each of the IEEE 802.11 WLAN modes) and transmitted off-line 50 times (for statistical purposes) over the IEEE 802.11 PHY layer for a wide range of fixed C/N power levels. For each sequence, for each GOP, and for each C/N, the average received distortion (MSE) is calculated and averaged over the 50 runs. This allows us to generate distortion performance curves which will constitute optimum link adaptation, where for each C/N the chosen operating mode is the mode with the lowest distortion. Figures 12, 13, and 14 show samples of the optimum link adaptation for GOP number 8 of *foreman* with *Set* (a), GOP

number 15 of *coastguard* with *Set* (b), and for GOP number 21 for *table* with *Set* (c), respectively.

By examining the PER curves in Figure 1, it can be seen that mode 2 (BPSK 3/4 rate) has worse performance than mode 3 (QPSK 1/2 rate), and that mode 4 (QPSK 3/4 rate) has a similar performance to mode 5 (16 QAM 1/2 rate). Moreover, both offer lower link-speeds (see Table 1). This explains why, using Figure 2 characterising the throughput under various conditions, modes 2 and 4 are never be used. This is also confirmed when examining the optimum link adaptation curves in Figures 12, 13, and 14, where modes 2 and 4 are similarly never used. As a consequence, BPSK 3/4 rate and QPSK 3/4 rate are no longer considered in the remainder of this paper.

#### 5.3. Behaviour of the proposed system

The sequences from Section 5.2 have been encoded with our encoder and the proposed cross-layer link adaptation mechanism. This allows the encoder to have knowledge of the C/N, which is in turn used to estimate the PER of the current mode and also of the adjacent modes. GOP distortion estimates were computed and the target bit rate and operating mode are updated as detailed in Section 4. It should be noted that, for a fixed C/N, the system behaviour over early GOPs will depend on the initial target and operating mode. This is illustrated in Figures 15, 16, and 17. Figure 15 compares the mode and distortion variations for coastguard encoded with a C/N of 15 dB with two different initial modes, with bit rates from Set (a). With mode 1 as the initial mode, the system upscales rapidly because of the favourable conditions, and then remains steady in mode 3. Whereas starting from mode 6, the system faces poor channel conditions, and needs to downscale to mode 5 and then to mode 3, where it remains. Similar conclusions can be drawn from Figure 16 with coastguard encoded with a C/N of 20 dB with rates in Set (b) and from Figure 17 with table encoded with a C/N of 21 dB with rates in Set (c). We also note that the selected mode adapts to channel conditions, but also to video content. For example, in Figure 17, for GOP number 3 to 11 (i.e., from frames 36 to 132) the camera zooms out in the table sequence. This part of the sequence is therefore less resilient to errors and the system automatically switches from mode 5 to 3; the sequence remains steady after GOP 11, where the system upscales to mode 5.

#### 5.4. Comparison with optimal link adaptation

This section compares the optimum modes with the modes selected by our algorithm, as well as the estimated and received distortions with the optimum ones. Figures 18, 19, and 20 compare the selected modes obtained for various C/N levels for GOP number 15 of *table* (*Set* a), GOP number 11 of *foreman* (*Set* b), and GOP number 9 of *akiyo* (*Set* c), respectively. It can be seen that the proposed mechanism offers very similar switching points compared to the optimum case. Similar mode switching curves were obtained with other GOP numbers at other C/N levels for the three rate sets.



FIGURE 21: Distortion comparison, *foreman*, GOP number 8, initial mode = 3, *Set* (a).



FIGURE 22: Distortion comparison, *table*, GOP number 15, initial mode = 3, *Set* (a).

The simulated curves were obtained by averaging over 50 runs for each video sequence encoded and for each C/N level. Figures 21, 22, 23, and 24 compare the optimum link adaptation distortion curves, with the estimated distortion from our system and with the simulated and received distortions, for rates from Sets (a), (b), and (c). First, it can be seen that the estimated and actual distortion levels are very similar, confirming the validity of the proposed model. Moreover, these curves smoothly follow the optimum case. Then, for a given C/N power level, the proposed system achieves the lowest video distortion, by adaptively choosing



FIGURE 23: Distortion comparison, *coastguard*, GOP number 15, initial mode = 3, *Set* (b).



FIGURE 24: Distortion comparison, *table*, GOP number 21, initial mode = 3, *Set* (c).

for each GOP the operating mode which minimises the overall distortion.

#### 6. CONCLUSIONS

In this paper, we have presented a novel link adaptation algorithm designed for low-latency video transmission over IEEE 802.11a/g without strong reliance on ARQ. Existing algorithms for link adaptation make extensive use of the retransmission mechanism at the MAC layer in order to improve the error-free data throughput without taking into account the bounded delay requirements of real-time video applications. Moreover, they do not incorporate the specific characteristics of video streams. Completely error-free communication is not essential if robust video compression techniques are used, and it is possible to obtain improved decoded video quality with a stream at a higher bit rate, using a higher link-speed, but with some degree of error rather than an error-free video stream at lower rate, using a lower link-speed. Based on these observations, a link adaptation mechanism minimising the overall transmission video distortion has been presented for low-latency video transmission.

Models were used to estimate the local rate distortion performance at the video encoder and to estimate the endto-end transmission distortion. These models were validated and shown to provide a reasonably accurate estimate of the video distortion. With the assumption that each operating mode carries a different bit rate, the proposed link adaptation uses the estimated overall distortion on the current operating mode, as well as on the lower and higher adjacent modes. For each GOP, the proposed algorithm effectively selects the mode that provides the lowest distortion. A crosslayer exchange of information is needed between the video encoder (at the application) and the link adapter (at the MAC layer).

The proposed system is extendable to other multirate systems such as WiMax and 3GPP LTE, which also support several link-speeds with different modulation and coding rates, each with different reliability levels. Future work will focus on the derivation of more sophisticated rate distortion source models and will also compare the proposed algorithm with known link adaptation techniques such as ARF, for a given received signal level trace. Validations of our approach will be performed using a real-time experimental platform.

#### ACKNOWLEDGMENTS

This work was partially funded by the UK TSB project VISUALISE and also by the EU FP6 project ASTRALS.

#### REFERENCES

- B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," *Wireless Communications* and Mobile Computing, vol. 2, no. 6, pp. 573–584, 2002.
- [2] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1752– 1763, 2003.
- [3] M. van der Schaar and D. S. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 185–197, 2007.
- [4] M. van der Schaar and S. N. Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 50–58, 2005.
- [5] "IEEE Standard 802.11b—part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer extension in the 2.4 GHz band," 1999.

- [6] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Crosslayer design of ad hoc networks for real-time video streaming," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 59–65, 2005.
- [7] I. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk, and H. Sips, "Optimized video streaming over IEEE 802.11 by cross layer signalling," *IEEE Communication Magazine*, vol. 44, no. 1, pp. 115–121, 2006.
- [8] I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips, "SNR-based rate control in WaveLAN," in *Proceedings of the* 10th Annual Conference of the Advanced School for Computing and Imaging (ASCI '04), Delft, The Netherlands, June 2004.
- [9] A. Kosentini, M. Naimi, and A. Guéroui, "Towards improvement of the H.264 video transmission over IEEE 802.11e through cross-layer architecture," *IEEE Communication Magazine*, vol. 44, no. 1, pp. 107–144, 2006.
- [10] Joint Video Team of ISO/IEC MPEG ITU-T VCEG, "ITU-T H.264—series H: audiovisual and multimedia aystems advanced video coding for generic audio visual services".
- [11] "IEEE Standard 802.11e; draft supplement to part 11: wireless medium access control (MAC) and physical layer (PHY) specifications: medium access control (MAC) enhancements for quality of services (QoS)," 2002.
- [12] "IEEE Standard 802.11g—part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: further high-speed physical layer in the 2.4 GHz band," d1.1 2001.
- [13] "IEEE Standard 802.11g; part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: highspeed physical layer in the 5 GHz band," d7.0, 1999.
- [14] J. Medbo and P. Schramm, "3ERI085B—channel models for HIPERLAN/2 in different indoor scenarios," ETSI EP BRAN, 1998.
- [15] "IEEE Std 802.11; part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," 1999.
- [16] I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips, "D3.16: application directed automatic 802.11 rate control," Tech. Rep., GigaMobile Project, Technische Universiteit Delft, Delft, The Netherlands, 2002.
- [17] P. Ferré, A. Doufexi, J. Chung-How, A. Nix, and D. Bull, "Video quality based link adaptation for low latency video transmission over WLANs," *Journal of Zhejiang University: Science A*, vol. 7, no. 5, pp. 847–856, 2006.
- [18] H. Zhu, M. Li, I. Chlantac, and B. Prabhakaran, "A survey of quality of service in IEEE 802.11e networks," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 6–14, 2004.
- [19] M. Lacage, H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," Tech. Rep., INRIA, Sophia Antipolis, France, 2004.
- [20] A. van der Vegt, "Auto rate fall back algorithm for IEEE 802.11a standard," Tech. Rep., HPC Group, Faculty of Physics and Astronomy, University of Utrecht, Utrecht, The Netherlands, 2002.
- [21] M. H. Manshaei, T. Turletti, and M. Krunz, "A mediaoriented transmission mode selection in 802.11 wireless LANs," in *Proceedings of the IEEE Wireless Communications* and Networking Conference (WCNC '04), vol. 2, pp. 1228– 1233, Atlanta, Ga, USA, March 2004.
- [22] A. Doufexi, S. Armour, M. Butler, A. Nix, and D. Bull, "A study of the performance of HIPERLAN/2 and IEEE 802.11a physical layers," in *Proceedings of the 53rd Vehicular Technology Conference (VTC '01)*, vol. 1, pp. 668–672, Rhodes, Greece, May 2001.

- [23] Z. Lin, G. Malmgren, and J. Torsner, "System performance analysis of link adaptation in HiperLAN type 2," in *Proceedings* of the 52nd Vehicular Technology Conference (VTC '00), vol. 4, pp. 1719–1725, Boston, Mass, USA, September 2000.
- [24] A. Doufexi, S. Armour, M. Butler, et al., "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 172–180, 2002.
- [25] D. Qiao, S. Choi, and K. G. Shin, "Goodput analysis and link adaptation for IEEE 802.11 a wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 1, no. 4, pp. 278–292, 2002.
- [26] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust rate adaptation for 802.11 wireless networks," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MOBICOM '06)*, vol. 2006, pp. 146–157, Los Angeles, Calif, USA, September 2006.
- [27] D. Qiao, S. Choi, A. Jain, and K. G. Shin, "MiSer: an optimal low energy transmission strategy for IEEE 802.11a/h," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MOBICOM '03)*, pp. 161– 175, San Diego, Calif, USA, September 2003.
- [28] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MOBICOM '01)*, pp. 236–251, Rome, Italy, July 2001.
- [29] I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips, "Hybrid rate control for IEEE 802.11," in *Proceedings of the 2nd International Workshop on Mobility Management and Wireless Access Protocols (MobiWac '04)*, pp. 10–18, Philadelphia, Pa, USA, October 2004.
- [30] I. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk, and H. Sips, "Automatic IEEE 802.11 rate control for streaming applications," *Wireless Communications and Mobile Computing*, vol. 5, no. 4, pp. 421–437, 2005.
- [31] J. Kim, S. Kim, S. Choi, and D. Qiao, "CARA: collision-aware rate adaptation for IEEE 802.11 WLANs," in *Proceedings of the* 25th IEEE International Conference on Computer Communications (INFOCOM '06), pp. 1–11, Barcelona, Spain, April 2006.
- [32] C. Hoffmann, H. Manshaei, and T. Turletti, "CLARA: closedloop adaptive rate allocation for IEEE 802.11 wireless LANs," in Proceedings of the International Conference on Wireless Networks, Communications, and Mobile Computing (WIRE-LESSCOM '05), vol. 1, pp. 668–673, Maui, Hawaii, USA, June 2005.
- [33] S. Ci and H. Sharif, "A variable data rate scheme to enhance throughput performance of wireless LANs," in *Proceedings* of the 3rd International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP '02), Staffordshire, UK, July 2002.
- [34] W. H. Yuen, H.-N. Lee, and T. D. Andersen, "A simple and effective cross layer networking system for mobile adhoc networks," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '02)*, vol. 4, pp. 1952–1956, Lisbon, Portugal, September 2002.
- [35] M. Lacage, M. H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach," in *Proceedings of the 7th Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '04)*, pp. 126–134, Venice, Italy, October 2004.
- [36] P. Ferré, A. Doufexi, J. Chung-How, A. Nix, and D. Bull, "Link adaptation for video transmission over COFDM based

WLANs," in *Proceedings of the 10th IEEE Symposium on Communications and Vehicular Technology (SCVT '03)*, Eindhoven, The Netherlands, November 2003.

- [37] K. Stuhlmüller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012– 1032, 2000.
- [38] S. Rane and B. Girod, "Analysis of error-resilient video transmission based on systematic source-channel coding," in *Proceedings of the 23rd Picture Coding Symposium (PCS '04)*, pp. 453–458, San Francisco, Calif, USA, December 2004.
- [39] S. Rane, P. Baccichet, and B. Girod, "Modeling and optimization of a systematic lossy error protection system based on H.264/AVC redundant slices," in *Proceedings of the 25th Picture Coding Symposium (PCS '06)*, vol. 2006, Beijing, China, April 2006.
- [40] D.-K. Kwon, M.-Y. Shen, and C.-C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 517–528, 2007.
- [41] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [42] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 598–603, 2003.
- [43] P. Ferré, D. Agrafiotis, and D. Bull, "Macroblock selection algorithms for error resilient H.264 video wireless transmission using redundant slices," in *Visual Communications and Image Processing*, vol. 6822 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2008.
- [44] K. Suhring, H.264/AVC Software Coordination, http://iphome .hhi.de/suehring/tml/.

## Research Article A Cross-Layer Approach for Maximizing Visual Entropy Using Closed-Loop Downlink MIMO

#### Hyungkeuk Lee, Sungho Jeon, and Sanghoon Lee

Wireless Network Laboratory, Yonsei University, Seoul 120-749, South Korea

Correspondence should be addressed to Sanghoon Lee, slee@yonsei.ac.kr

Received 1 October 2007; Revised 27 March 2008; Accepted 8 May 2008

Recommended by David Bull

We propose an adaptive video transmission scheme to achieve unequal error protection in a closed loop multiple input multiple output (MIMO) system for wavelet-based video coding. In this scheme, visual entropy is employed as a video quality metric in agreement with the human visual system (HVS), and the associated visual weight is used to obtain a set of optimal powers in the MIMO system for maximizing the visual quality of the reconstructed video. For ease of cross-layer optimization, the video sequence is divided into several streams, and the visual importance of each stream is quantified using the visual weight. Moreover, an adaptive load balance control, named equal termination scheduling (ETS), is proposed to improve the throughput of visually important data with higher priority. An optimal solution for power allocation is derived as a closed form using a Lagrangian relaxation method. In the simulation results, a highly improved visual quality is demonstrated in the reconstructed video via the cross-layer approach by means of visual entropy.

Copyright © 2008 Hyungkeuk Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

#### 1. INTRODUCTION

The ongoing broadband wireless networks have attractive advantages for providing a variety of multimedia streaming applications while guaranteeing the quality of service (QoS) for mobile users.

Nevertheless, many limitations for adapting the magnificent growth of multimedia traffic into expensive and capacity-limited wireless channels continue to exist. The multiple input multiple output (MIMO) system is capable of increasing channel throughput drastically by using multiple transmit and multiple receive antennas [1, 2]. Since the MIMO channel is composed of multiple parallel subchannels with different quality, more efficient radio resource management can be developed by exploiting such different channel characteristics. If higher and lower quality subchannels are used for more and less important data, respectively, from the perspective of cross-layer optimization, a better performance could be expected.

Some recent papers have highlighted issues of cross-layer optimization for achieving a better quality of source over a capacity-limited wireless channel [3–7]. If source-dependent information exchanges across the top and bottom protocol

layers are used, more improved performance can be obtained even if the exchanges may not be available in traditional layered architectures in [3].

The authors in [4] presented a high-level framework for resource-distortion optimization, that jointly considered factors across the network layer, including source coding, channel resource allocation, and error concealment. In [5], a framework of cross-layer design for supporting delay critical traffic over ad-hoc wireless networks was proposed and its benefits for video streaming were analyzed. In [7], a modified moving picture experts group (MPEG)-4 coding scheme was employed for progressive data transmission by controlling the number of subcarriers over a multicarrier system. Besides, the authors in [8–15] exploited joint transmission and coding schemes over MIMO systems using not only the layered coding, but also the multiple description coding (MDC). In [8], an unequal power allocation scheme for transmission of joint photographic experts group (JPEG) compressed images employing spatial multiplexing was proposed, so a significant image quality improvement was achieved compared to other schemes. Similarly, in [9], the unequal spatial diversity scheme was proposed for providing unequal error protection, which was based on



(a) PSNR = 22.3 Visual entropy = 8538.0



(b) PSNR = 23.6 Visual entropy = 10490.0



(c) PSNR = 25.1 Visual entropy = 11812.5







(e) PSNR = 23.6 Visual entropy = 5232.2



(f) PSNR = 25.7 Visual entropy = 6386.6

FIGURE 1: Quality assessment using PSNR versus visual entropy.

the combined use of turbo codes and space-time codes. It could also provide a reduction in average transmission time and a image quality improvement compared with no spatial diversity, but the criteria was not suggested. Authors in [10] presented the gains arising from transmitting MDC over spatial multiplexing (SM) systems. Authors in [11] showed that the layered coding might outperform MDC under certain conditions when an error-free environment or an environment with a very low-error rate can be guaranteed for the base layer. Nevertheless, it is presented that MDC can be one of the realistic MIMO transmission scenarios as good as the layered coding can in [12]. Authors in [13] observed that the general water-filling power allocation, while optimizing the capacity of MIMO singular value decomposition (SVD) system, may not be optimal for video.

From the perspective of cross-layer optimization, the major drawback in the previous research is the lack of the specific criteria defining the importance of each information bit. Moreover, the heuristic algorithm without the use of a mathematical proof is only presented. In order to adapt a bulky multimedia traffic to a capacity-limited wireless channel, it is necessary to generate layered video bitstreams and then to transmit more visually important data to higher quality subchannels and vice versa. Even if it is easy to conceive such idea, the main issue is how the radio resource control can be conducted based on which criterion. The most widely used quality criterion peak signal-to-noise ratio (PSNR) does not characterize the quality of the visual data perfectly. Figure 1 illustrates the defect in the PSNR value. Even though, the PSNR values shown in Figures

1(a), 1(b), and 1(c) are approximately the same as those shown in Figures 1(d), 1(e), and 1(f), respectively, the visual qualities for them are significantly different because the PSNR criterion cannot determine where distortion comes from. Therefore, the PSNR as a quality assessment does not accurately represent visual quality. However, the PSNR is known as the dominant quality assessment because, in spite of this defect, no clear quality criterion exists as an alternative. Therefore, the current technical limitation lies in the lack of quality criteria for evaluating the performance gain attained by the cross-layer approach.

In agreement with the human visual system (HVS), we recently defined "visual entropy" as the expected number of bits required to represent image information over the human visual coordinates [16, 17]. Stemming from this, a new quality metric, termed the FPSNR (Foveal PSNR) was defined, and the video coding algorithms were optimized by means of the quality criterion [18, 19]. The main attractive advantage of visual entropy lies in quantifying the visual gain as a concrete quantity such as bit.

In this paper, we explore a theoretical approach to crosslayer optimization between multimedia and wireless network layers by means of a quality criterion termed "visual entropy" for the closed-loop downlink MIMO system, using a wavelet coding algorithm. We propose an efficient unequal power allocation scheme for improving visual quality as well as for maintaining a QoS requirement. The proposed framework does not involve a redesign of existing protocols, but rather adapts existing standards seamlessly with simple configuration for multimedia transmission over the MIMO system.



FIGURE 2: Block diagram for the rate control-based closed-loop MIMO system: transmitter and receiver.

From the perspective of the HVS, an optimal power allocation set is determined for delivering the maximal visual entropy by utilizing *Lagrangian relaxation*. As a result, the power level associated with each subband is determined according to the layer of wavelet domain for maximizing visual throughput, which leads to a better visual quality by the numerical and simulation results.

In addition, due to channel variations, transmissions using different antennas may experience different packet loss rates using the optimal receiver. In this case, the greater visual quality can be obtained by transmitting the more important data via the best quality channel. Therefore, it is necessary to measure the amount of visual information for each bitstream and then to load the bitstream to a suitable antenna path according to the amount. To quantify the visual importance, visual entropy is introduced. Based on this value, the video data with a more important information is transmitted over a high-quality channel and vice versa. Besides, an adaptive load balance control scheme named equal termination scheduling (ETS) is proposed to give a privilege for high-priority data by avoiding inevitable channel errors over an error-prone channel.

#### 2. SYSTEM OVERVIEW AND ASSUMPTION

#### 2.1. The background area

Generally, the video sequence is coded into a single or multiple bitstreams according to the coding architecture, which is composed of different codewords including different degrees of importance. It is quite noticeable that each codeword contains different visual information so that the bitstream with different importance can be treated differently for provisioning higher quality services. In other words, the loss of important data may result in a severe degradation of the decoded video quality. In contrast, the loss of less important data may be tolerable. Therefore, it is necessary to provide better protection to important data, which is the basic idea of unequal error protection (UEP).

Essentially, the UEP method implicates the distribution of errors in order that more important data can experience fewer bit errors without demanding extra resource consumption. It has been widely demonstrated that the UEP is an efficient method in delivering error sensitive video over error-prone wireless channels [20]. Common approaches for the UEP are based on forward error correction (FEC) [21] or modulation scheme, such as hierarchical quadrature amplitude modulation (QAM) [22]. In [23], a UEP scheme based on subcarrier allocations in a multicarrier system is also proposed.

In this work, we propose the new UEP technique based on the HVS using the unequal power allocation and exploit the difference in visual importance of each bit stream by means of visual entropy using unequal power allocation among multiple antennas. To achieve this main goal, a wavelet-based video coding is used to encode the video sequence into multiple bitstreams with different visual contents. For example, in the two-layer video, the base layer with a high weight carries more important visual information as an independently decodable expression with acceptable quality, but the enhancement layer with a low weight carries additional detailed visual information for quality improvement. In addition, the video coder based on the wavelet transform has the desirable property of generating naturally-layered bitstreams, which are composed of low- and high-frequency components. Therefore, the UEP provides stronger protection to the layer, which contains the important visual information.

#### 2.2. At the transmitter side

Figure 2 depicts the block diagram of the MIMO system with  $M_T$  and  $M_R$  antennas at the transmitter and receiver, respectively. In addition, we assume spatially multiplexing transmission in which  $M_T$  independent data streams are sent from each transmit antenna.

Using a progressive wavelet video encoder, for example, set partitioning in hierarchical trees (SPIHT) or embedded block coding with optimized truncation (EBCOT), each layer can be constructed by scanning wavelet coefficients [24, 25]. In this case, each coefficient has a different visual importance according to the associated spatial and frequency weight. After obtaining the sum of the visual weights for each layer, the value can be included in the header. In terms of the weighted value, it is assumed that the communication system can recognize the importance of each layer.

It is assumed that the source data is divided into several independent layers by using the spatial demultiplexer as shown in Figure 2. These layers are subsequently coded, modulated separately, and then transmitted simultaneously on the same frequency. The coding, modulation, and transmit power of each layer are subject to the capacity maximization according the feedback information and the visual information which each layer contains, as depicted in Figure 2. In this paper, the optimization for the maximal capacity experienced over the wireless channel is obtained by using the Shannon capacity. Since the Shannon capacity is a theoretical upper bound afforded by using communication techniques, such as the automatic repeat-request (ARQ), forward error correction (FEC), and modulation schemes, it is assumed that the proposed system employs the best ARQ, FEC, and modulation schemes. We assume that a combination of coding and modulation at each antenna is the same. The only difference is the level of allocated power at each transmit antenna. If any power is not allocated to the *k*th antenna, the *k*th antenna is not used for transmission.

The power allocation under the total transmit power constraint is one of the roles in the preprocessing stage. It divides the streams into nonoverlapping blocks. The power optimization algorithm then runs on each of these blocks independently with respect to the amount of the visual information. The detail in the optimization procedure will be discussed later. Thus, an optimal power level is allocated to each block by taking into account the visual weight for transmitting data as much as possible from the visual quality point of view.

#### 2.3. The channel model

For numerical analysis, let  $p_k$  be the allocated power to the *k*th transmit antenna. The signal vector to be sent from the transmitter is expressed as  $\mathbf{x} = [x_1, \dots, x_{M_T}]^T$ , with  $E[\mathbf{x}\mathbf{x}^H] = \text{diag}(p_1, p_2, \dots, p_{M_T})$  subject to  $\sum_i^{M_T} p_i = \overline{P}$ , where  $\overline{P}$  is the total transmit power. The channel response between the transmitter and the receiver is represented by an  $M_R \times M_T$  MIMO channel matrix as

$$\mathbf{H} = \begin{pmatrix} h_{11} & \cdots & h_{1M_T} \\ \vdots & \ddots & \vdots \\ h_{M_R 1} & \cdots & h_{M_R M_T} \end{pmatrix}, \qquad (1)$$

where  $h_{mn}$   $(1 \le m \le M_R, 1 \le n \le M_T)$  is modeled as a complex Gaussian variable with zero-mean and unit variance representing the channel response between the *n*th transmit antenna and the *m*th receive antenna. A spatially uncorrelated channel model is assumed to be used in this paper.

Accordingly, the  $M_R \times 1$  received signal vector is then

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},\tag{2}$$

where **n** denotes the  $M_R \times 1$  independent and identically distributed zero-mean circularly symmetric complex gaussian (ZMCSCG) noise vector with the covariance matrix  $E[\mathbf{nn}^H] = N_o \mathbf{I}_{M_R}$  [26–28]. The received signal vector, **y**, is then sent to the linear receiver.

#### 2.4. At the receiver side

At the receiver, we assume that the channel is perfectly estimated for the closed-loop MIMO system. Here, three alternative receiver schemes are considered: singular value decomposition (SVD) detection, zero-forcing (ZF) detection, and minimum mean square error (MMSE) detection [29]. For ease of analysis, it is assumed that the most powerful channel estimation technique is used. Based on the information at the receiver, the estimated channel value needed to determine the allocated power is then feedback to adjust the corresponding transmission parameters as mentioned before. Authors in [14] showed that a delay in feeding the channel status information(CSI) back to the transmitter causes severe degradation in the performance of SVD systems, and the effect from this was quantified in [15]. Since this effect is beyond the scope in this paper, it is assumed that there is neither delay nor error in the feedback channel.

The channel is modeled as a complex Gaussian random variable with zero-mean and unity variance, which is also assumed to be flat fading and quasistatic so that the channel remains constant over the transmission during the execution for the power allocation after the feedback information. It is also assumed to use the optimal channel realization technique for ease of analysis.

After detecting the symbol and deciding the bits at each antenna, the raw data bitstream is then passed to the multiplexing block. The block converts these  $M_R$  bitstreams into serial streams corresponding to the number of transmit antennas. Finally, the multiplexer combines those streams into a single received bitstream.

#### 2.5. The definition of visual entropy

To measure the visual importance of each layer at the preprocessing stage, it is necessary to decide the cross-layer optimization constraint or criterion. Here, a normalized weight will be adopted as the criterion to quantify the visual importance of each layer. In [16, 17], we defined "visual entropy" as the expected number of bits required to represent image information mapped over human visual coordinates. The visual entropy in [17] is written as

$$H_d^w(a[m]) = w_m^t H_d(a[m]) = w_m^t (\log_2 \sigma_m + \log_2 \sqrt{2e^2}), \quad (3)$$

where *m* is the index of wavelet coefficients, a[m] is a random variable of coefficient with the index *m*,  $H_d(a[m])$  is the entropy of a[m],  $w_m^t$  is the visual weight, and  $\sigma_m$  is the variance when a[m] has a Laplacian distribution. Since  $H_d(a[m])$  is the minimum number of bits needed to represent a[m], the visual entropy can be expressed as a weighted version of  $H_d(a[m])$  associated with the visual weight  $w_m^t$ .

The visual weight  $w_m^t$  is characterized by using two visual components: one for the spatial domain  $w_m^s$ , and the other for the frequency domain  $w_m^f$  as shown in Figure 3.

According to the wavelet decomposition in Figure 3(a), the levels of the weights are presented in Figures 3(b), 3(c),



FIGURE 3: (a) Wavelet decomposition, (b) the weight of the spatial domain, (c) the weight of the frequency domain, and (d) the total weight wavelet domain. The brightness in the figures represents the level of visual importance.

and 3(d), respectively. When spatial visual information such as a region of interest, an object or objects, the nonuniform sampling process of the human eye can be utilized to obtain  $w_m^s$  over the spatial domain. In addition, the human visual sensitivity can be characterized by  $w_m^f$  over the frequency domain by measuring the contrast sensitivity of the human eye [30]. Based on this measurement, the total weight over the two domains can be obtained by  $w_m^t = w_m^f \cdot w_m^s$ . In the layered video coding based on the frequency band division without the use of foveation, the weight of each layer becomes  $w_m^t = w_m^j$ . In the region-based, object-based, or foveation-based video coding without the use of the layered structure, the weight becomes  $w_m^t = w_m^s$ . In the hybrid video coding based on an object-based layered mechanism, the weight over the spatial and frequency domains needs to be taken into account. In this case,  $w_m^t = w_m^f \cdot w_m^s$  The details about  $w_m^f$  and  $w_m^s$  are discussed in [17].

Since the entropy H(a[m]) is a constant value, the sum of visual entropy for M coefficients yields

$$\sum_{m=0}^{M-1} H^w(a[m]) = M \cdot H(a[m]) \sum_{m=0}^{M-1} w_m^t$$

$$= M \cdot H(a[m]) \cdot \overline{w}^t = C_w,$$
(4)

where  $C_w$  is the sum of the delivered visual entropies for each coefficients. The details are described in [17].

Since the HVS is insensitive for distortions in the fastmoving region to a considerable extent, some considerations can be applied to the visual weight for an "I-frame" or a "P-frame," respectively, according to the temporal activity of video, which is computed as the mean value of motion vectors in the frame. Authors in [31] proposed a quality metric for video quality assessment using the amplitude of motion vectors and evaluated it in accordance with a subjective quality assessment method such as double-stimulus continuous quality scale (DSCQS) and single-stimulus continuous quality evaluation (SSCQE) [32]. Therefore, it is necessary to consider the temporal extent using motion vectors for obtaining visual entropy for the video sequence. The temporal activity of the *i*th frame  $TA_i$  is, then, defined as

$$TA_{i} = \overline{|mv_{x,i}(x, y)|} + \overline{|mv_{y,i}(x, y)|}, \quad 1 \le x \le W, \ 1 \le y \le H,$$
(5)

where  $|mv_{x,i}(x, y)|$  and  $|mv_{y,i}(x, y)|$  represent the mean values of the horizontal and vertical components of the motion vector at the spatial domain (x, y) in the *i*th frame, and *W* and *H* are the width and height of the video sequence, respectively.

Reflecting the temporal activity, the visual weight  $w'_m$  can be redefined as

$$w'_{m} = \frac{w_{m}}{\left(c_{1} + \left(\max\left(TA_{i}, c_{2}\right)^{2}\right)/c_{3}\right)},$$
(6)

where  $c_1$ ,  $c_2$ , and  $c_3$  are constants determined by experiments and are used by "2.5," "5," and "30" in [31]. For brevity, it is assumed that  $w'_m$  is expressed by  $w_m$  through this paper.

## 2.6. The unequal power allocation with multiple antennas

The UEP can be implemented by utilizing the differences in the channel quality among the multiple antennas. The general UEP method has taken only the dynamics of the channel situation into account, and the UEP based on the water-filling method has been known as an optimal solution for maximum channel throughput [8, 9]. In contrast, in this paper, the amount of visual information is used as the optimal value of the object function for a given power constraint.

In the scheme, the video sequence is decoded into several bitstreams using a layered wavelet video. Each layer includes a different degree of importance which is quantified by means of visual entropy. An unequal power allocation (UPA) algorithm may be then performed in real-time. However, in general, intensive computation may be required to obtain an optimal solution. To reduce the computational complexity, we derive a closed numerical form of the optimal power for the power allocation method.

The proposed UPA technique consists of two steps: antenna selection based on the channel gain, and optimal power allocation according to the visual weight in Figure 3. The multiple antennas can be classified and ordered based on the metric of the channel gain. To perform this antenna selection at any instantaneous channel realization, we measure the channel for each antenna using a channel estimation. More specifically, the antenna with the best channel gain is labeled as the 4th antenna, and the antenna with the second best antenna as the 3rd antenna, and so on, if  $M_T = 4$ .



Step 3) Arranged packets are divided by the divisor (the number of antennas). Then, the scheduler makes an index for each packet.



Step 2) All packets are virtually arranged by the DL scheduler as if they are stacked in a single queue.



Step 4) The DL scheduler makes a plan for transmitting packets: how much packets are taken out from each queue at a certain time slot.

	Q1	Q2	Q3
Time slot 1	2	0	1
Time slot 2	1	1	1

Step 5) The DL scheduler transmits the packet taken out from the queue in accordance with the table plan in step 4.



FIGURE 4: A conceptual example of the ETS algorithm.

After performing the antenna selection and assignment for different streams, a power is then allocated to each antenna according to the visual weight of the associated video layer. Hence, more power can be allocated to more important layer, resulting in a further increase in the overall visual throughput. Therefore, the visually important data will experience less packet errors, and vice versa.

#### 2.7. The adaptive load control using the ETS algorithm

It is assumed that each layer consists of the packets, and the number of packets in each layer may be different from those of the others. In the downlink scheduler, each layer is stacked into the corresponding queue as the unit of the packet according to its priority. Since the priority is determined based on the visual importance carried in the packet so that the packet classification is accomplished through queues in the scheduler.

The procedure of the ETS algorithm is described in detail as follows.

(1) Step 1: based on the visual weight, which each packet contains, the transmission priority is determined so that it can be stacked in the corresponding queue. In Figure 4, the queue of Q1 has the highest priority, which contains three

packets notated *A*, *B*, and *C*. the priority is decreased in the order of Q1, Q2, and Q3.

(2) Step 2: all the packets in the queues are virtually arranged by the scheduler as if they are stacked in a single queue as shown in Figure 4.

(3) Step 3: the arranged packets are divided by the divisor which is the number of transmit antennas. The scheduler then makes an index for each packet. It is assumed that three channels are available so that the arranged packets are divided into three subgroups.

(4) Step 4: the scheduler makes a plan for transmitting the packets: how many packets are drawn in each queue at each time slot. For example, the total number of packets is 6 over the three available antennas so that two-time slots are required to transmit all the packets. In Q1, two packets are transmitted at the first time slot and one packet is transmitted in the second time slot. In case of Q2, no packet is transmitted in the first time slot, and the remaining packet is transmitted in the second time slot.

(5) Step 5: the scheduler transmits the packet from the queue in accordance with the table obtained in step 4.

Based on the explanation of the procedure, it can be seen that the transmit order is strictly controlled by the scheduler based on the virtual map. The main issue is how to drop

Step 1) Different priority data are stacked in a different priority downlink queue.



FIGURE 5: Tail packets are discarded regardless of their weights in the ETS algorithm.

packets if the channel capacity is not enough to transmit all the packets. The issue is how to deal with remaining packets and the solution, *the tail packet discarding*, is proposed as depicted in Figure 5.

For example, Figures 5(a) and 5(b) are the cases of requiring 3 time slots with 2 antennas, and Figure 5(c) is the case of requiring 2 time slots with 3 antennas. The remainder occurs when the number of packets is not exactly divided by the divisor. In such a case, the remaining packets are discarded regardless of its visual weight, since the visual weight of the remaining packets are relatively smaller for the previous queueing and virtual arrangement. Thus, utilizing the ETS algorithm, the throughput of visually important data can be maintained while delivering the packets in the order of arrival at the scheduler. The policy of tail packet dropping contributes an efficient use of resources for delay sensitive but loss tolerant video traffic.

#### 3. OPTIMAL POWER CONTROL USING LAGRANGIAN RELAXATION

In this section, a numerical analysis for cross-layer optimization is described to maximize the amount of the transmitted data over the MIMO system. In particular, we make an effort to transmit the visual information as much as possible for a given channel capacity. Thus, in the optimization problem, the source rate is expressed by means of visual entropy, and the channel capacity is calculated by Shannon theorem.

To maximize visual entropy, an optimization problem can be formulated as follows:

(A) 
$$\max \sum_{m=1}^{M} H^{w}(a[m]),$$
  
subject to 
$$\sum_{m=1}^{M} H(a[m]) \leq C,$$
 (7)

where H(X) is the entropy of a random variable X,  $H^w(X)$  is the visual entropy of X, m is the index of coefficients, and C is the channel capacity. This objective function for the optimization will be more specified according to the type of the receiver as follows.



FIGURE 6: Utilizing precoder and decoder via decomposition of **H** when the channel is known to both transmitter and receiver.

#### 3.1. SVD (singular value decomposition) receiver

In [29], the eigen-mode spatial multiplexing method is studied by performing singular value decomposition (SVD) on the channel response matrix. Through precoding at the transmitter and decoding at the receiver, the channel matrix is converted into a matrix as

$$\Sigma = \mathbf{U}^{H} \mathbf{H} \mathbf{V}$$

$$= \begin{pmatrix} \sqrt{\lambda_{1}} & \mathbf{0} & \mathbf{0} \\ & \ddots & & \\ \mathbf{0} & \sqrt{\lambda_{r}} & \\ \mathbf{0} & \cdots & \mathbf{0}_{(M_{T}-r) \times (M_{R}-r)} \end{pmatrix}, \qquad (8)$$

where  $r \leq \min\{M_T, M_R\}$  is the rank of **H**, and  $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the eigenvalues of the channel matrix **HH**<sup>*H*</sup>. Terms **U**<sup>*H*</sup> and **V** are the  $M_R \times M_R$  and  $M_T \times M_T$  unitary matrices that are used as the decoding and precoding matrices, respectively. Therefore, (2) becomes

$$y = Hx + n$$
  
=  $U\Sigma V^{H}x + n.$  (9)

By multiplying **V** and **U**<sup>*H*</sup> to **x** and **y**, (9) is transformed into

$$\mathbf{U}^{H}\mathbf{y} = \widetilde{\mathbf{y}}$$
  
=  $\mathbf{U}^{H}\mathbf{H}\mathbf{V}\widetilde{\mathbf{x}} + \mathbf{U}^{H}\mathbf{n}$   
=  $\mathbf{U}^{H}\mathbf{H}\mathbf{V}\widetilde{\mathbf{x}} + \widetilde{\mathbf{n}}$  (10)  
=  $\mathbf{U}^{H}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{H}\mathbf{V}\widetilde{\mathbf{x}} + \widetilde{\mathbf{n}}$   
=  $\Sigma\widetilde{\mathbf{x}} + \widetilde{\mathbf{n}}$ .

Figure 6 shows the schematic channel model of *eigen-mode transmission* when the channel is known to the transmitter and receiver.

Equation (10) shows that  $\mathbf{H}$  can be explicitly decomposed into r parallel single input single output (SISO) channels satisfying

$$\widetilde{y}_k = \sqrt{\lambda_k} \widetilde{x}_k + \widetilde{n} \tag{11}$$

when the transmitter knows the channel matrix.

Since  $\mathbf{U}^H$  is a unitary matrix,  $\mathbf{U}^H \mathbf{n}$  has the same covariance as  $\mathbf{n}$ , and thus the postprocessing SNR for the *k*th data stream is

$$\mathrm{SNR}_k = \frac{p_k}{N_o} \lambda_k,\tag{12}$$

where  $p_k = E\{|x_k|^2\}, \sum_{k=1}^{M_T} p_k \leq \overline{P}, \lambda_k \text{ is 0 if } k > r. p_k$ reflects the transmit energy in the *i*th subchannel and satisfies  $\sum_{k=1}^{M_T} p_k \leq \overline{P}.$  From (12), it is clear that the received SNR of each data stream is proportional to its transmit power. Furthermore, since the transmission rate is continuous, the optimum strategy for power allocation is simply based on the waterfilling theory [1].

To obtain the optimum power value using SVD, (7) can be transformed to a new problem by (12) as follows:

(B1) 
$$\max_{p_k} \sum_{k=1}^{r} \overline{w}_k^t \cdot \log_2\left(1 + \frac{p_k}{N_o}\lambda_k\right),$$
  
subject to 
$$\sum_{k=1}^{r} p_k \le \overline{P}, \ p_k \ge 0$$
 (13)

where  $\overline{P}$  is a total transmit power with respect to all transmit antennas, and  $\overline{w}_k^t$  is the value of the visual weight in the transmitted layer corresponding to the assigned *k*th transmit antenna. The solution in (13) is an optimal power set,  $\{p_1, p_2, \ldots, p_{M_T}\}$ . Because (13) is a convex problem, we can apply to the Karush-Kuhn-Tucker (KKT) condition with respect to  $p_k$  to obtain an optimal power set which is a globally optimum solution.

Using a Lagrangian relaxation,

$$L(p_k, \nu) = \sum_{k=1}^{r} \overline{w}_k^t \cdot \log_2\left(1 + \frac{p_k}{N_o}\lambda_k\right) + \nu\left(\overline{P} - \sum_{k=1}^{r} p_k\right), \quad (14)$$

where  $\nu$  is a nonnegative Lagrangian multiplier. Taking the derivatives with respect to  $p_k$  and  $\nu$  can be obtained as follows:

$$\frac{\partial L}{\partial p_k} = \overline{w}_k^t \cdot \frac{\lambda_k / N_o}{\left(1 + p_k \lambda_k / N_o\right) \ln 2} - \nu \le 0, \tag{15}$$

$$p_k \cdot \frac{\partial L}{\partial p_k} = 0, \tag{16}$$

$$\nu\left(\overline{P} - \sum_{k=1}^{r} p_k\right) = 0.$$
(17)

From (15) and (16), if power  $p_k$  is allocated to the *k*th data stream (i.e.,  $p_k \ge 0$ ), the complementary slackness condition is then satisfied as follows:

$$\overline{w}_{k}^{t} \cdot \frac{\lambda_{k}/N_{o}}{\left(1 + p_{k}\lambda_{k}/N_{o}\right)\ln 2} = \nu.$$
(18)

In addition, the optimal values of  $p_k$  and its multiplier  $\nu$  are given by

$$p_k = \frac{\overline{w}_k^t}{\nu \ln 2} - \frac{N_o}{\lambda_k}.$$
(19)

Substituting (17) with (19),

$$\frac{1}{\nu \ln 2} = \frac{\overline{P} + N_o \sum_{k=1}^r (1/\lambda_k)}{\sum_{k=1}^r \overline{w}_k^t}.$$
 (20)

Substituting (21) with (20),

$$p_k = \frac{\overline{w}_k^t}{\sum_{k=1}^r \overline{w}_k^t} \left( \overline{P} + N_o \sum_{k=1}^r \frac{1}{\lambda_k} \right) - \frac{N_o}{\lambda_k}.$$
 (21)

#### 3.2. MMSE (minimum mean square error) receiver

The MMSE matrix filter for extracting the received signal into the *k*th component transmitted stream is given by

$$\mathbf{G}_{\mathrm{MMSE}} = \mathbf{h}_{k}^{H} \left( N_{o} \mathbf{I}_{M_{R}} + \sum_{i \neq k}^{M_{T}} p_{i} \mathbf{h}_{i} \mathbf{h}_{i}^{H} \right)^{-1}, \qquad (22)$$

where  $\mathbf{h}_k$  is the *k*th column of  $\mathbf{H}$ , that is,  $M_R \times 1$  vector. Thus, the SINR for the *k*th data stream can be expressed as

$$\operatorname{SINR}_{k} = p_{k} \mathbf{h}_{k}^{H} \left( N_{o} \mathbf{I}_{M_{R}} + \sum_{i \neq k}^{M_{T}} p_{i} \mathbf{h}_{i} \mathbf{h}_{i}^{H} \right)^{-1} \mathbf{h}_{k} = p_{k} g_{k}, \quad (23)$$

where  $g_k = \mathbf{h}_k^H (N_o \mathbf{I}_{M_R} + \sum_{i \neq k}^{M_T} p_i \mathbf{h}_i \mathbf{h}_i^H)^{-1} \mathbf{h}_k$ .

To obtain the optimum power value using the MMSE receiver, (7) can be transformed to a new problem using (23) as follows:

(B3) 
$$\max_{p_k} \sum_{k=1}^{M_T} \overline{w}_k^t \cdot \log_2(1 + p_k g_k),$$
  
subject to 
$$\sum_{k=1}^{M_T} p_k \le \overline{P}, \ p_k \ge 0.$$
 (24)

Equation (24) is also a convex problem, we can apply to the KKT condition with respect to  $p_k$  to obtain an optimal power set. By using a Lagrangian relaxation,

$$L(p_k,\nu) = \sum_{k=1}^{M_T} \overline{w}_k^t \cdot \log_2(1+p_k g_k) + \nu \left[\overline{P} - \sum_{k=1}^{M_T} p_k\right], \quad (25)$$

where  $\nu$  is a nonnegative Lagrangian multiplier. Taking the derivatives with respect to  $p_k$  and  $\nu$ , respectively, then

$$\frac{\partial L}{\partial p_k} = \overline{w}_k^t \cdot \frac{g_k}{(1+p_k g_k) \ln 2} - \nu \le 0, \tag{26}$$

$$p_k \cdot \frac{\partial L}{\partial p_k} = 0, \qquad (27)$$

$$\nu\left(\overline{P} - \sum_{k=1}^{M_T} p_k\right) = 0.$$
(28)

Using (26) and (27), the complementary slackness condition is given by

$$\overline{w}_k^t \cdot \frac{g_k}{\left(1 + p_k^* g_k\right) \ln 2} = \nu.$$
<sup>(29)</sup>

The optimal power is obtained by

$$p_k^* = \frac{1}{g_k} \left[ -1 + \frac{\overline{w}_k^t \cdot g_k}{\nu \ln 2} \right].$$
(30)

Using (28) and (30),

$$\frac{1}{\nu \ln 2} = \frac{\overline{P} + \sum_{k=1}^{M_T} (1/g_k)}{\sum_{k=1}^{M_T} \overline{w}_k^t \cdot g_k}.$$
 (31)

Using (30) and (31),

$$p_k = \frac{1}{g_k} \left[ -1 + \frac{\overline{w}_k^t \cdot g_k}{\sum_{k=1}^{M_T} \overline{w}_k^t \cdot g_k} \left( \overline{P} + \sum_{k=1}^{M_T} \frac{1}{g_k} \right) \right].$$
(32)

TABLE 1: Visual weight for each layer.

	Layer 1	Layer 2	Layer 3	Layer 4
Visual weight (I-frame)	0.09236	0.12258	0.17951	0.45107
Visual weight (P-frame)	0.12568	0.16728	0.24783	0.45920

#### 3.3. ZF (zero forcing) receiver

The zero forcing (ZF) matrix filter for extracting the received signal into its component transmitted streams is given by

$$\mathbf{G}_{\mathrm{ZF}} = \left(\mathbf{H}^{H}\mathbf{H}\right)^{-1}\mathbf{H}^{H},\tag{33}$$

where  $G_{ZF}$  is an  $M_T \times M_R$  pseudo-inverse matrix that simply inverts the channel. The output of the ZF receiver is given by

$$\mathbf{G}_{ZF}\mathbf{y} = \mathbf{x} + \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}.$$
 (34)

Thus, the postprocessing SNR for the *k*th data stream in [26–28] can be expressed as

$$\mathrm{SNR}_{k} = \frac{p_{k}}{N_{o} [\mathbf{H}^{H} \mathbf{H}]_{k,k}^{-1}}.$$
(35)

To obtain the optimum power value using the ZF receiver, (7) can be transformed to a new problem using (35) as follows:

(B2) 
$$\max_{p_k} \sum_{k=1}^{M_T} \overline{w}_k^t \cdot \log_2 \left( 1 + \frac{p_k}{N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}} \right),$$
  
subject to 
$$\sum_{k=1}^{M_T} p_k \le \overline{P}, \ p_k \ge 0.$$
 (36)

The solution of the optimization problem in (36) is an optimal power set,  $\{p_1, p_2, ..., p_{M_T}\}$  for each antenna. Because (36) is a convex problem, we apply the KKT condition with respect to  $p_k$  to obtain an optimal power set which is a globally optimum solution.

By using a Lagrangian relaxation,

$$L(p_k, \nu) = \sum_{k=1}^{M_T} \overline{w}_k^t \cdot \log_2 \left( 1 + \frac{p_k}{N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}} \right) + \nu \left( \overline{P} - \sum_{k=1}^{M_T} p_k \right),$$
(37)

where  $\nu$  is a nonnegative Lagrangian multiplier. Taking the derivatives with respect to  $p_k$  and  $\nu$ , respectively, yields the KKT conditions as follows:

$$\frac{\partial L}{\partial p_k} = \overline{w}_k^t \cdot \frac{1/N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}}{(1 + p_k/N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}) \ln 2} - \nu \le 0, \qquad (38)$$

$$p_k \cdot \frac{\partial L}{\partial p_k} = 0, \tag{39}$$

$$\nu\left(\overline{P} - \sum_{k=1}^{M_T} p_k\right) = 0. \tag{40}$$

From (38) and (39), if  $p_k$  is allocated to the *k*th data stream (i.e.,  $p_k \ge 0$ ), the complementary slackness condition is then satisfied as follows:

$$\overline{w}_{k}^{t} \cdot \frac{1/N_{o} [\mathbf{H}^{H} \mathbf{H}]_{k,k}^{-1}}{(1 + p_{k}^{*}/N_{o} [\mathbf{H}^{H} \mathbf{H}]_{k,k}^{-1}) \ln 2} = \nu.$$
(41)

The optimal value of  $p_k^*$  is given by

$$p_k^* = \frac{\overline{w}_k^t}{\nu \ln 2} - N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}.$$
 (42)

Substituting (40) with (42),

$$\frac{1}{\nu \ln 2} = \frac{\overline{P} + N_o \sum_{k=1}^{M_T} [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}}{\sum_{k=1}^{M_T} \overline{w}_k^t}.$$
(43)

Substituting (44) with (43), the optimal power can be obtained by

$$p_k^* = \frac{\overline{w}_k^t}{\sum_{k=1}^{M_T} \overline{w}_k^t} \left( \overline{P} + N_o \sum_{k=1}^{M_T} [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1} \right) - N_o [\mathbf{H}^H \mathbf{H}]_{k,k}^{-1}.$$
(44)

In short, the optimal power sets for maximizing visual entropy for the cases of SVD, MMSE, and ZF receivers are (21), (32), and (44), respectively.

#### 4. NUMERICAL RESULTS

In the simulation, the three different types of linear receivers are adopted for performance comparison. First of all, the major parameters used for the simulation are SNR: 0 dB, the number of transmit antennas: 4, the number of receive antennas: 4, and the total transmit power: 1. The "Lena" (frame size -256 by 256) is used to apply the proposed algorithm to the I-frame analysis, and the "Stefan" (frame size -352 by 240, frame rate -15 frame/second) is used to apply it to the P-frame analysis. The total transmit power is normalized to analyze with ease.

We made the encoded data from the "Lena" image using the modified SPIHT in [33]. First, after extracting the coefficients from the first sorting and refinement pass, the visual weight of these data is obtained. Similarly, the visual weights are calculated for the next three data extracted from the next passes, and four layers were loaded to the transmit antenna according to the visual weight.

In addition, the visual weight  $\overline{w}_k^t$  for each layer or bitstream in (4) is used for the simulation as listed in Table 1, and the amount of visual information can be different according to the visual weight in Table 1 ((a) and (b) represent the visual weight for the "Lena" and "Stefan," resp.) These values are consistent to the results in Figure 7.



FIGURE 7: The reconstructed images without the 1st, 2nd, 3rd, and 4th layer data, from (a) to (d), respectively.



FIGURE 8: The sum capacity versus the sum of visual entropy according to the receiver configuration.

Figure 7 represents the images reconstructed without the 1st, 2nd, 3rd, and 4th layer data, respectively, assuming that the higher number layer has more important data, which will load to an antenna with a higher number. In other words, each subfigure represents the reconstructed data without information as much as the visual weight,  $\overline{w}_1^t$ ,  $\overline{w}_2^t$ ,  $\overline{w}_3^t$ , and  $\overline{w}_4^t$ , respectively. Whereas the image in Figure 7(a) without the 1st information has a relatively small degradation for quality, the image in Figure 7(d) has the poorest quality among all the images due to the loss of the information in the 4th layer, and this shows that the 4th layer has the most visually important data. The quantity of this information can be calculated by means of the visual weight.

A common channel matrix of H, the ZMCSCG channel is used, and the uncorrelated channel is only considered in the numerical analysis.

Figure 8 shows the sum rate of the capacity and the total visual entropy according to the linear receiver. The sum rate is measured by Shannon capacity theorem [26] for

the unequal power allocation scheme and by the conventional water-filling scheme. As mentioned, the general UEP methods have used only the channel quality metric to apply the water-filling scheme, but the proposed method achieves a maximal visual throughput via visual entropy. Although an absolute maximal volume of the transmitted data for the proposed method can be lower than that of the waterfilling scheme, the proposed system can obtain greater visual information compared to the water-filling scheme.

In addition, it can be seen in Figure 8 that the channel throughput of the proposed scheme is greater than that of the conventional water-filling scheme regardless of the receiver type, but a higher visual entropy can be obtained. Consequently, although the proposed method entails a certain loss of transmitted bits from the Shannon capacity point of view, the throughput gain in terms of the visual entropy is increased up to about 20%. In other words, the proposed technique does not obtain the maximal mutual information compared to the water-filling algorithm for a


FIGURE 9: The amplitude of the allocated power, the number of transmit bits, and the related visual entropy according to the type of different receivers: (a)-(c) SVD receiver, (d)-(f) MMSE receiver, and (g)-(i) ZF receiver. The 4th layer has the highest visual weight, and the 1st layer has the least visual weight.



FIGURE 10: "Lena" images using (a) the proposed, (b) the water-filling, and (c) the equal power methods.



FIGURE 11: The 2nd frame for "Stefan" using (a) the proposed, (b) the water-filling, and (c) the equal power methods.

given channel condition, but the visual QoS is significantly enhanced from the users point of view.

Figures 9(a), 9(d), and 9(g) show optimal power sets using (19), (42), and (30), which are the solutions of (13), (36), and (24), respectively. In the ETS scheme, the optimal set of power is determined according to the visual weight carried in each packet. Although the same amount of data is delivered over each band, each bitstream has a different visual information. Since the 4th layer has the most sensitive visual information in terms of the HVS, it can be seen that the highest power is allocated to the 4th SISO channel. The power patterns for the rest of the layers are relatively smaller compared to other power allocation algorithms.

The findings show that an increase in the allocated power of the 4th layer results in an improvement in throughput as shown in Figure 9. Since the visual weight of the layer is the greatest compared to the other layers, it is expected that much higher visual entropy can be delivered by using the unequal power allocation according to the visual importance.

Figures 9(b), 9(e), and 9(h) show the number of transmitted bits using (13), (36), and (24), respectively, where the value of  $\overline{w}_b^k$  is assumed to "one." Under the given channel environment, the UPA based on the water-filling can transmit the more number of bits over the antenna arrays. The proposed scheme allocates a higher power for the 4th layer, it can be seen that the throughput of the layer is relatively lower than that of the water-filling case.

Figures 9(c), 9(f), and 9(i) show the values of visual entropy using (13), (36), and (24), respectively. In the view

of visual entropy, it can be found that the proposed method demonstrates the best performance. Moreover, additional visual entropy gain can be achieved because a greater power is allocated to the bitstream of the 4th band. The similar tendencies can be founded regardless of the receiver type.

Even if the number of the total transmit bits for the proposed method is lower than that of the water-filling scheme, the throughput of visual data can be significantly increased due to the UPA for each layer in the order of visual entropy.

Figure 10 shows the reconstructed images using the proposed, water-filling, and equal power methods when the SVD transmission is employed as the linear receiver. Due to the less throughput of visual entropy in the other schemes, the visual quality is much degraded compared to the proposed scheme. In case of the proposed method, even though it does not receive any 1st or 2nd information, the received image can have the best quality by protecting the most important data. It can be seen that these results are, therefore, consistent to the numerical results in Figures 8 and 9.

Figure 11 shows the reconstructed frames for "Stefan" using the proposed, water-filling, and equal power methods when the SVD transmission is employed as the linear receiver. It is assumed that the previous I-frame (i.e., the 1st frame) is transmitted with an error-free channel, and data with respect to only the motion vector is loaded to the MIMO antennas. These results are consistent to the previous results as shown in Figure 10.

## 5. CONCLUSION

In this paper, we considered the realization of UEP in the MIMO system using the channel feedback, in which data can be transmitted simultaneously through multiple antennas. We proposed an effective way to improve the error resilience of compressed video based on a cross-layer approach. Due to two-dimensional characteristics of video, that is, different portions of video data have different importance, video data can be divided in the metric of visual entropy. In this work, we employ an image quality metric and visual entropy to quantify the image quality. Due to channel variations and the amount of the allocated power, transmissions on different antennas may experience different packet loss rates. Thus, to achieve the different error distribution according to data with different visual weight, data with higher priority is transmitted in order to achieve higher channel gain for lower loss and error rate, and data with lower priority is on the remaining channel. Meanwhile, an adaptive load balance control scheme is proposed to give a privilege for highpriority data by passing transmission errors to data with lower priority for avoiding inevitable channel errors over an error-prone channel. The simulation results demonstrate that the proposed adaptive transmission scheme achieves significantly better performance than existing conventional systems.

## ACKNOWLEDGMENTS

This work was supported by the Korea Science and Engineering Foundation Grant funded by the Korea government (MOST) (no. R01-2007-000-11708-0), and Seoul Research & Business Development Program (11136M0212351).

#### REFERENCES

- I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [3] H. Zheng, "Optimizing wireless multimedia transmissions through cross layer design," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '03)*, vol. 1, pp. 185–188, Baltimore, Md, USA, July 2003.
- [4] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packetbased real-time video transmission," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 135–146, 2005.
- [5] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Crosslayer design of ad hoc networks for real-time video streaming," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 59–64, 2005.
- [6] H. Jiang, W. Zhuang, and X. Shen, "Cross-layer design for resource allocation in 3G wireless networks and beyond," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 120–126, 2005.
- [7] N. Conci, G. B. Scorza, and C. Sacchi, "A cross-layer approach for efficient MPEG-4 video streaming using multicarrier spread-spectrum transmission and unequal error protection," in *Proceedings of the IEEE International Conference on Image*

Processing (ICIP '05), vol. 1, pp. 201–204, Genova, Italy, September 2005.

- [8] M. F. Sabir, R. W. Heath Jr., and A. C. Bovik, "Unequal power allocation for JPEG transmission over MIMO systems," in *Proceedings of the 39th Asilomar Conference on Signals, Systems* and Computers, pp. 1608–1612, Pacific Grove, Calif, USA, October-November 2005.
- [9] M. F. Sabir, R. W. Heath Jr., and A. C. Bovik, "An unequal error protection scheme for multiple input multiple output systems," in *Proceedings of the 36th Asilomar Conference on Signals Systems and Computers*, vol. 1, pp. 575–579, Pacific Grove, Calif, USA, November 2002.
- [10] M. Tesanovic, D. Bull, A. Doufexi, and A. Nix, "Analysis of IEEE 802.11n-like transmission techniques with and without prior CSI for video applications," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '07)*, vol. 6, pp. 493–496, San Antonio, Tex, USA, September 2007.
- [11] N. Gogate, D.-M. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 777–792, 2002.
- [12] S. Lin, A. Stefanov, and Y. Wang, "Joint source and spacetime block coding for MIMO video communications," in *Proceedings of the 60th IEEE Vehicular Technology Conference* (*VTC '04*), vol. 4, pp. 2508–2512, Los Angeles, Calif, USA, September 2004.
- [13] M. Tesanovic, D. R. Bull, A. Doufexi, and A. R. Nix, "H.264-based multiple description coding for robust video transmission over MIMO systems," *Electronics Letters*, vol. 42, no. 18, pp. 1028–1030, 2006.
- [14] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 757–764, 2005.
- [15] M. Tesanovic, D. R. Bull, A. Doufexi, V. Sgardoni, and A. R. Nix, "Impact of CSI latency on video quality in MIMO systems employing singular value decomposition," *Electronics Letters*, vol. 43, no. 18, pp. 972–973, 2007.
- [16] H. Lee and S. Lee, "Visual data rate gain for wavelet foveated image coding," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 41–44, Genova, Italy, September 2005.
- [17] H. Lee and S. Lee, "Visual entropy gain for wavelet image coding," *IEEE Signal Processing Letters*, vol. 13, no. 9, pp. 553– 556, 2006.
- [18] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 977–992, 2001.
- [19] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Transactions on Multimedia*, vol. 4, no. 1, pp. 129–132, 2002.
- [20] L. Hanzo and J. Streit, "Adaptive low-rate wireless video phone schemes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 4, pp. 305–318, 1995.
- [21] D. G. Daut and J. W. Modestino, "Two-dimensional DPCM image transmission over fading channels," *IEEE Transactions* on Communications, vol. 31, no. 3, pp. 315–328, 1983.
- [22] ETSI, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television (DVB-T)," Tech. Rep. ETSI EN 300 744, V1.4.1, European Telecommunication Standard Institute, Sophia Antipolis, France, 2001.
- [23] G.-H. Yang, D. Shen, and V. O. K. Li, "UEP for video transmission in space-time coded OFDM systems," in *Proceedings*

of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04), vol. 2, pp. 1200– 1210, Hongkong, March 2004.

- [24] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.
- [25] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamental, Standards and Practice*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
- [26] A. Paulraj, R. Nabir, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.
- [27] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, Cambridge, UK, 2005.
- [28] A. Goldsmith, Wireless Communications, Cambridge University Press, Cambridge, UK, 2005.
- [29] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2198–2206, 2001.
- [30] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.
- [31] F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective no-reference metric for digital video quality assessment," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 685–688, 2005.
- [32] ITU-T Recommendation BT.500-10, "Methodology for the subjective assessment of the quality of television pictures," 2000.
- [33] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1397–1410, 2001.

## Research Article

## Joint Video Summarization and Transmission Adaptation for Energy-Efficient Wireless Video Streaming

## Zhu Li,<sup>1</sup> Fan Zhai,<sup>2</sup> and Aggelos K. Katsaggelos<sup>3</sup>

<sup>1</sup> Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong

<sup>2</sup>DSP Systems, ASP, Texas Instruments Inc., Dallas, TX 75243, USA

<sup>3</sup> Department of Electrical Engineering & Computer Science (EECS), Northwestern University, Evanston, IL 60208, USA

Correspondence should be addressed to Zhu Li, zhu.li@ieee.org

Received 13 October 2007; Accepted 25 February 2008

Recommended by Jianfei Cai

The deployment of the higher data rate wireless infrastructure systems and the emerging convergence of voice, video, and data services have been driving various modern multimedia applications, such as video streaming and mobile TV. However, the greatest challenge for video transmission over an uplink multiaccess wireless channel is the limited channel bandwidth and battery energy of a mobile device. In this paper, we pursue an energy-efficient video communication solution through joint video summarization and transmission adaptation over a slow fading wireless channel. Video summarization, coding and modulation schemes, and packet transmission are optimally adapted to the unique packet arrival and delay characteristics of the video summaries. In addition to the optimal solution, we also propose a heuristic solution that has close-to-optimal performance. Operational energy efficiency versus video distortion performance is characterized under a summarization setting. Simulation results demonstrate the advantage of the proposed scheme in energy efficiency and video transmission quality.

Copyright © 2008 Zhu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

The rapid increase in channel bandwidth brought about by new technologies such as the present third-generation (3G), the emerging fourth-generation (4G) wireless systems, and the IEEE 802.11 WLAN standards is enabling video streaming in personal communications and driving a wide range of modern multimedia applications such as video telephony and mobile TV. However, transmitting video over wireless channels from mobile devices still faces some unique challenges. Due to the shadowing and multipath effect, the channel gain varies over time, which makes reliable signaling difficult. On the other hand, a major limitation in any wireless system is the fact that mobile devices typically depend on a battery with a limited energy supply. Such a limitation is especially of concern because of the high energy consumption rate for encoding and transmitting video bit streams. Therefore, how to achieve reliable video communications over a fading channel with energy efficiency is crucial for the wide deployment of wireless video-based applications.

Energy-efficient wireless communications is a widely studied topic. For example, a simple scheme is to put the device into sleep mode when not in use, as in [1, 2]. Although the energy consumption on circuits is being driven down, as the VLSI design and integrated circuit (IC) manufacturing technologies advance, the communication energy cost is lower bounded by information theory results. In [3], the fundamental tradeoff between average power and delay constraint in communication over fading channels is explored and characterized. In [4], optimal power control schemes for communication over fading channels are developed. In [5, 6], optimal offline and near optimal online packet scheduling algorithms are developed to directly minimize energy usage in transmitting a given amount of information over fading channels with certain delay constraints.

Video streaming applications typically have different quality of service (QoS) requirements with respect to packet loss probability and delay constraints, which differentiate them from traditional data transmission applications. Approaches of cross-layer optimization of video source coding/adaptation and communication decisions have been widely adopted. Taking advantage of the specific characteristics of video source and jointly adapting video source coding decisions with transmission power, modulation and coding schemes can achieve substantial energy efficiency compared with nonadaptive transmission schemes. Examples of this type of work are reported in [7-11]. In those studies, sourcecoding controls are mostly based on frame and/or macroblock (MB) level coding mode and parameter decisions.

When both bandwidth and energy are severely limited for video streaming, sending a video sequence over with severe distortion is not desirable. Instead, we consider joint video summarization and transmission approaches to achieve the required energy efficiency. Video summarization is a video adaptation technique that selects a subset of video frames from the original video sequence based on some criterion, e.g., some newly defined frame loss distortion metric [12], specified by the user. It generates a shorter yet visually more pleasing sequence than traditional technologies that usually focus on the optimization of quantization parameters (QP) [12], which can have serious artifacts at reconstruction at very low bit rates.

Video summarization may be required when a system is operating under limited bandwidth conditions, or under tight constraints in viewing time or storage capacity. For example, for a remote surveillance application in which video must be recorded over long lengths of time, a shorter version of the original video sequence may be desirable when the viewing time is a constraint. Video summarization is also needed when important video segments must be transmitted to a base station in real time in order to be viewed by a human operator. Examples of the video summarization and related shot segmentation work can be found in [13–18], where a video sequence is segmented into video shots, and then one or multiple key frames per shot are selected based on certain criterion for the summary.

In this work, we consider the application of video summarization over wireless channels. In particular, we consider using the scheme of video summarization together with other adaptations including transmission power and modulations to deal with problems in uplink wireless video transmission arising from the severe limitation in both bandwidth and transmission energy. Since the summarization process inevitably introduces distortion, and the summarization "rate" is related to the conciseness of the summary, we formulated the summarization problem as a rate-distortion optimization problem in [12], and developed an optimal solution based on dynamic programming. We extended the formulation to deal with the situation where bit rate is used as summarization rate in [19]. In [20, 21], we formulated the energy-efficient video summarization and transmission problem as an energy-summarization distortion optimization problem; the solution of which is found through jointly optimizing the summarization and transmission parameters/decisions to achieve the operational optimality in energy efficiency. In this paper, we further extend the work in [20, 21] to consider the maximum frame drop distortion case for energy-efficient streaming. We also propose a heuristic solution, which is a greedy method that approximates well the performance of the optimal solutions. The rest of the paper is organized as follows. In Section 2, we describe the assumptions on the communication over fading wireless channels and formulate the problem as an energy-summarization distortion optimization problem. In Section 3, we develop an optimal solution based on Lagrangian relaxation and dynamic programming, as well as a heuristic solution. In Section 4, we present simulation results. Finally, in Section 5 we draw conclusions and discuss the future work in this area.

## 2. ASSUMPTIONS AND PROBLEM FORMULATION

In this section, we describe the channel model used in this work, carry out delay analysis for video summary packets, and provide the problem formulations.

#### 2.1. Wireless channel models and assumptions

In this work, we assume that the wireless channel can be modeled as a band-limited, additive white Gaussian noise (AWGN) channel with discrete time, and slow block fading. The output  $y_k$  is a function of the input  $x_k$  as

$$y_k = \sqrt{h_k} x_k + n_k, \tag{1}$$

where  $h_k$  is the channel gain for time slot k and  $n_k$  is the additive Gaussian noise with power spectrum density N. We assume that the channel gain stays constant for time  $T_c$ , the channel coherent time, and that the symbol duration  $T_s$  satisfies  $T_s \ll T_c$ , thus the channel is slow fading and there are many channel uses during each time slot. The variation of the channel state is modeled as a finite state Markov channel (FSMC) [22], which has a finite set of possible states,  $H = \{h_1, h_2, \ldots, h_m\}$ , and transitions every  $T_c$  second with probability given by the transition probability matrix  $A = |a_{ii}|$ , where  $a_{ii} = \text{Prob} \{\text{transition from } h_i \text{ to } h_i\}$ .

To reliably send R information bits over the fading channel in one channel use, the minimum power needed with optimal coding is given as [23]

$$P = N(2^{2R} - 1)/h,$$
 (2)

where *h* represents the channel gain. Similarly to the analysis in [5], let x = 1/R be the number of transmissions needed to send one bit over the channel; we can characterize the energy-delay tradeoff as  $E_b$ , energy per bit as a function of *x* as

$$E_b(x,h) = xP = xN(2^{2/x} - 1)/h.$$
 (3)

Examples of the energy efficiency functions with different fading states are shown in Figure 1. The range of *x* in Figure 1 corresponds to the received signal-to-noise tatio (SNR) of 2.0 dB to 20 dB, a typical operating range for wireless communication. To send a data packet with *B* bits and deadline  $\tau$ , assuming  $\tau \gg T_c$ , the number of transmissions available is equal to  $2W\tau$ , where *W* is the signaling rate. Then



FIGURE 1: Energy-efficiency over fading channels.

the expected energy cost will be

$$E(B,\tau) = \mathbf{E}_H \{ E_b(2W\tau/B,h)B \mid A,H,h_0 \}.$$
(4)

In (4), the expectation  $\mathbf{E}_H$  is with respect to all possible channel states, which are governed by an FSMC specified by the state set H, the transition probability matrix A, and the initial state  $h_0$ . The function in (4) can be implemented as a lookup table for a given channel model in simulations. A closed form solution may also be possible, under some optimal coding and packet scheduling assumptions. More details for a 2-state FSMC channel analysis can be found in the appendix.

# 2.2. Summarization and packet delay constraint analysis

Let a video sequence of *n* frames be denoted by  $V = \{f_0, f_1, \ldots, f_{n-1}\}$  and its video summary of *m* frames by  $S = \{f_{l_0}, f_{l_1}, \ldots, f_{l_{m-1}}\}$ . Obviously, the video summarization process has an implicit constraint that  $0 \le l_0 < l_1 < \cdots < l_{m-1} \le n-1$ . Let the reconstructed sequence  $V'_S = \{f'_0, f'_1, \ldots, f'_{n-1}\}$  be obtained by substituting missing frames with the most recent frame that is in the summary *S*, that is,  $f'_k = f_{i=\max(l): \text{ s.t. } l \in \{l_0, l_1, \ldots, l_{m-1}\}, i \le k}$ . Let the summarization rate be

$$R(S) = \frac{m}{n},\tag{5}$$

taking values in  $\{1/n, 2/n, ..., n/n\}$ . The summarization distortion can be computed as the average frame distortion between the original sequence and the reconstructed sequence from the summary

$$D(S) = \frac{1}{n} \sum_{k=0}^{n-1} d(f_k, f'_k),$$
(6)

where  $d(f_k, f'_k)$  is the distortion of the reconstructed frame  $f'_k$  and n is the number of frames in the video sequence. Various distortion metrics can be utilized here to capture the impact of frame-loss-induced distortion,  $d(f_k, f'_k)$ . In this work, we use the Euclidean distance of scaled frames in PCA space, as discussed in [12]. This is an effective metric that matches the perception of frame losses well.

In video summarization studies [24], we also found that in addition to the average frame loss distortion metric, the maximum frame loss distortion-based metric is also very effective in matching the subjective perception, especially the jerkiness in playback. Therefore, the video summarization distortion can also be defined as

$$D(S) = \max_{k} d(f_{k}, f_{k}').$$
 (7)

The loss of frames in high activity segments of video sequence will typically result in a large D(S) in this case. The average  $(l_2)$  and maximum  $(l_{\infty})$  metrics for video summarization compliment each other in characterizing the distortion.

For the encoding of the video summary frames, we assume a constant Peak SNR (PSNR) or QP coding strategy, with frame bit budget  $B_{l_i}$  given by some rate profiler see, for example, [25]. Packets from different summary frames have different delay tolerances. Without loss of generality, we assume that the first frame of the original sequence,  $f_0$ , is always selected for the summary and intracoded with some  $B_0$  bits. The delay tolerance  $\tau_0$  is determined by how much initial streaming delay is allowed in an application. For packets generated by the summary frame  $f_{l_i}$ , with  $l_i > 0$ , if the previous summary frame  $f_{l_{j-1}}$  is decoded at time  $t_{j-1}$ , then the packet needs to arrive by the time  $t_i = t_{i-1} + t_{i-1}$  $(l_i - l_{i-1})/F$ , where F is the frame rate of the original video sequence. Therefore, the delay tolerance for frame  $f_{l_i}$  is  $\tau_{l_i} =$  $(l_i - l_{i-1})/F$ . This is a simplified delay model, not accounting for minor variations in frame encoding and other delays. The energy cost to transmit a summary S of *m* frames is therefore given by

$$E(S) = \sum_{k=0}^{m-1} E(B_{l_k}, \tau_{l_k}) = E(B_0, \tau_0) + \sum_{k=1}^{m-1} E(B_{l_k}, \tau_{l_k}), \quad (8)$$

where  $B_{l_k}$  is the number of bits needed to encode summary frame  $f_{l_k}$ , and  $\tau_{l_k}$  is the delay tolerance for frame  $f_{l_k}$ .

There are tradeoffs between the summary transmission energy cost, E(S), and the summarization distortion, D(S). The more frames selected into the summary, the smaller the summarization distortion. On the other hand, the more frames in the summary, the more bits needed to be spent in encoding the frames, and the packet arrival pattern gets more dense, which can be translated into higher bit rate and smaller delay tolerance. The transmission of more bits with more stringent deadline can incur higher transmission energy cost.

In the next subsection, we will characterize the relationship between the summarization distortion and energy cost, and formulate the energy-efficient video summarization and transmission problem as an energy-distortion (E-D) optimization problem.

#### 2.3. Energy-efficient summarization formulations

The energy-efficient summarization problem can be formulated as a constrained optimization problem. For a given constraint on the summarization distortion, we need to find the optimal summary that minimizes the transmission energy cost, while satisfying the distortion constraint,  $D_{max}$ . That is, the Minimizing Energy Optimal Summarization (MEOS) formulation is given by

$$S^* = \arg\min_{S} E(S), \quad \text{s.t. } D(S) \le D_{\max}.$$
(9)

We can also formulate the energy efficiency problem as a Minimizing Distortion Optimal Summarization (MDOS) problem. That is, for a given energy constraint,  $E_{max}$ , we want to find the optimal summary that minimizes the summarization distortion:

$$S^* = \arg\min_{S} D(S), \quad \text{s.t. } E(S) \le E_{\max}.$$
(10)

The optimal solutions to the formulations in (9) and (10) can be achieved through Dynamic Programming (DP) for the maximum frame loss distortion case in (7), by exploiting the structure of the summarization problem. As for the average distortion metric case in (6), a convex hull optimal solution can be found via Lagrangian relaxation and DP, which are discussed in more detail in the next section.

## 3. SOLUTION ALGORITHMS

Solving the constrained problems in (9) and (10) directly is usually difficult due to the complicated dependencies and large searching space for the operating parameters. For the average distortion case, we introduce the Lagrange multiplier relaxation to convert the original problem into an unconstrained problem. The solution to the original problem can then be found by solving the resulting unconstrained problem with the appropriate Lagrange multiplier that satisfies the constraint. This gradient-based approach has been widely used in solving a number of coding and resource allocation problems in video/image compression [8, 26]. For the maximum distortion case, a direct DP solution can provide us with the optimal solution at polynomial computational complexity. Finally, we introduce a heuristic algorithm that approximates the E-D performance of the optimal solutions at a fraction of the computational cost.

## 3.1. Average distortion problems

Considering the MEOS formulation with the average distortion metric in (4), by introducing the Lagrange multiplier, the relaxed problem is given by

$$S^*(\lambda) = \arg \min_{S} \{ E(S) + \lambda D(S) \}, \tag{11}$$



FIGURE 2: An example of DP trellis for the average distortion minimization problem.

in which the optimal solution  $S^*$  becomes a function of  $\lambda$ . From [27], we know that by varying  $\lambda$  from zero to infinity, we sweep the convex hull of the operational E-D function  $E(D(S^*(\lambda)))$ , which is also monotonic with respect to  $\lambda$ . Therefore, a bisection search algorithm on  $\lambda$  can give us the optimal solution within a convex hull approximation. In real-world applications, the E-D operational point sets are typically convex, and the optimal solution can indeed be found by the algorithm described above.

Solving the relaxed problem in (11) by exhaustive search is not feasible in practice, due to its exponential computational complexity. Instead, we observe that there are built-in recursive structures that can be exploited for an efficient dynamic programming solution of the relaxed problem with polynomial computational complexity.

First, let us introduce a notation on segment distortion introduced by missing frames between summary frame  $l_t$  and  $l_{t+1}$ , which is given by

$$G_{l_t}^{l_{t+1}} = \sum_{k=l_t}^{l_{t+1}-1} d(f_{l_t}, f_k).$$
(12)

Let the *state* of a video summary have *t* frames, and the last frame  $f_k$  be the minimum of the relaxed objective function given by

$$J_{t}^{k}(\lambda) = \min_{S: \text{ s.t. } |S|=t, l_{t-1}=k} \{D(S) + \lambda E(S)\}$$
$$= \min_{l_{1}, l_{2}, \dots, l_{t-2}} \left\{ G_{0}^{l_{1}} + G_{l_{1}}^{l_{2}} + \dots + G_{l_{t-2}}^{k} + G_{k}^{n} + \lambda \sum_{k=0}^{t-1} E(B_{l_{k}}, \tau_{l_{k}}) \right\},$$
(13)

where |S| denotes the number of frames in S. Note that  $l_0 = 0$ , as we assume the first frame is always selected. The

minimization process in (11) has the following recursion:

$$\begin{split} J_{t+1}^{k}(\lambda) &= \min_{S: s.t. |S|=t+1, l_{t}=k} \{D(S) + \lambda E(S)\} \\ &= \min_{l_{1}, l_{2}, \dots, l_{t-1}} \{G_{0}^{l_{1}} + G_{l_{1}}^{l_{2}} \cdots + G_{l_{t-1}}^{k} + G_{k}^{n} \\ &+ \lambda [E(B_{0}, \tau_{0}) + E(B_{l_{1}}, (l_{1} - 0)/F) \\ &+ \dots + E(B_{l_{t-1}}, (l_{t-1} - l_{t-2})/F) \\ &+ E(B_{k}, (k - l_{t-1})/F)]\} \\ &= \min_{l_{1}, l_{2}, \dots, l_{t-1}} \left\{ \underbrace{G_{0}^{l_{1}} + G_{l_{1}}^{l_{2}} \cdots + G_{l_{t-2}}^{l_{t-1}} + G_{l_{t-1}}^{n} - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_{k}^{n} \\ &+ \lambda \left[ \underbrace{E(B_{0}, \tau_{0}) + E(B_{l_{1}}, (l_{1} - 0)/F)}_{E_{t}^{l_{t-1}}} \\ &+ \dots + E(B_{l_{t-1}}, (l_{t-1} - l_{t-2})/F) \\ &+ \dots + E(B_{l_{t}, (k - l_{t-1})/F)} \right] \right\} \\ &= \min_{l_{1}, l_{2}, \dots, l_{t-1}} \left\{ D_{t}^{l_{t-1}} + \lambda E_{t}^{l_{t-1}} \\ &+ \underbrace{\lambda E(B_{k}, (k - l_{t-1})/F)}_{E_{t}^{l_{t-1}}} - G_{l_{t-1}}^{n} + G_{l_{t-1}}^{k} + G_{k}^{n} \\ &+ \underbrace{\lambda E(B_{k}, (k - l_{t-1})/F)}_{E_{t}^{l_{t-1}}} - \frac{1}{E_{t-1}^{l_{t-1}}} + \frac{1}{E_{t$$

The recursion has the initial condition given by

$$J_1^0(\lambda) = G_0^n + \lambda E(B_0, \tau_0).$$
(15)

The cost of transition is given by the edge cost  $e^{l_{t-1}k}$  in (14), which is a function of  $\lambda$ ,  $l_{t-1}$  and k as

$$e^{l_{t-1},k} = \begin{cases} \lambda E(r_k, (k-l_{t-1})/F) - G_{l_{t-1}}^n + G_{l_{t-1}}^k + G_k^n, & \text{intracoding,} \\ \lambda E(r_{k,l_{t-1}}, (k-l_{t-1})/F) - G_{l_{t-1}}^n + G_{l_{t-1}}^k + G_k^n & \text{intercoding,} \end{cases}$$
(16)

where  $r_k$  and  $r_{k,l_{t-1}}$  are the estimated bit rates obtained from a rate profiler (e.g., [25]) to intracode the frame  $f_k$ , and intercode frame  $f_k$  with backward prediction from frame  $f_{l_{t-1}}$ , respectively. The DP solution starts with the initial node  $J_1^0$ , and propagates through a trellis with arcs representing possible transitions. At each node, we compute and store the optimal incoming arc and the minimum cost. Once all nodes with the final virtual frame  $f_n$ ,  $\{J_t^n(\lambda) \mid t = 1, 2, ..., n\}$ , are computed, the optimal solution to the relaxed problem in (11) is found by selecting the minimum cost

$$S^*(\lambda) = \arg\min\left\{J_t^n(\lambda)\right\},\tag{17}$$

and backtracking from the resulting final virtual frame nodes for the optimal solution. This is similar to the Viterbi algorithm [28]. An example of a trellis for n = 5 and  $\lambda = 1.0e-4$  is shown in Figure 2, where all possible state transitions are plotted. For each state node, the minimum incoming cost is plotted as solid line, while other incoming arcs are plotted as dotted lines. For example, the node  $J_3^4$  is computed as  $J_3^4 = \min_{j \in \{1,2,3\}} \{J_2^j + e^{j,4}\}$ , and its incoming arc with the minimum cost is from node  $J_2^2$ . The virtual final frame nodes are all at the top of the trellis.

The Lagrange multiplier controls the tradeoff between summarization distortion and the energy cost in transmitting the summarized video frames. By varying the value of  $\lambda$  and solving the relaxed problem in the inner loop, we can obtain the optimal solution that minimizes the transmission energy cost while meeting certain distortion constraints. Since the operational energy-distortion function  $E(D(S^*(\lambda)))$  is monotonic with respect to  $\lambda$ , a fast bisection search algorithm can be applied to find the optimal  $\lambda^*$ , which results in the tightest bound on the distortion constraint  $D_{\text{max}}$ , that is,  $D(S^*(\lambda^*))$  is the closest to  $D_{\text{max}}$ . The algorithm can perform even faster by reusing the distortion and energy cost results that only need to be computed once in the iteration. The solution to the MEOS formulation can also be solved in the same fashion.

The complexity of the optimal inner loop solution is polynomial in frame number n, and the outer loop bisection search complexity depends on the choice of initial search window size and location. But overall, for small n < 60, the complexity can be well handled by mobile devices with more powerful modern processors.

## 3.2. Maximum distortion problems

When the maximum distortion metric in (6) is used, the problem has a simpler structure due to less complex dependencies. Let us consider the MEOS problem first. The objective here is to minimize the energy cost of transmitting a segment of the video summary, with the given constraint on the maximum frame distortion allowed. Unlike the complicated structures in the average distortion case, this given distortion constraint can be used to prune the infeasible edges in the summary state trellis similarly to the previous case, and then a search and back tracking algorithm can be derived.

Let us define the summarization distortion for the video segment between video summary frames  $l_t$  and  $l_{t+1}$  as

$$D_{l_t}^{l_{t+1}} = \max_{j \in [l_t, l_{t+1}-1]} d(f_{l_t}, f_j).$$
(18)

This is the maximum frame distortion between the previous summary frame  $l_t$ , and the subsequent missing frames before

the next summary frame  $l_{t+1}$ . It is clear that the placement of summary frames will have a major impact on the resulting video summary distortion. Generally, the larger the distance between the two summary frames  $l_t$  and  $l_{t+1}$ , the larger the resulting distortion. Where the summary frames are placed is also important. For example, if the summary frames  $l_t$  and  $l_{t+1}$  astride two different video shots, there will be a spike in the distortion  $D_{l_t}^{l_{t+1}}$ .

A frame loss distortion larger than  $D_{\text{max}}$  is not allowed in this case; we can reflect this constraint by defining the energy cost for the segment as

$$E_{l_t}^{l_{t+1}} = \begin{cases} E(B_{l_{t+1}}, (l_{t+1} - l_t)/F), & \text{if } D_{l_t}^{l_{t+1}} \le D_{\max}, \\ \infty, & \text{otherwise.} \end{cases}$$
(19)

With this, any summary frame selections with resulting segment distortion greater than  $D_{\text{max}}$  are excluded from the MEOS solution.

For the maximum energy minimization problem, let us also explore the structure of the energy cost of the optimal video summary solution ending with frame  $l_t$ :

$$E_{l_t} = \min_{l_1, l_2, \dots, l_{t-1}} \left\{ E_0^{l_1} + E_{l_1}^{l_2} + \dots + E_{l_{t-1}}^{l_t} \right\}.$$
 (20)

This includes any combination of choices of summary frames between  $f_0$  and  $f_{l_t}$ . Similarly to the relaxed cost case in average distortion minimization, it also has a recursive structure as

This recursive relationship is illustrated by an example in Figure 3. A small scale problem with n = 6 frames from the "foreman" sequence is considered. The  $D_{\text{max}}$  is 15 in this case, which prunes out  $[l_t, l_{t+1}]$  summary segments that have resulting distortion  $D_{l_t}^{l_{t+1}} > D_{\text{max}}$ . The optimal solution is therefore found by searching through all feasible transitions in energy cost trellis, recording the minimum energy cost arcs as we compute the next stage in trellis expansion, and then backtracking for the optimal solution in a Viterbi algorithmic fashion [28]. The optimal summary for the problem in Figure 3 consists of frames  $f_0$  and  $f_4$ .

Notice that the summary found is optimal, as compared with the convex-hull approximately optimal in the average distortion case. The resulting distortion  $d(f_k, f'_k)$ has interesting patterns as shown in Figure 4, for the 120frame "foreman" sequence segment (frames 120~249). The



FIGURE 3: An example of DP trellis for the max distortion minimization problem.



FIGURE 4: MEOS summary example.

distortion threshold  $D_{\text{max}} = 12$ , and the resulting summary consists of 45 frames.

Figure 4(a) is the sequence activity level profile as differential frame distance,  $d(f_k, f_{k-1})$ , and the summary frame selections are plotted in red vertical lines. Figure 4(b) is the summary distortion plot  $d(f_k, f'_k)$ . Notice that the placement of summary frames brings the maximum distortion for each segment below  $D_{\text{max}}$  indeed. The density of the summary frames also reflects well the activity level in the sequence, as expected.

To solve the maximum distortion minimization problem, instead of searching on the Lagrange multiplier as in the average distortion case, we develop a bisection search algorithm that searches on the maximum distortion constraint,  $D_{max}$ , in the outer loop, and in the inner loop, and solves the MEOS problem as a function of the threshold  $D_{max}$ , that is,

$$S^*(D_{\max}) = \arg\min_{s} E(S), \quad \text{s.t. } D(S) \le D_{\max}.$$
 (22)

To find the minimum distortion summary that meets the given energy constraint  $E_{max}$ , the bisection search stops when the resulting energy cost  $E(S^*(D_{max}))$  is the closest to the  $E_{max}$ . This is similar to the Lagrangian relaxation and DP solution to the average distortion case in structure.

### 3.3. Heuristic greedy solution

The DP solution has polynomial computational complexity  $O(n^2)$ , with *n* the number of frames in the sequence, which may not be practical for mobile devices that usually have limited power and computation capacity. A heuristic solution is thus developed to generate energy-efficient video summaries for both average and maximum distortion cases.

The heuristic algorithm selects the summary frames such that all summarization distortion segments  $G_{l_{t-1}}^{l_t}$ ,

$$G_{l_{l}}^{l_{l+1}} \begin{cases} \sum_{k=l_{l}}^{l_{l+1}-1} d(f_{l_{l}}, f_{k}), & \text{avg distortion,} \\ \max_{k \in [l_{l}, l_{l+1}-1]} d(f_{l_{l}}, f_{k}), & \text{max distortion,} \end{cases}$$
(23)

between successive summary frames satisfy  $G_{l_{l-1}}^{l_t} \leq \Delta$ , for a preselected step size  $\Delta$ . Notice that this applies to both average and maximum distortions. The algorithm is greedy and operates in an one-pass fashion for a given  $\Delta$ . The pseudocode of the proposed heuristic algorithm is then shown in Algorithm 1.

This replaces the DP algorithm in the optimal solution, and a bisection search on  $\Delta$  can find the solution that satisfies the summarization distortion or the energy cost constraints. The computational complexity is O(n) for the greedy algorithm solution. Simulation results with both the optimal and the heuristic algorithms are presented and discussed in Section 4.

#### 4. SIMULATION RESULTS

To simulate a slow fading wireless channel, we model the channel fading as a two-state FSMC with channel states  $h_0$  and  $h_1$ . The channel has transition probabilities, p and q, for state transition from  $h_0$  to  $h_1$ , and  $h_1$  to  $h_0$ , respectively, and the channel state transitional probability is given by  $A = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$ . The steady-state channel state probability is therefore computed as  $\pi_0 = q/(p+q)$  and  $\pi_1 = q/(p+q)$ . Assuming that the deadline  $\tau$  is much greater than the channel coherent time,  $T_c$ , that is,  $\tau \gg T_c$ , and the signaling rate is W (W is selected to simulate typical SNR operating range in wireless communications), then out of the total  $2W\tau$  channel uses,  $(p/(p+q))2W\tau$  are in channel state  $h_1$ .

Assuming that the channel state is known to both the transmitter and the receiver, with the optimal coding and packet scheduling, then the expected energy cost of transmitting *B* bits with delay constraint  $\tau$  can then be computed as

$$E(B, \tau) = \mathbf{E}_{H} \{ E_{b}(2W\tau/B, h)B \}$$
  
=  $\min_{0 \le z \le 1} \{ f(z; B, W, \tau, p, q, h_{0}, h_{1}) \}$   
=  $\min_{0 \le z \le 1} \left\{ zBE_{b} \left( \frac{q}{p+q} 2W\tau/(zB), h_{0} \right) + (1-z)BE_{b} \left( \frac{p}{p+q} 2W\tau/(B(1-z)), h_{1} \right) \right\}.$  (24)

In (24), we need to find an optimal bits splitting factor, z in  $\begin{bmatrix} 0 & 1 \end{bmatrix}$ , of the total bits B, with zB bits transmitted optimally while the channel state is  $h_0$ , and (1 - z)B bits transmitted optimally while the channel state is  $h_1$ .

Note that (24) can be implemented as a lookup table in a practical system with more complex channel models. For simple channel models such as the two-state FSMC, a closed form solution can be derived. Once the conditions based on the first- and second-order derivatives (see the appendix for more detail) are satisfied for the minimization problem in (24), the optimal splitting of the bits is given by

$$z^* = \frac{w\tau pq}{B(p+q)^2} \left[ \log_2\left(\frac{h_0}{h_1}\right) + \frac{(p+q)}{w\tau p}B \right]$$
  
$$= \frac{w\tau pq}{B(p+q)^2} \log_2\left(\frac{h_0}{h_1}\right) + \frac{q}{(p+q)},$$
 (25)

and the minimum energy cost is given by

$$E(B,\tau) = f(z^*; B, W, \tau, p, q, h_0, h_1)$$
  
=  $z^* BE_b \left( \frac{q}{p+q} 2W\tau/(z^*B), h_0 \right)$   
+  $(1-z^*)BE_b \left( \frac{p}{p+q} 2W\tau/(B(1-z^*)), h_1 \right).$  (26)

Equation (26) can be implemented as a lookup table for the energy-distortion optimization algorithm.

The performance of the proposed algorithms has been studied in experiments as well. Some representative results are presented next. The implementation of the algorithms was done with a mix of C and Matlab.

In Figure 5, the QCIF-sized "foreman" sequence (frames  $150 \sim 299$ ) was utilized. The channel state is modeled as  $h_0 = 0.9$ ,  $h_1 = 0.1$ , p = 0.7, q = 0.8. Signaling rate is set as W = 20 kHz. The background noise power is assumed to be N = 1 mJ per channel use. The summary frames are intracoded

$L = 0; S = \{f_0\}.$	% select 1 <sup>st</sup> frame
For $k = 1$ : $n - 1$	04 check the comment distortion value
$S = S + \{f_k\}$	% check the segment distortion value
L = k	
End	
End	

ALGORITHM 1: Heuristic algorithm pseudo code.



FIGURE 5: Examples of energy-efficient video summarization for the average distortion case.

with constant PSNR quality using the H.263 codec based on the TMN5 rate control. Summarization distortion and average power during transmissions are plotted for two different values of the Lagrange multiplier, with  $\lambda_1 = 1.0e-5$ and  $\lambda_2 = 6.0e-5$ . For larger Lagrange multiplier,  $\lambda_2$ , more weight is placed on minimizing the energy cost, therefore the associated energy cost (area under the average power plot) is smaller than that of a smaller value  $\lambda_1$ . On the other hand, the summarization distortion is larger for  $\lambda_1$  than for  $\lambda_2$ , as expected.

In the second set of experiments, the overall performance is characterized as the E-D and Energy-Rate (E-R) curves in Figures 6(a) and 6(b), respectively, for both W = 10 kHz and 20 kHz, as well as inter- and intracoding cases. Figure 6(a) characterizes the relationship between the summarization

TABLE 1: Computational complexity of the DP solution.

<i>n</i> = 150	<i>n</i> = 120	<i>n</i> = 90	<i>n</i> = 60	<i>n</i> = 45	<i>n</i> = 30
$t = 15.47 \mathrm{s}$	$t = 9.82 \mathrm{s}$	$t = 5.78  \mathrm{s}$	$t = 2.78  \mathrm{s}$	$t = 1.59  \mathrm{s}$	t = 0.6  s

TABLE 2: Energy-summary quality tradeoff subjective evaluation.

Summary name	λ	R(S)	D(S)	E(S)
"S1.263"	4.8e - 08	0.80	06.32	7.55e + 08
"S2.263"	2.0e - 07	0.68	09.75	2.62e + 08
"S3.263"	6.0e - 07	0.55	13.14	1.18e + 08
"S4.263"	3.0e - 06	0.39	18.91	4.46e + 07
"S5.263"	1.0e - 05	0.26	29.08	1.44e + 07
"S6.263"	1.0e - 04	0.12	49.68	2.53e + 06

distortion and the total energy cost in  $\log_{10}(mJ)$  scale. As the summarization distortion goes up linearly, the energy cost drops exponentially. Figure 6(b) characterizes the relationship between the energy cost and the summarization rate. In the typical operating range of the video summarization, for example, R(S) = [0.1, 0.9], the energy cost can change from 2 to 6 orders of magnitude. This clearly indicates that summarization can be an effective energy conserving scheme for wireless video communications.

The E-D performance for the maximum distortion metric is also summarized in Figure 7 for the optimal DP and greedy algorithms. Notice that the greedy solution performs closer to the optimal solution in this case.

The computational complexity of the DPsolution is indeed significantly larger than that of the greedy solution, especially as the size of the problem becomes larger. The execution times for the DP algorithm for various video segment lengths are summarized in Table 1.

These results are obtained with nonoptimized Matlab code running on a 2.0 GHz Celeron PC. Notice that the average execution time for the greedy algorithm is 0.11 s on the same computer for n = 150.

In Table 2 the summary rate, distortion, and energy cost are shown for various values of the Lagrange multiplier, along with the corresponding names of the summary sequences (based on the same 150-frame "foreman" sequence segment, intercoding, with W = 10 kHz) generated with the optimal DP algorithm. The sequences are also available for subjective evaluation of the tradeoffs between visual quality and energy cost in transmitting the sequence.



FIGURE 6: Energy-distortion performance for the average distortion minimization case.

Based on the visual evaluation of the results in Table 2, the graceful degradation of the video summary visual quality is clearly demonstrated. As the Lagrange multiplier value increases, more weight is placed on the energy cost during minimization. In the typical operating range of 0.12 to 0.80 for the video summarization rate, the energy cost differs by a factor of around 300 times. This demonstrates that video summarization is indeed an effective energy conservation scheme for wireless video streaming applications.

## 5. CONCLUSION AND FUTURE WORK

In this work, we formulated the problem of energy-efficient video summarization and transmission and proposed an

optimal (within a convex hull approximation) algorithm for solving it. The algorithm is based on Lagrangian relaxation and dynamic programming in the average distortion metric case, and bisection search on distortion threshold and dynamic programming in the maximum distortion metric case. A heuristic algorithm to reduce the computational complexity has also been developed. The simulation results indicate that this is a very efficient and effective method in energy-efficient video transmission over a slow fading wireless channel.

The next step of the work is to have more realistic channel models for commercially deployed wireless systems, for example, WiMAX, and consider a multiuser setup and exploit diversity gains among users.



FIGURE 7: Energy-distortion performance for the maximum distortion case.

## **APPENDIX**

## DERIVATION OF THE OPTIMAL SPLIT IN TRANSMISSION

Assuming the channel state is known to both the transmitter and the receiver, the expected energy cost of transmitting *B* bits with delay  $\tau$  is computed as

$$E(B, \tau) = \mathbf{E}_{H} \{ E_{b}(2W\tau/B, h)B \}$$
  
=  $\min_{0 \le z \le 1} \{ f(z; B, W, \tau, p, q, h_{0}, h_{1}) \}$   
=  $\min_{0 \le z \le 1} \left\{ zBE_{b} \left( \frac{q}{p+q} 2W\tau/(zB), h_{0} \right) + (1-z)BE_{b} \left( \frac{p}{p+q} 2W\tau/(B(1-z)), h_{1} \right) \right\}.$   
(A.1)

Consequently, we have

$$f(z) = zBE_b(2W\tau\pi_0/(zB), h_0) + (1-z)BE_b(2W\tau\pi_1/((1-z)B), h_1) = (2\pi_0W\tau/h_0)(2^{zB/\pi_0W\tau} - 1) + (2\pi_1W\tau/h_1)(2^{(1-z)B/\pi_1W\tau} - 1).$$
(A.2)

Let

$$a_0 = 2\pi_0 W \tau / h_0,$$
  $a_1 = 2\pi_1 W \tau / h_1,$   
 $b_0 = \frac{B}{\pi_0 W \tau},$   $b_1 = \frac{B}{\pi_1 W \tau}.$  (A.3)

We have  $f(z) = a_0(2^{b_0 z} - 1) + a_1(2^{b_1(1-z)} - 1)$ . To minimize f(z), let the first-order derivative be zero, which leads to

$$f'(z) = a_0 b_0 \ln(2) 2^{b_0 z} - a_1 b_1 \ln(2) 2^{b_1(1-z)}$$
  
= 0,  $\implies z^* = \frac{1}{b_0 + b_1} \left( \log_2 \left( \frac{a_1 b_1}{a_0 b_0} \right) + b_1 \right).$   
(A.4)

Because the second-order derivative is always nonnegative as below

$$f''(z) = a_0 b_0^2 \ln^2(2) 2^{b_0 z} + a_1 b_1^2 \ln^2(2) 2^{b_1(1-z)} \ge 0, \quad \forall 0 \le z \le 1,$$
(A.5)

the optimal bit splitting ratio is then

$$z^* = \pi_0 \pi_1 \log_2 \left(\frac{h_0}{h_1}\right) \frac{W\tau}{B} + \pi_0,$$
 (A.6)

and the optimal energy cost is given by

$$E(B,\tau) = z^* B E_b (2\pi_0 W \tau / (z^* B), h_0) + (1 - z^*) B E_b (2\pi_1 W \tau / (B(1 - z^*)), h_1).$$
(A.7)

## ACKNOWLEDGMENT

Part of this work was presented at SPIE VCIP 2005.

### REFERENCES

- Wireless LAN Medium Access Control (MAC) Physical Layer (PHY), Specification of IEEE 802.11 Standard, 1998.
- [2] R. Kravets and P. Krishnan, "Application-driven power management for mobile communication," *Wireless Networks*, vol. 6, no. 4, pp. 263–277, 2000.
- [3] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.
- [4] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information The*ory, vol. 45, no. 5, pp. 1468–1489, 1999.
- [5] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '02)*, vol. 3, pp. 1773–1782, New York, NY, USA, June 2002.
- [6] E. Uysal-Biyikoglu, B. Prabhakar, and A. El Gamal, "Energyefficient packet transmission over a wireless link," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 487–499, 2002.
- [7] Y. S. Chan and J. W. Modestino, "Transport of scalable video over CDMA wireless networks: a joint source coding and power control approach," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 973–976, Thesaloniki, Greece, October 2001.
- [8] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy-efficient wireless video communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 411–424, 2002.

- [9] Z. He, J. Cai, and C. W. Chen, "Joint source channel ratedistortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 511–523, 2002.
- [10] I.-M. Kim and H.-M. Kim, "An optimum power management scheme for wireless video service in CDMA systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 1, pp. 81– 91, 2003.
- [11] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy-efficient wireless video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1710– 1720, 2003.
- [12] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Rate-distortion optimal video summary generation," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1550–1560, 2005.
- [13] N. D. Doulamis, A. D. Doulamis, Y. S. Avrithis, and S. D. Kollias, "Video content representation using optimal extraction of frames and scenes," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 875–879, Chicago, Ill, USA, October 1998.
- [14] A. Hanjalic and H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1280–1289, 1999.
- [15] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 90–105, 2002.
- [16] R. Lienhart, "Reliable transition detection in videos: a survey and practioner's guide," *International Journal of Image and Graphics*, vol. 1, no. 3, pp. 469–486, 2001.
- [17] H. Sundaram and S.-F. Chang, "Constrained utility maximization for generating visual skims," in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (*CBAIVL '01*), pp. 124–131, Kauai, Hawaii, USA, December 2001.
- [18] Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, "Adaptive key frame extracting using unsupervised clustering," in *Proceedings* of the IEEE International Conference on Image Processing (ICIP '98), vol. 1, pp. 866–870, Chicago, III, USA, October 1998.
- [19] Z. Li, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, "Bit constrained optimal video summarization," in *Proceedings of the IEEE International Conference on Image Processing (ICIP* '04), Singapore, October 2004.
- [20] Z. Li, F. Zhai, A. K. Katsaggelos, and T. N. Pappas, "Energyefficient video summarization and transmission over a slow fading wireless channel," in *Image and Video Communications* and Processing, vol. 5685 of Proceedings of SPIE, pp. 940–948, San Jose, Calif, USA, January 2005.
- [21] Z. Li, F. Zhai, and A. K. Katsaggelos, "Video summarization for energy-efficient wireless streaming," in *Visual Communications* and Image Processing, vol. 5960 of Proceedings of SPIE, pp. 763– 774, Beijing, China, July 2005.
- [22] H. S. Wang and N. Moayeri, "Finite-state Markov channela useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163– 171, 1995.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunication, John Wiley & Sons, New York, NY, USA, 1991.
- [24] Z. Li, G. M. Schuster, and A. K. Katsaggelos, "MINMAX optimal video summarization," *IEEE Transactions on Circuits*

and Systems for Video Technology, vol. 15, no. 10, pp. 1245-1256, 2005.

- [25] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1221– 1236, 2001.
- [26] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression, Optimal Video Frame Compression and Object Boundary Encoding*, Kluwer Academic Publishers, Norwell, Mass, USA, 1997.
- [27] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 160–175, 1993.
- [28] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

## **Research Article**

## Optimal JPWL Forward Error Correction Rate Allocation for Robust JPEG 2000 Images and Video Streaming over Mobile Ad Hoc Networks

# Max Agueh,<sup>1</sup> Jean-François Diouris,<sup>1</sup> Magaye Diop,<sup>2</sup> François-Olivier Devaux,<sup>3</sup> Christophe De Vleeschouwer,<sup>3</sup> and Benoit Macq<sup>3</sup>

<sup>1</sup> Institut de Recherche en Electrotechnique et Electronique de Nantes Atlantique (IREENA), Equipe Communications Numériques et Radiofréquences, Rue Christian Pauc, La chantrerie, BP 50609, 44306 Nantes cedex 3, France

<sup>2</sup> Ecole Supérieure Polytechnique, Université Cheikh Anta Diop de Dakar (UCAD), BP 5085 Dakar, Senegal

<sup>3</sup> Communications and Remote Sensing Laboratory, FSA/TELE, Bâtiment Stévin, Place du Levant 2,

B-1348 Louvain-la-Neuve, Belgium

Correspondence should be addressed to Max Agueh, max.agueh@gmail.com

Received 1 October 2007; Revised 12 February 2008; Accepted 26 April 2008

Recommended by Jianfei Cai

Based on the analysis of real mobile ad hoc network (MANET) traces, we derive in this paper an optimal wireless JPEG 2000 compliant forward error correction (FEC) rate allocation scheme for a robust streaming of images and videos over MANET. The packet-based proposed scheme has a low complexity and is compliant to JPWL, the 11th part of the JPEG 2000 standard. The effectiveness of the proposed method is evaluated using a wireless Motion JPEG 2000 client/server application; and the ability of the optimal scheme to guarantee quality of service (QoS) to wireless clients is demonstrated.

Copyright © 2008 Max Agueh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Nowadays, there is an increasing demand of multimedia applications which integrate wireless transmission functionalities. Wireless networks are suitable for those types of applications, due to their ease of deployment and because they yield tremendous advantages in terms of mobility of user equipment (UE). However, wireless networks are subject to a high level of transmission errors because they rely on radio waves whose characteristics are highly dependent on the transmission environment.

In wireless video streaming applications like the one considered in this paper (Figure 1), effective data protection is a crucial issue.

JPEG 2000, the newest image representation standard completing the existing JPEG standard [1], addresses this issue. Part 1 of this standard defines several tools allowing the decoder to detect errors in the transmitted codestream, and to resynchronize the decoding in order to avoid erroneous decoding and crashes. Even if these tools give a first level of protection from transmission errors, they become ineffective when the transmission channel experiences high bit error rate. To overcome this limitation, wireless JPEG 2000 (JPWL, JPEG 2000 11th part) defines techniques to increase the resilience of the codestream to transmission errors in wireless systems. JPWL specifies error resilience tools such as forward error correction (FEC), interleaving, and unequal error protection.

In [2], the description of the JPWL system is presented and the performance of its error protection block (EPB) is evaluated. A fully JPEG 2000 part 1 compliant backward compatible error protection scheme is proposed in [3]. A memoryless binary symmetric channel (BSC) is used for simulations both in [2, 3]. However, as packets errors mainly occur in bursts, the channel model considered in these works is not realistic. Moreover, JPEG 2000 codestream interleaving is not considered in [3].

In this paper we present a wireless JPEG 2000 images/video streaming system based on the recommendations of JPWL final draft [4]. To the best of our knowledge, the present work is the first to rely on an analysis of real 802.11 data traces and to derive an optimal JPWL



802.11 ad hoc network

FIGURE 1: Wireless video streaming system.

compliant FEC rate allocation method for robust JPEG 2000 images/video streaming over wireless channel. It is worth noting that the performance of this method is evaluated using a Motion JPEG 2000 video streaming application over real MANET channel traces.

The paper is arranged as follows. In Section 2, the proposed JPWL-based system is described. Section 3 is dedicated to the analysis and modeling of real MANET channel traces. In Section 4, the FEC rate allocation problem is formalised, and an optimal FEC rate allocation method is proposed. In Section 5, experimental results are derived from JPEG 2000 frames transmission over wireless channel traces. Finally, some conclusions are provided in Section 6.

## 2. A WIRELESS JPEG 2000 IMAGES/VIDEO STREAMING SYSTEM

### 2.1. System functionalities

The functionalities of the proposed JPWL-based system are presented in Figure 2. The aim of this system is to efficiently transmit a Motion JPEG 2000 (MJ2) video sequence through MANET channel traces.

The system is described as follows.

The input of the JPWL codec is a Motion JPEG 2000 (MJ2) file. The JPEG 2000 codestreams included in the MJ2 file are extracted and indexed.

These indexed codestreams are transmitted to the JPWL encoder ([4] presents a more accurate description of the used JPWL encoder) which applies FEC at the specified rate and adds the JPWL markers in order to make the codestream compliant to wireless JPEG 2000 standard. At this stage, frames are still JPEG 2000 part 1 compliant, which means that any JPEG 2000 decoder is able to decode them.

To increase JPWL frames robustness, an interleaving mechanism is processed before each frame transmission through the error-prone channel. This is a recommended mechanism for transmission over wireless channel where errors occur in burst (contiguous long sequence of errors). Thanks to interleaving, the correlation between error sequences is reduced.

The interleaving step is followed by RTP packetization. In this process, JPEG 2000 codestream data and other types of data are integrated into RTP packets as described in [5].

RTP packets are then transmitted through the wireless channel which is modelled in this work by a Gilbert channel model. This channel model will be further presented in Section 3.2.

At the decoder side, after depacketization, the JPWL decoder corrects and decodes the received JPWL codestreams and rebuilds the JPEG 2000 frames. At this stage, parameters



FIGURE 2: JPWL-based system functionalities.

such as packet error rate (PER) are extracted, increasing the knowledge of the channel state. The decoder sends extracted parameters back to the JPWL encoder via the Uplink.

The last process of the transmission chain is the evaluation of the peak signal-to-noise ratio (PSNR) which measures the distortion between the transmitted and the decoded image/video.

## 2.2. JPEG 2000 codestreams transmission over the proposed JPWL system

Figure 3 presents the structure of JPEG 2000 codestreams when transmitted through our proposed JPWL system.

After the indexation of the Motion JPEG 2000 file, the original JPEG 2000 codestreams are introduced in the system. Then, our FEC rate allocation scheme selects the optimal Reed-Solomon codes and calculates the resulting JPWL protection headers. In Figure 3 this step corresponds to the JPWL protection, where redundant data are added to original codestreams. Protected data are then interleaved in order to reduce the impact of transmission errors (interleaving process). A detailed description of the interleaving process is presented in Section 5.1. Interleaved data are then RTP-packetized (Figure 3). In this work, we do not assume a particular RTP packetization scheme. It is worth noting that Futemma et al. proposed in [6] an RTP payload format for JPEG 2000 streams. This work under progress defines an intelligent JPEG 2000 packets fragmentation into RTP payload for robust images/video streaming. An interesting extension to our work could be to integrate this new RTP packetization scheme in our proposed system. In our system, we do not emphasize a cross-layer approach meaning that channel errors are handled at lower layers and are not



FIGURE 3: JPEG 2000 codestreams transmission through the proposed JPWL system.

transmitted to upper layers. Thus, only correctly received data packets are transmitted to the application layer.

RTP packets are transmitted through a wireless channel subject to losses (in Figure 3, packet 8 is corrupted). At

the receiver side, RTP packets are depacketized and the extracted data are de-interleaved. At the following step (JPWL correction), redundant data are used to correct the corrupted part of the codestream. After JPWL correction, the

transmitted codestreams are recovered and can be compared to the original codestreams.

As a better knowledge of the characteristic of the wireless channel can significantly improves the design of the FEC rate allocation mechanism, we dedicate the following section to the analysis and modeling of real MANET channel traces.

## 3. ANALYSING AND MODELING MANET CHANNEL TRACES

In this section we analyze loss patterns of a mobile ad hoc network channel and derive application level models to emulate transmission error occurrences in the considered system. We first describe the loss pattern generation scenario and then focus our study on modeling these patterns with Gilbert model based on first-order Markov chains.

The interest of this section is to derive conclusions on accurate transmission errors modeling at application level. The generated models allow refinement of error protection strategies.

## 3.1. MANET loss patterns generation

The platform used to generate the loss patterns is presented in Figure 4. It consists of a client/server software pair running on two Windows XP laptops connected in ad hoc network using two PCMCIA IEEE 802.11 b/g cards (at 2,4 GHz). As the platform only contains two laptops, no collision occurs with other stations.

The set of generated loss patterns covers different transmission scenarios (mobile or static). Each pattern corresponds to a specific carrier-to-noise ratio C/N (C/N is the ratio between the desired signal and the total received noise power).

The mode used at the physical layer of the wireless link is the mode 4 where the modulation is QPSK. The coding rate is 3/4 and the nominal data rate  $R_{\text{Nominal}}$  is 18 Mbps. In the considered loss patterns, *C/N* varies between 20 dB and 11 dB, which corresponds to a packet error rate ranging from  $5.1 \times 10^{-3}$  to  $2.662 \times 10^{-1}$ . Generated traces are available in [7].

### 3.2. Modeling loss patterns with Gilbert model

## 3.2.1. Gilbert model

The Gilbert model was first introduced by Gilbert in [8]. Elliot proposes an extension of the Gilbert model in [9], the last model is commonly known as Gilbert-Elliot (GE). In GE model, the modeled wireless channel has two states: good and bad. In the good state (g), the channel provides a constant and low error probability ( $P_G$ ); whereas in the bad state (b), the channel experiences a high error probability ( $P_B$ ). Hence we have  $P_G \ll P_B$  for GE,  $P_G = 0$  and  $P_B = 1$  for the Gilbert channel. In other words Gilbert model is a simplified GE model.

In this work, we use an 8-bit symbol oriented Gilbert model to emulate the correlated error characteristics of wireless channel. Therefore, our wireless channel is modeled as a two-state Markov process (Figure 5).



FIGURE 4: Loss patterns generation platform.



FIGURE 5: Two-state Markov process scheme.

With this model, the channel produces error bursts because when in bad state, the probability of staying in this state is greater than the probability of returning to good state.

In Markov chains with finite state space, the transition probability distribution can be represented by a matrix called transition matrix *P*. The  $(i, j)^{ieme}$  element of *P* is  $P(X_{n+1} = j/X_n = i)$  with  $i, j \in \{0, 1\}$ . Hence the transition matrix of the model presented in Figure 5 is

$$P = \begin{bmatrix} p_{gg} & p_{bg} \\ p_{gb} & p_{bb} \end{bmatrix} = \begin{bmatrix} p_{gg} & 1 - p_{bb} \\ 1 - p_{gg} & p_{bb} \end{bmatrix}.$$
 (1)

From *P* we derive the stationary distribution  $\pi = [\pi_G \ \pi_B]$  which satisfies the condition  $\pi \cdot P = \pi$ :

$$\pi_{G} = \frac{1 - p_{bb}}{1 - p_{bb} + 1 - p_{gg}},$$

$$\pi_{B} = \frac{1 - p_{gg}}{1 - p_{bb} + 1 - p_{gg}}.$$
(2)

Let  $L_G$  and  $L_B$  be respectively the mean length of error free and erroneous sequences, then we have

$$L_G = \frac{1}{1 - p_{gg}}, \qquad L_B = \frac{1}{1 - p_{bb}}.$$
 (3)



FIGURE 6: Error bursts distribution.

Applying Markov process at symbol level, the symbol error rate (SER) for GE is

SER = 
$$P_G \pi_G + P_B \pi_B = \frac{P_G (1 - p_{bb}) + P_B (1 - p_{gg})}{(1 - p_{bb} + 1 - p_{gg})}.$$
 (4)

For the Gilbert model, we have  $P_G = 0$  and  $P_B = 1$ , so the SER is given by

SER = 
$$\frac{1 - p_{gg}}{1 - p_{bb} + 1 - p_{gg}}$$
. (5)

A comprehensive description of Markov-based wireless channel modeling is available in [10].

#### 3.2.2. Traces analysis under Gilbert framework

It is worth noting that in the considered traces, each RTP packet has a fixed length of 1128 symbols (bytes). Hence, in our case the symbol error rate (SER) is equal to the packet error rate (PER). Therefore, packet oriented Gilbert models derived from our traces have the same characteristics and same parameters as the 8-bit symbol oriented Gilbert models used to emulate the wireless channel at application level. As loss patterns are applied on RTP packets, we present a packet oriented analysis of the traces.

In the loss patterns, good state (G) and bad state (B) are represented, respectively, by 0 and 1. Hence 0 corresponds to a well-received RTP packet and 1 to an erroneous packet.

The distribution of error burst length is presented in Figure 6 for different loss patterns.

From Figure 6 we notice that the error burst length is often less than 10 packets. So we consider  $L_B^{\text{max}} = 10$  as the upper bound of the error burst length.



FIGURE 7: Error-free burst length distribution.

By evaluating the error-free burst length distribution in Figure 7, we show that the upper bound  $L_G^{\text{max}} = 100$  is ten times higher than the error burst length upper bound. This is due to the fact that despite in case where the wireless channel experiences fading (burst of errors), the transmission is often successful.

The number of error-free bursts is lower than the number of error bursts, but this gap is compensated by the time spent in error-free state (error-free burst length) which is much longer than the one in error state (error burst length). So in our models, the mean time in the good state G should be sensibly greater than the mean time in the bad state B.

We rely on this analysis to derive accurate Gilbert model parameters  $p_{gb}$  and  $p_{bg}$  using the relation verified by Jain [11]:

$$p_{gb} = \frac{1}{L_G}, \qquad p_{bg} = \frac{1}{L_B}.$$
 (6)

This analysis allows a better characterization of transmission errors, improving by the way the design of the FEC rate allocation scheme.

## 4. OPTIMAL FORWARD ERROR CORRECTION RATE ALLOCATION

Making an analogy between the FEC rate allocation problem and the multiple choice Knapsack problem (MCKP) leads to the conclusion that both problems are NP-hard. Hence, most of the algorithms proposed in the literature such as the one presented by Thomos et al. [12] lead to exhaustive search among different FEC rate solutions, exponentially increasing their complexity. These algorithms are thus interesting for an offline video streaming but are unpractical for real-time applications.

To overcome this limitation, Guo et al. proposed in [13] a slightly complex layered unequal error protection scheme for robust Motion JPEG 2000 streaming over wireless network. However, this algorithm is not JPWL compliant and was designed based on the assumption that the channel is a memoryless binary symmetric channel (uncorrelated error occurrence) which is not realistic because wireless channels have correlated errors sequence. Hence, we have proposed in [14] a dynamic layer-based unequal error protection FEC rate allocation methodology for efficient JPEG 2000 streaming over MANET. The proposed scheme improved the performance by about 10% compared to a priori selection of channel coding. However the main drawback of both methodologies is that the FEC rate allocation is suboptimal. In fact, in both schemes the protection strategy is layerbased which implies that a selected FEC rate is applied to all the substreams belonging to the same layer. This limits the effectiveness of those protection strategies especially for fast varying channels where the selected FEC rate may need to be updated from one substream to another.

In this paper we propose a slightly complex, packet-based optimal FEC rate allocation algorithm for robust Motion JPEG 2000 video streaming over wireless channel.

In Section 4.1 we formalize the FEC rate allocation problem and introduce in Section 4.2 the initial incremental reduction of distortion  $(RD_i^0)$  associated to the decoding of packet *i*. This metric is of central importance in our scheme and is derived from the JPEG 2000 encoding scheme. Section 4.3 introduces evaluation of the decoding error probability when using *t*-error correcting Reed-Solomon codes to protect JPEG 2000 codestreams.

We then present the proposed optimal FEC allocation algorithm in Section 4.4.

## 4.1. Problem formalization

The goal is to optimally protect JPEG 2000 images/video for robust streaming over wireless channel.

Considering that JPEG 2000 codestreams are constituted by a set of *S* substreams, the optimal FEC allocation problem can be resumed by answering the question of how to optimally protect each substream so as to minimize the transmitted image distortion under a rate constraint determined by the available bandwidth in the system.

Since the JPEG 2000 standard specifies that packets are byte-aligned, it is especially interesting to work with Galois field  $GF(2^8)$  to provide error correction capabilities. In this context, JPWL final draft [4] recommends the use of Reed-Solomon (RS) codes as FEC codes and fixes a set of RS default codes for substream protection before transmission over wireless channels.

Let  $\gamma$  be a substream protection level selected in the range  $0 \le \gamma \le \gamma_{max}$ , each protection level corresponds to a specific RS code selected between JPWL default RS codes ( $\gamma = 0$  means that the substream is not transmitted,  $\gamma = 1$  means transmission with protection level 1, higher values imply increasing channel code capacity with  $\gamma$ ).

Let  $B_{av}$  be the byte budget constraint corresponding to the available bandwidth in the system.

Let  $l_i$  be the length in bytes of the *i*th packet of the *S* substreams and RS(*n*, *k*) the Reed-Solomon code used for its protection, the corresponding protection level is  $\gamma$  and the FEC coding rate is R = k/n. We define fec = 1/R = n/k as the invert of the channel coding rate, so  $l_i \times$  fec represents, in byte, the increase of the *i*th packet length when protected at level  $\gamma$ .

The correct decoding of packet *i* at the receiver yields a reduction of the distortion on the transmitted image. Let  $RD_i^0$  be the reduction of distortion associated to decoding of packet *i*, and  $RD_{i,y}$  the reduction of distortion achieved when packet *i* is protected at level  $\gamma$  ( $RD_{i,y}$  will be further formalized). We define the gain as the ratio between the image quality improvement  $RD_{i,y}$  and the associated cost in terms of bandwidth consumption  $l_i \times \text{fec.}$ 

Thus, the FEC rate allocation problem can be stated as: how to optimally select substream *i* protection level  $\gamma$  in order to maximize the associated reduction of distortion RD<sub>*i*, $\gamma$ </sub> under a budget constraint *B*<sub>av</sub>.

\$ ----

This problem is formalized by the following:

maximize 
$$\sum_{i=1}^{S} \frac{\text{RD}_{i,y}}{l_i \cdot \text{fec}_i},$$
subject to 
$$\sum_{i=1}^{S} l_i \cdot \text{fec}_i \leq B_{\text{av}}.$$
(7)

#### 4.2. Reduction of distortion metric

Taubman and Rosenbaum [15] and Descampe et al. [16] characterize a JPEG 2000 packet by its precinct indices r and p (where r and p are, resp., its resolution and spatial location), and by its layer index q, s.t  $0 \le q \le Q$ , with Q denoting the total number of quality layers. Defining RD(r, p, q) to be the amount by which the distortion, measured on the whole original image, is decreased if packet (r, p, q) is decoded compared to the distortion if only the packets  $(r, p, \alpha), \alpha < q$ , are decoded. Descampe et al. come to the conclusion that the metric RD(r, p, q) is additive, meaning that the gain in quality provided on the entire image by multiple packets has to be equal to the sum of the gain provided by each individual packet. So approximating the additive distortion by the mean square error (MSE) defined in [17], they derive the distortion  $D^q_{\alpha}$  associated to the reconstruction of the codeblock  $B_{\alpha}$  from its first q quality layers:

$$D^{q}_{\alpha} = w^{2}_{b_{\alpha}} \sum_{(x,y)\in B_{\alpha}} \left[ \hat{c}^{q}_{\alpha}(x,y) - c_{\alpha}(x,y) \right]^{2}, \tag{8}$$

where  $c_{\alpha}(x, y)$  denotes the subband coefficient in the codeblock  $B_{\alpha}$ ,  $\hat{c}^{q}_{\alpha}(x, y)$  denotes the quantized representation of these coefficients associated to the first q quality layers, and  $w_{b_{\alpha}}$  denotes the L2-norm of the wavelet basis functions for the subband to which the codeblock  $B_{\alpha}$  belongs. Denoting  $\Gamma(r, p)$  the set of codeblocks belonging to precinct (r, p), the incremental reduction of distortion RD(r, p, q) associated to the decoding of packet (r, p, q) is given by

$$\operatorname{RD}(r, p, q) = \sum_{\alpha \in \Gamma(r, p)} D_{\alpha}^{(q-1)} - \sum_{\alpha \in \Gamma(r, p)} D_{\alpha}^{q}.$$
 (9)

The FEC allocation algorithm is based on this central metric RD(r, p, q) derived from a codestream index file. The codestream index file is generated by the Open JPEG library (http://www.openjpeg.org/) and defines the gain in quality and the range of bytes corresponding to each packet. In the following we denote RD(r, p, q) as RD<sup>*i*</sup><sub>pack</sub>, the incremental reduction of distortion associated to decoding of packet *i* (packet *i* is characterized by the corresponding r and p).

## 4.3. Decoding error probability estimation

Considering an 8-bit oriented Gilbert model in [18], Yee and Weldon derive the symbol error rate (SER), thanks to the formula (5). Defining  $\varphi$  to be the correlation between two consecutive error symbols  $X_1$  and  $X_2$ , they show that

$$\varphi = \frac{E((X_1 - \text{SER})(X_2 - \text{SER}))}{\sigma^2},$$

$$\varphi = p_{bb} + p_{gg} - 1.$$
(10)

Solving (5) and (10), we have

$$p_{gg} = 1 - \text{SER}(1 - \varphi),$$
  
 $p_{bb} = 1 - (1 - \text{SER})(1 - \varphi).$ 
(11)

Thus the transition matrix is expressed by

$$P = \begin{bmatrix} 1 - \text{SER}(1 - \varphi) & (1 - \text{SER})(1 - \varphi) \\ \text{SER}(1 - \varphi) & 1 - (1 - \text{SER})(1 - \varphi) \end{bmatrix}.$$
 (12)

Yee and Weldon also consider the impact of interleaving data to level *I*. In this case they show that  $\varphi$  is replaced by  $\varphi^I$  and *P* along with the transmission probabilities become

$$P = \begin{bmatrix} 1 - \operatorname{SER}(1 - \varphi^I) & (1 - \operatorname{SER})(1 - \varphi^I) \\ \operatorname{SER}(1 - \varphi^I) & 1 - (1 - \operatorname{SER})(1 - \varphi^I) \end{bmatrix}.$$
 (13)

Hence, we obtain the following:

$$p_{gg,I} = 1 - \text{SER}(1 - \varphi^{I}),$$
  
 $p_{bb,I} = 1 - (1 - \text{SER})(1 - \varphi^{I}).$ 
(14)

Relying on the double recursion method in [18], we derive P(m, n), the probability of *m* errors in a sequence of *n* symbols:

$$P(m,n) = P_G(m,n) + P_B(m,n),$$
 (15)

where  $P_G(m, n)$  is the probability of *m* errors in *n* transmissions with the channel ending in state *G* and  $P_B(m, n)$  the probability of *m* errors in *n* transmissions with the channel ending in state *B*.

For the simplified Gilbert channel,  $P_G = 0$  and  $P_B = 1$  and we have the following.

For 
$$n = 1, 2, 3, ...$$
 and  $m = 0, 1, 2, ..., n$ ,  
 $P_G(m, n) = P_G(m, n - 1)p_{gg} + P_B(m, n - 1)(1 - p_{bb})$ ,  
 $P_B(m, n) = P_B(m - 1, n - 1)p_{bb} + P_G(m - 1, n - 1)(1 - p_{gg})$ .  
(16)

The initials conditions of the double recursion are

$$P_B(0,0) = \frac{1 - p_{gg}}{1 - p_{bb} + 1 - p_{gg}},$$

$$P_G(0,0) = \frac{1 - p_{bb}}{1 - p_{bb} + 1 - p_{gg}}$$
(17)

with  $P_B(m, 0) = P_G(m, 0) = 0$  for  $m \neq 0$ .

From these developments we derive  $P_e$  the decoding error probability of an *n*-symbol sequence protected with a channel code of capacity *t*:

$$P_e = \sum_{m=t+1}^{n} P(m, n).$$
 (18)

In our system, the channel code is a Reed-Solomon code defined by RS(n,k) and its corresponding capacity is t = (n-k)/2. Hence, the information word is a *k*-symbol packet.

It is worth noting that the JPWL final draft [4] defines 16 Reed-Solomon codes for JPEG 2000 data protection. All those recommended RS(n,k) codes have a fixed k =32 bytes. Considering each JPEG 2000 packets as an  $\eta_{w}$ information word packet and denoting  $P_e$  as the probability that a decoded word is incorrect, we derive the JPEG 2000 packet decoding error probability  $P_{pack}$ .

$$\eta_w(\text{number of word}) = \frac{\text{packet length (bytes)}}{k(= 32 \text{ bytes})},$$
(19)

 $P_{\text{pack}} = (\text{Probability that 1 word is incorrect})$ 

and  $(\eta_w - 1)$  words are well decoded)

+ (Probability that 2 words are incorrect

and  $(\eta_w - 2)$  words are well decoded) + · · ·

+ (Probability that all 
$$(\eta_w)$$
 words are incorrect)

$$P_{\text{pack}} = C_{\eta_{w}}^{1} (1 - P_{e})^{\eta_{w} - 1} (P_{e})^{1} + C_{\eta_{w}}^{2} (1 - P_{e})^{\eta_{w} - 2} (P_{e})^{2} + \cdots + C_{\eta_{w}}^{\eta_{w}} (1 - P_{e})^{\eta_{w} - \eta_{w}} (P_{e})^{\eta_{w}}.$$
(20)

Hence, we have  $P_{\text{pack}} = \sum_{i=1}^{n_w} C_{n_w}^i (1 - P_e)^{n_w - i} (P_e)^i$ .

Evaluating  $P_{\text{pack}}$  for each transmitted substream *i* and for different protection levels *y* leads to deriving a set of possible decoding error probabilities  $P_{\text{pack}}^{i,y}$ . Each of these  $P_{\text{pack}}^{i,y}$  metrics is of central importance when designing the optimization scheme in the following section.

## 4.4. Optimization

Since the optimization problem can be solved by finding the optimal protection for each substream of JPEG 2000 codestreams under a budget constraint, we define  $G_{i,y}$  as the gain in quality of the transmitted image obtained at the receiver side when packet *i* is decoded. Let  $RD_{i,1}$  and  $RD_{i,y}$  be the reduction of distortion obtained when packet *i* is transmitted respectively with protection level 1 and with protection level *y*, we have

$$RD_{i,1} = (1 - P_{pack}^{i,1}) \cdot RD_{pack}^{i},$$

$$RD_{i,\gamma} = (1 - P_{pack}^{i,\gamma}) \cdot RD_{pack}^{i}.$$
(21)

The resulting gain is

$$G_{i,1} = \frac{\text{RD}_{i,1}}{l_i} = \frac{(1 - P_{\text{pack}}^{i,1}) \cdot \text{RD}_{\text{pack}}^i}{l_i}.$$
 (22)

Similarly, any transmission between two consecutive protection levels ( $\gamma$  and  $\gamma$  + 1) yields an improvement in terms of reduction of distortion but has a budget cost equal to (fec<sub> $\gamma$ +1</sub> - fec<sub> $\gamma$ </sub>) ×  $l_i$ , hence we have

$$G_{i,y} = \frac{\mathrm{RD}_{i,y} - \mathrm{RD}_{i,y-1}}{(\mathrm{fec}_{y} - \mathrm{fec}_{y-1}) \cdot l_{i}},$$

$$G_{i,y} = \frac{(P_{\mathrm{pack}}^{i,y-1} - P_{\mathrm{pack}}^{i,y}) \cdot \mathrm{RD}_{\mathrm{pack}}^{i}}{(\mathrm{fec}_{y} - \mathrm{fec}_{y-1}) \cdot l_{i}}.$$
(23)

Protection levels incremental gains  $G_{1,1}$  to  $G_{S,y}$  are derived for each packet and stored in *S* different vectors (V1, V2, ..., VS) as presented in Figure 8. For each vector, the gains are expected to be decreasing so that the ratedistortion curve corresponding to a specific substream is always convex and that the FEC allocation is always optimal. For example, raising substream *i*'s protection level  $\gamma$  to  $\gamma$  + 1 yields more gain than going from level  $\gamma$  – 1 to  $\gamma$ , we have to merge the two elements in an average gain value  $\hat{G}$  given by:

$$\hat{G} = \frac{\text{RD}_{i,y+1} - \text{RD}_{i,y-1}}{(\text{fec}_{y+1} - \text{fec}_{y-1}) \cdot l_i},$$

$$\hat{G} = \frac{(P_{\text{pack}}^{i,y-1} - P_{\text{pack}}^{i,y+1}) \cdot \text{RD}_{\text{pack}}^i}{(\text{fec}_{y+1} - \text{fec}_{y-1}) \cdot l_i}.$$
(24)

After the merging step where all the vectors are filled with strictly decreasing gains, all the vectors  $(V1, V2, V3, \ldots, VS)$  are collected into an overall big vector  $(V\_all)$ . Then, this vector is reorganized in decreasing order of gain. The last step is to select the elements of the now strictly decreasing gains vector  $(V\_all\_ordered)$  and their corresponding protection level. For each packet, the optimal protection level is derived from the maximum related gain value selected when meeting the rate constraint (bandwidth available  $B\_av$ ).

## 4.5. Synopsis of the FEC rate allocation scheme and algorithm

Synopsis of the optimal FEC rate allocation algorithm (see Algorithm 1).

## 4.6. Proposed scheme complexity

In order to derive the complexity of the proposed FEC rate allocation scheme, we divide the algorithm into three parts.

Packet 1	Packet 2	Packet 3		Packet S
V1	V2	V3		VS
$G_{1,\gamma}$	G <sub>2,γ</sub>	G <sub>3,y</sub>		$G_{S,\gamma}$
$G_{1,1}$	G <sub>2,1</sub>	G <sub>3,1</sub>		$G_{S,1}$
G <sub>1,2</sub>	G <sub>2,2</sub>	G <sub>3,2</sub>		<i>G</i> <sub><i>S</i>,2</sub>
G <sub>1,3</sub>	G <sub>2,3</sub>	G <sub>3,3</sub>		<i>G</i> <sub><i>S</i>,3</sub>
÷	÷	÷		÷
÷	:		:	
$G_{1,\gamma_{\max}}$	$G_{2,\gamma_{\max}}$	G <sub>3,ymax</sub>		$G_{S,\gamma_{\max}}$

FIGURE 8: JPEG 2000 data packets and possible gain associated to their protection.



FIGURE 9: Gains selection by decreasing order of importance.

The first one consists in the evaluation of the gain vectors. The second part corresponds to the merging step and the last part is dedicated to ordering vector  $V_{-}$ all. Let remind that the number of JPEG 2000 codestreams is *s*; and the number of protection levels is  $\gamma_{max}$  ( $\gamma_{max}$  is fixed to 16 in [4]). Hence, we have

complexity of gains vectors estimation:  $O(s \cdot \gamma_{max})$ ; complexity of merging step:  $O(s \cdot ((\gamma_{max})^2/2))$ ; complexity of *V*\_all ordering:  $O((s \cdot \gamma_{max})^2)$ .

We conclude that the overall complexity of our scheme is  $O((s \cdot \gamma_{max})^2)$ . The complexity of layer-based FEC rate allocation scheme such as the one proposed in [13] is low and is generally of order  $O((L \cdot \gamma_{max})^2)$ , where *L* stands for the number of JPEG 2000 layers. Thus, we can infer that our scheme is slightly more complex as far as the ratio between the number of substreams and the number of JPEG 2000 layers is low. However, if the number of substreams is significantly higher compared to the number of layers, the proposed scheme may not be suitable for highly delay-constrained video streaming applications. An For each JPEG 2000 image - Model the channel with a Gilbert model and for each possible protection level y, evaluate the probability of incorrect word decoding  $P_{\text{pack}}^{i,\gamma}$ - For i = 1 to i = S (Number of JPEG 2000 packets) For  $\gamma = 1$  to  $\gamma = \gamma_{max}$ Estimate  $\text{RD}_{i,y} = (1 - P_{\text{pack}}^{i,y}) \cdot \text{RD}_{\text{pack}}^{i}$   $G_{i,y} = \frac{\text{RD}_{i,y} - \text{RD}_{i,y-1}}{(\text{fec}_{y} - \text{fec}_{y-1}) \cdot l_{i}}$  $V(i)[\gamma] = G_{i,\gamma}$ End For - Merging V(i) vectors protection levels if necessary to ensure that V(i) vectors are constituted of strictly decreasing gains values - Collecting  $V_{all} = V(i)$ End For - Ordering V\_all on decreasing order of importance values (V\_all\_ordered) - Selecting each gain value, corresponding to a specific protection level, up to meeting the rate constraint - Optimally protect JPEG 2000 packets with the corresponding Reed-Solomon codes End For

Algorithm 1

interesting extension to this work could be to combine both algorithms in a smart FEC rate allocation scheme. In this smart scheme, the packet oriented unequal error protection scheme proposed in this paper could be used for JPEG 2000 frames with reasonable number of substreams ( $s \le 1000$ ), while layer-based unequal error protection scheme will be preferred when the number of JPEG 2000 substreams significantly increases.

### 4.7. A practical scenario

Let consider the following scenario to illustrate how our optimal packet oriented FEC rate allocation algorithm works.

## Scenario

Available bandwidth:  $B_{av} = 100$  bytes.

Gilbert model parameters derived from traces analysis:

$$p_{bg} = 0.9445, \qquad p_{gb} = 0.0618.$$
 (25)

Two JPEG 2000 images codestreams packets: pack1 and pack2.

Pack1 had a length  $l_1 = 20$  bytes and it yields a reduction of distortion  $RD_{pack}^1 = 100$ .

Pack2 had a length  $l_2 = 40$  bytes and it yields a reduction of distortion  $RD_{pack}^2 = 50$ .

- Estimating the decoding error probability leads to: for protection level  $\gamma = 1$  we have fec<sub>1</sub> = 38/32 = 1.1875and estimated  $P_e = 0.008112$ for protection level  $\gamma = 2$  we have  $fec_2 = 40/32 = 1.25$ and estimated  $P_e = 0.000625$ for protection level  $\gamma = 3$  we have fec<sub>3</sub> = 45/32 = 1.40625and estimated  $P_e = 0.000007$ - Estimating JPEG 2000 decoding error probability  $P_{\text{pack}}^{i,\gamma}$ we have  $P_{\text{pack}}^{2,1} = 0.016$   $P_{\text{pack}}^{2,2} = 0.001$   $P_{\text{pack}}^{2,3} = 1.39 \cdot 10^{-5}$  $P_{\text{pack}}^{1,1} = 0.008$  $P_{\text{pack}}^{1,2} = 0.001$  $P_{\text{pack}}^{1,2} = 0.001$  $P_{\text{pack}}^{1,3} = 7.10^{-6}$ - Estimating reduction of distortion and gains vectors: Reduction of distortion  $RD_{i,y}$ :  $RD_{1,1} = 83.52$  $RD_{2,1} = 8.28$  $RD_{1,2} = 79.95$  $RD_{2,2} = 7.99$  $RD_{2,3} = 7.11$  $RD_{1,3} = 71.11$ Corresponding gains vectors estimation: (V1)(V2) $G_{1,1} = 3.5169$  $G_{2,1} = 0.3488$  $G_{1,2} = 63.96$  $G_{2,2} = 3.196$  $G_{1,3} = 22.75$  $G_{2,3} = 1.1376$ Merging vectors Step 1 (V1)(V2) $\hat{G}_{1,2} = 3.19$  $\hat{G}_{2,2} = 0.16$  $G_{2,3} = 1.14$  $G_{1,3} = 22.75$ Step 2 (V2)(V1) $\hat{G}_{1,3} = 0.81$  $\hat{G}_{2,3} = 0.13$ Building vector V\_all:  $(V_{all})$ V10.81 V2 0.13 Ordering vector V\_all into V\_all\_ordered: (V\_all\_ordered)  $\hat{G}_{1,3} = 0.81$  $\hat{G}_{2,3} = 0.13$ 

#### Algorithm 2

Let assume that there are 3 possible protection levels  $\gamma$  = 1, 2, 3 corresponding, respectively, to RS(38,32), RS(40,32), and RS(45,32).

How to optimally select the FEC rate?

We apply our FEC rate allocation algorithm (see Algorithm 2).

Selecting the gains values up to meeting the budget constraint:

 $\hat{G}_{1,3} = 0.81$  Cost<sub>1,3</sub> = 28.12 bytes (bandwidth needed 1)

 $\hat{G}_{2,3} = 0.13$  Cost<sub>2,3</sub> = 56.25 bytes (bandwidth needed 2) (26)

Bandwidth needed 1 = 28.12 bytes Bandwidth needed 2 = 84.37 bytes



Deriving each packet FEC rate:

 $\hat{G}_{1,3}$  means Pack1 should be protected with RS(45,32)

 $\hat{G}_{2,3}$  means Pack2 should also be protected with RS(45,32).

It is worth noting that having less available bandwidth,  $B_{-}av = 70$  for example, would have led to selecting only  $\hat{G}_{1,3}$ , and so, protecting and transmitting only Pack1 with RS(45,32).

## 5. JPEG 2000 IMAGE AND VIDEO STREAMING OVER REAL MANET TRACES WITH OPTIMAL FEC RATE ALLOCATION

The goal of this section is to show the results achieved while streaming JPEG 2000-based images/video over real MANET traces and to highlight the practical interest of the proposed JPWL-based system associated to our optimal FEC rate allocation algorithm.

The considered wireless channel traces are analysed in Section 3 and the video sequence used is *speedway.mj2* [19] containing 200 JPEG 2000 frames generated with an overall compression ratio of 20 for the base layer, 10 for the second layer, and 5 for the third layer. When dealing with a single image transmission, the corresponding image is *speedway\_0.j2k* ( $352 \times 288$ , 3 layers) which is the first image extracted from *speedway.mj2*. This image is constituted of 16 data packets.

As error occurrence in the transmission channel is a random process, different runs are made for each trial and the mean square error (MSE) between the original image  $(I_o)$  and the decoded image  $(I_d)$  is averaged over all runs in order to have statistically representative metrics.

The measured peak signal-to-noise ratio (PSNR) is obtained as follows:

$$MSE(I_o, I_d) = \frac{1}{M \cdot N} \sum_{x=1}^{M} \sum_{y=1}^{N} |I_o(x, y) - I_d(x, y)|^2,$$
  
$$\overline{MSE} = \frac{MSE}{N_{\text{frames}}},$$
  
$$PSNR = 10 \times \log_{10} \left(\frac{255^2}{\overline{MSE}}\right),$$
  
$$(27)$$

where  $\overline{\text{MSE}}$  is the mean square error over all the  $N_{\text{frames}}$  considered images. In the case of Motion JPEG 2000 streaming,  $N_{\text{frames}}$  represents the 200 JPEG 2000 frames constituting the video sequence and in the single image transmission case,  $N_{\text{frames}}$  represents the number of trials needed to have a statistically representative metric. Each PSNR measure is associated to a successful decoding rate metric which corresponds to decoder crash avoidance on the basis of 1000 transmission trials.

#### 5.1. On JPEG 2000 codestreams interleaving

In this section, we evaluate the impact of data interleaving in the effectiveness of the FEC rate allocation scheme. Thanks

Interleaving	PSNR (dB)	Successful
degree I	I SINK (dD)	decoding rate
I = 1	24.1	77.5
I = 2	24.6	89.8
I = 4	25.2	92.1
I = 8	31.8	93.4
I = 16	38.7	94.5
I = 32	44.33	94.7
I = 64	44.38	94.9
I = 128	44.37	94.8

to the interleaving matrix presented in Figure 10, protected JPEG 2000 data are decorrelated before being sent through the wireless channel. Hence, the impact of consecutive channel errors sequences on the transmitted codestreams is reduced. In Figure 10, the protected JPEG 2000 codestream is divided into Px packets of length N. Then, the interleaving process consists in storing M consecutive packets into an  $M \times N$  matrix; and to read the columns of this matrix so that two initially consecutive symbols are separated by a distance of I = M (symbols). We refer to I as the interleaving degree.

The considered channel is a real mobile ad hoc network channel experiencing PER =  $3.88 \times 10^{-2}$  and the interleaving degrees are 1, 2, 4, 8, 16, 32, 64, and 128. Table 1 shows the PSNR evolution as function of interleaving degree *I*. The considered image is *speedway\_0.j2k* protected with our optimal JPWL compliant scheme.

The interest of interleaving is shown in Table 1 in the sense that the PSNR and the successful decoding rate increase with the interleaving degree *I*.

The results in Table 1 are valid for a Gilbert channel with a specific error correlation factor and are no longer the same when this factor changes. For the considered channel, we observe that for  $I \le 8$ , interleaving has no noticeable impact because the interleaving degree *I* is smaller than the average error burst length. In fact, we show in Section 3.2.2 that the upper bound of the mean error burst length is  $L_B^{max} = 10$ bytes. Hence, in order to be efficient, the interleaving degree should be higher than 10 bytes. Hence, when *I* is increased to 16 or more, we notice an improvement of both the PSNR and the successful decoding rate. However, we observe that higher values of *I* (128) yield only slight improvement in terms of PSNR while consuming considerable memory resources leading to the conclusion that reasonable interleaving degree (typically I = 16 or I = 32) is a good compromise.

## 5.2. JPEG 2000 image/video streaming over real MANET channel traces

#### 5.2.1. Optimal FEC rate allocation

Figure 11 presents incremental reduction of distortion  $(RD_i^0)$  associated to decoding of the 16 packets of *speedway\_0.j2k* image. We observe that packets from 0 to 5 have the most important reduction of distortion values, therefore they are the most important packets. Hence they should be protected

TABLE 1: Interleaving degree and associated image PSNR.



FIGURE 10: Interleaving process.



FIGURE 11: Reduction of distortion of JPEG 2000 packets.

by Reed-Solomon codes with higher error correcting ability than other packets.

Figure 12 shows the RS(*n*, *k*) codes error correcting ability t = (n-k)/2 applied for the protection of the 16 packets of *speedway\_0.j2k*. The considered system experiences a carrierto-noise ratio C/N = 17 dB corresponding to PER =  $2.4 \times 10^{-2}$ . In this figure, we compare the results achieved when applying the dynamic FEC allocation rate heuristic proposed in [14], the layered unequal error protection presented in [13], the equal error protection (EEP with RS(40,32), protection rate 4/5) and our optimal FEC rate allocation scheme. The available bandwidth in the system is 6 Mbps.

We observe that the EEP scheme applies the same protection rate to all JPEG 2000 packets whereas the optimal FEC rate scheme allocates more powerful codes from packet 0 to packet 5 (first layer) and protects the other packets at lower level which is coherent when considering the importance of each packet. Moreover, from packets 6 to 15 (second layer and third layer) which yield low reduction of distortion, therefore they are less protected because they are less important.

The dynamic FEC rate allocation highly protects the first important packets (layer 1 and layer 2) but does not protect the last layer due to restricted bandwidth budget. The layered UEP proposed by Guo applies less powerful RS codes so that all the layers are protected. Contrarily to the proposed optimal scheme, both layered oriented schemes protect all the packets of the same layer at the same rate but they do not manage to take the difference between packets into account. In other words, in case of fast varying channel, the layered oriented protection scheme may not be sufficiently efficient to guarantee QoS to wireless clients.

## 5.2.2. Performance of the optimal FEC rate allocation methodology: application to wireless Motion JPEG 2000 video transmission over real MANET channel traces

In this section, performance of the optimal FEC rate allocation methodology is evaluated using *speedway.mj2* [19] video streaming over real MANET channel traces [7]. The available bandwidth in the system is 6 Mbps.

In Figure 13, we present the successful decoding rate of the transmitted video for different carrier-to-noise ratio. For  $C/N \le 14$  dB, we observe that the optimal FEC rate allocation performs from 2% to 10% better than layered UEP and EEP in terms of successful decoding rate. For the dynamic FEC rate allocation methodology, for C/N = 11 dB, we notice that the successful decoding rate is about 50%. It means that we lost half of the transmitted frames,



FIGURE 12: Correcting ability of the RS(n,32) codes used for JPEG 2000 data packets—bandwidth of 6 Mbps.



FIGURE 13: Successful decoding rate.

which is intolerable for video streaming applications. Hence, for highly noised channel, typically  $C/N \le 14$  dB, the proposed optimal scheme yields a sensitive improvement of the successful decoding rate when compared to layered UEP, dynamic scheme, and EEP.

However for noisy and slightly noisy channels where the carrier-to-noise ratio is, respectively, between  $14 \text{ dB} < C/N \le 18 \text{ dB}$  and  $C/N \ge 18 \text{ dB}$ ; the performances of all the presented methodologies in terms of successful decoding rate are close. This is because less protection is required to correct



FIGURE 14: PSNR versus carrier-to-noise ratio.

transmission errors, so even suboptimal selection of RS codes could help avoiding decoder crashes.

Figure 14 presents the PSNR of decoded video at user equipment for different channel conditions (carrier-to-noise ratio ranging from 11 dB to 20 dB).

We notice that the proposed optimal FEC rate allocation mechanism allows robust JPEG 2000 codestream streaming over mobile ah hoc networks. In fact, in terms of PSNR it performs significantly better than existing FEC rate allocation schemes thanks to efficient selection of RS codes. Hence, for highly noised channel, typically  $C/N \leq 14 \, \text{dB}$ which corresponds to PER  $\leq 9.9 \times 10^{-2}$ , the dynamic FEC rate allocation presented in [14] outperforms the layered UEP scheme proposed by Guo et al. [13]. However, for both schemes the peak signal-to-noise ratio is still bellow 30 dB, leading to unpleasant video quality. Both layered oriented methodologies are less effective than the optimal scheme, because the last one manages to take into account the importance of each packet constituting a JPEG 2000 frame. Hence, while both methodologies apply a selected RS code for a layer, the optimal schemes applies different selected RS codes for JPEG 2000 packets leading to more accurate protection level selection.

We also notice that EEP is not effective for wireless channel subjected to high level of transmission errors because of bad performance in terms of PSNR (PSNR  $\leq$  25 dB).

For noisy channel, typically  $14 \text{ dB} < C/N \le 18 \text{ dB}$ , the proposed optimal FEC rate allocation scheme performs from 4 dB to 11 dB more than layered UEP and from 9 dB to 13 dB more than EEP scheme. It is interesting to notice that the effectiveness of dynamic FEC rate allocation scheme is increased up to meeting the optimal point. This is due to the fact that FEC rate is dynamically adapted to transmission

condition which leads to better selection of RS codes. For slightly noisy channel, typically  $18 \text{ dB} < C/N \le 20 \text{ dB}$ , our proposed scheme outperforms both EEP and layered UEP scheme; and the dynamic FEC rate allocation scheme is the only one to achieve similar performance in terms of PSNR.

The main advantage of the proposed optimal FEC is its high ability to adapt channel coding to transmission environment. Hence, thanks to an efficient selection of RS codes, our optimal scheme maintains the video quality at a high and stable quality level (between 35 dB and 42 dB). Contrary to the scheme proposed in this paper, the dynamic FEC rate allocation, the layered UEP scheme, and other suboptimal schemes such as EEP, yield significant variation of the streamed video quality resulting in disgraceful visualization at user equipment. Hence, the optimal rate allocation proposed in this paper allows guaranteeing quality of service to wireless client.

## 6. CONCLUSION

In this paper, a JPWL compliant system based on an optimal FEC rate allocation scheme for robust transmission of JPEG 2000 images and video over MANET is presented.

The paper starts by an overview of JPEG 2000 and wireless JPEG 2000 (JPWL—Part 11 of JPEG 2000 standards) and then the proposed system functionalities are presented.

We analyze real mobile ad hoc network traces. Then we discuss the problem of FEC rate allocation and propose an optimal JPWL compliant methodology for FEC rate allocation.

Interesting results are then presented to illustrate the effectiveness of the proposed scheme. The impact of data interleaving is also investigated. We then demonstrate that the proposed optimal FEC allocation methodology outperforms existing layer-oriented unequal error protection schemes, using an application of Motion JPEG 2000 video streaming over real MANET channel traces.

Summarizing we can say that JPEG 2000, including the JPWL features, is a good point of departure to achieve robust video transmission over noisy channels. Hence, we consider the proposed JPWL compliant system, based on our optimal FEC rate allocation scheme, as a valid step toward guaranteeing quality of service in JPEG 2000-based wireless multimedia systems.

## ACKNOWLEDGMENTS

The video sequence used for the results presented in this document (speedway test sequence) has been generated by the Université catholique de Louvain (UCL), in the context of the MODEST project.

#### REFERENCES

- D. S. Taubman and M. W. Marcellin, JPEG 2000 Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
- [2] F. Dufaux and D. Nicholson, "JPWL: JPEG 2000 for wireless applications," in Applications of Digital Image Processing

*XXVII*, A. G. Tescher, Ed., vol. 5558 of *Proceeding of SPIE*, pp. 309–318, Denver, Colo, USA, August 2004.

- [3] D. Nicholson, C. Lamy-Bergot, X. Naturel, and C. Poulliat, "JPEG 2000 backward compatible error protection with Reed-Solomon codes," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 855–860, 2003.
- [4] JPEG 2000 part 11 Final Draft International Standard, ISO/IEC JTC 1/SC 29/WG 1 N3797.
- [5] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," STD 64, RFC 3550, July 2003.
- [6] S. Futemma, A. Leung, and E. Itakura, "RTP Payload Format for JPEG 2000 video streams," draft-ietf-avtrtp-JPEG 2000-18, September 2007, http://www.ietf.org/ internet-drafts/draft-ietf-avt-rtp-jpeg2000-19.txt.
- [7] Wireless Cameras and Audio-Visual Seamless Networking, "Loss patterns acquired during the WCAM Annecy 2004 measurement campaigns," IST-2003-507204, http://www.ist-wcam.org/.
- [8] E. N. Gilbert, "Capacity of a burst noise channel," *Bell System Technical Journal*, vol. 398, pp. 1253–1266, 1960.
- [9] O. Elliot, "Estimates of error rates for codes on burst-noise channel," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, 1963.
- [10] J. Aráuz and P. Krishnamurthy, "Markov modeling of 802.11 channels," in *Proceedings of the 58th IEEE Vehicular Technology Conference (VTC '03)*, vol. 2, pp. 771–775, Orlando, Fla, USA, October 2003.
- [11] R. Jain, The Art of Computer Systems Performance Analysis, John Wiley & Sons, New York, NY, USA, 1991.
- [12] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Wireless transmission of images using JPEG 2000," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 4, pp. 2523–2526, Singapore, October 2004.
- [13] Z. Guo, Y. Nishikawa, R. Y. Omaki, T. Onoye, and I. Shirakawa, "A low-complexity FEC assignment scheme for motion JPEG 2000 over wireless network," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, pp. 81–86, 2006.
- [14] M. Agueh, J.-F. Diouris, M. Diop, and F.-O. Devaux, "Dynamic channel coding for efficient Motion JPEG 2000 video streaming over Mobile Ad hoc Networks," in *Proceedings of the 3rd International Mobile Multimedia Communications Conference* (*MobiMedia '07*), Nafpaktos, Greece, August 2007.
- [15] D. Taubman and R. Rosenbaum, "Rate-distortion optimized interactive browsing of JPEG 2000 images," in *Proceedings* of the IEEE International Conference on Image Processing (ICIP '03), vol. 3, pp. 765–768, Barcelona, Spain, September 2003.
- [16] A. Descampe, C. De Vleeschouwer, C. Iregui, B. Macq, and F. Marques, "Prefetching and caching strategies for remote and interactive browsing of JPEG 2000 images," *IEEE Transactions* on *Image Processing*, vol. 16, no. 5, pp. 1339–1354, 2006.
- [17] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.
- [18] J. R. Yee and E. J. Weldon Jr., "Evaluation of the performance of error-correcting codes on a Gilbert channel," *IEEE Transactions on Communications*, vol. 43, no. 8, pp. 2316–2323, 1995.
- [19] Speedway video sequences have been generated by UCL. http://euterpe.tele.ucl.ac.be/WCAM/public/Speedway%20 Sequence/.

# Research Article Scalable and Media Aware Adaptive Video Streaming over Wireless Networks

## Nicolas Tizon<sup>1, 2</sup> and Béatrice Pesquet-Popescu<sup>1</sup>

<sup>1</sup> Signal and Image processing Department, TELECOM ParisTech, 46 Rue Barrault, 75634 Paris, France <sup>2</sup> R&D Department, Société Française du Radiotéléphone (SFR), 1 Place Carpeaux, Tour Séquoia, 92915 La Défense, France

Correspondence should be addressed to Béatrice Pesquet-Popescu, pesquet@tsi.enst.fr

Received 29 September 2007; Accepted 6 May 2008

Recommended by David Bull

This paper proposes an advanced video streaming system based on scalable video coding in order to optimize resource utilization in wireless networks with retransmission mechanisms at radio protocol level. The key component of this system is a packet scheduling algorithm which operates on the different substreams of a main scalable video stream and which is implemented in a so-called media aware network element. The concerned type of transport channel is a dedicated channel subject to parameters (bitrate, loss rate) variations on the long run. Moreover, we propose a combined scalability approach in which common temporal and SNR scalability features can be used jointly with a partitioning of the image into regions of interest. Simulation results show that our approach provides substantial quality gain compared to classical packet transmission methods and they demonstrate how ROI coding combined with SNR scalability allows to improve again the visual quality.

Copyright © 2008 N. Tizon and B. Pesquet-Popescu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Streaming video applications are involved in an increasing number of communication services. The need of interoperability between networks is crucial and media adaptation at the entrance of bottleneck links (e.g., wireless networks) is a key issue. In the last releases of 3G networks [1], jointly with a high speed transport channel, the high speed downlink packet access (HSDPA) technology provides enhanced channel coding features. On the one hand, packet scheduling functionalities of the shared channel located close to the air interface allow to use radio resources more efficiently. On the other hand, error correction mechanisms like hybrid automatic repeat request (HARQ) or forward error correction (FEC) contribute to build an error resilient system. However, these enhancements are designed to be operational through a large collection of services without considering subsequent optimizations. In the best case, a QoS framework would be implemented with network differentiated operating modes to provide a class of services [2]. To guarantee continuous video playout, streaming services are constrained by strictly delay bounds. Usually, guaranteed bitrates (GBR) are negotiated to maintain required bandwidth in case of congestion.

Moreover, to guarantee on-time delivery, the retransmission of lost packets must be limited, leading to an over allocation of resources to face the worst cases. The main drawback of a QoS-oriented network is that it requires a guaranteed bitrate per user and thus it does not allow to take advantage of rate variability of encoded videos. In [3], a streaming system is proposed with QoS differentiation in order to optimize experienced quality at client side in the case of degraded channel quality. Assuming that the bandwidth allocated to the user is not large enough with respect to negotiated GBR, this study shows that prioritization of packets following the regions of interest (ROI) can achieve a substantial gain on perceived video quality.

In the scope of packetized media streaming over besteffort networks and more precisely channel adaptive video streaming, [4] proposes a review of recent advances. The closest approach from our works is the well-known ratedistortion optimized packet scheduling method. However, in this technical review, scalable-based solutions are considered as inefficient due to the fact that poor compression performances and wireless networks are not really studied with their most important specificities at radio link layer like radio frame retransmissions. In [5], Chou and Miao have addressed the problem of rate-distortion optimized packet scheduling conducted as an error-cost optimization problem. In their approach, encoded data partitioned into dependent data units, which can be a scalable stream, are represented as a directed acyclic graph. This representation is used with channel error rate measurements as input parameters of a Lagrangian minimization algorithm. This general framework can be adapted in terms of channel model and transmission protocol between the server and the client. For example in [6], the error process of a wireless fading channel is approximated by a first-order Markov process. Then, in order to choose the optimal scheduling policy, the server uses this model combined with video frame-based acknowledgment (ACK/NACK) from the client to compute the expected distortion reduction to be maximized. In [7], a similar approach is proposed considering a measure of congestion instead of the previous distortion. Besides, packet scheduling algorithms can switch between different versions of the streamed video, encoded with different qualities, instead of pruning the previous set of dependent data units. Then, These methods based on rate (congestion)distortion optimized packet scheduling are in theory likely to provide an optimal solution to media aware scheduling problem. However, without simplification, the Lagrangian optimization is computationally intensive and the channel estimation (delay, capacity) may be more difficult when packets are segmented and retransmitted below application layer (e.g., ARQ at radio link control (RLC) layer). Moreover, in a wireless system, packet scheduling on the shared resource occurs at MAC or RLC layers independently of the application content.

In [7], media bitrate adaptation problem is set as a tradeoff between the current stream pruning and stream switching among a set of videos with different qualities. In order to provide more flexible schemes, the scalable extension of H.264/AVC, namely, scalable video coding (SVC), [8] allows to encode in the same bitstream a wide range of spatiotemporal and quality layers. In [9], a generic wireless multiuser video streaming system uses SVC coding in order to adapt the input stream at the radio link layer as a function of the available bandwidth. Thanks to a mediaaware network element (MANE) that assigns priority labels to video packets, in the proposed approach, a drop prioritybased (DPB) radio link buffer management strategy [10] is used to keep a finite queue before the bottleneck link. The main drawback of this method is that the efficiency of source bitrate adaptation depends on buffer dimensioning and with this approach, video packets are transmitted without considering their reception deadlines.

In this paper, our approach is to exploit the SVC coding in order to provide a subset of hierarchically organized substreams at the RLC layer entry point and we propose an algorithm to select scalable substreams to be transmitted to RCL layer depending on the channel transmission conditions. The general idea is to perform a fair scheduling between scalable substreams until the deadline of the oldest unsent data units with higher priorities is approaching. When this deadline is expected to be violated, fairness is no longer maintained and packets with lower priorities are delayed in a first time and later dropped if necessary. In order to do this, we propose an algorithm located in a so-called media aware network element (MANE) which performs a bitstream adaptation between RTP and RLC layers based on an estimation of transport channel conditions. This adaptation is made possible thanks to the splitting of the main scalable stream into different substreams. Each of these substreams conveys a specific combination of SNR and/or temporal layers which corresponds to a specific combination of high-level syntax elements. In addition, SVC coding is tuned, leading to a generalized scalability scheme including regions of interest. ROI coding combined with SNR and temporal scalability provides a wide range of possible bitstream partitions that can be judiciously selected in order to improve psychovisual perception.

The paper is organized as follows: in the next section we describe the scalable video coding context and the related standardized tools. In Section 3, we address the problem of ROI definition and propose an efficient way to transmit partitioning information requiring only a slight modification of the compressed bitstream syntax. Then, in Section 4, we present our developed algorithm to perform bitstream adaptation and packet scheduling at the entrance of RLC layer. Finally, simulation results are presented in Section 5 and we conclude in Section 6.

## 2. SCALABLE VIDEO CODING CONTEXT

#### 2.1. SVC main concepts

To serve different needs of users with different displays connected through different network links by using a single bitstream, a single coded version of the video should provide spatial, temporal, and quality scalability. As a distinctive feature, SVC allows a generation of an H.264/MPEG-4 AVC compliant, that is, backwards-compatible, base layer and one, or several, enhancement layer(s). Each enhancement layer can be turned into an AVC-compliant standalone (and not anymore scalable) bitstream, using built-in SVC tools. The base-layer bitstream corresponds to a minimum quality, frame rate, and resolution (e.g., QCIF video), and the enhancement-layer bitstreams represent the same video at gradually increased quality and/or increased resolution (e.g., CIF) and/or increased frame rate. A mechanism of prediction between the various enhancement layers allows the reuse of textures and motion-vector fields obtained in preceding layers. This layered approach is able to provide spatial scalability but also a coarse-grain SNR scalability. In a CGS bitstream, all layers have the same spatial resolution but lower layers coefficients are encoded with a coarser quantization steps. In order to achieve a finer granularity of quality, a so-called medium grain scalability (MGS), identical in principle to CGS, allows to partition the transform coefficients of a layer into up to 16 MGS layers. This increases the number of packets and the number of extraction points with different bitrates. Coding efficiency of SVC depends on the application requirements but the goal is to achieve a rate-distortion performance that is comparable to nonscalable H.264/MPEG-4 AVC. The design of the scalable



FIGURE 1: Additional bytes in SVC NAL unit header.

H.264/MPEG4-AVC extension and promising application areas are pointed out in [8].

#### 2.2. Bitstream adaptation

An important feature of the SVC design is that scalability is provided at the bitstream level. Bitstreams for a reduced spatial and/or temporal resolution can be simply obtained by discarding NAL units (or network packets) from a global SVC bitstream that are not required for decoding the target resolution. NAL units of progressive refinement slices can additionally be dropped or truncated in order to further reduce the bitrate and the associated reconstruction quality. In order to assist an MANE (e.g., a network gateway) in bitstream manipulations, the one-byte NAL unit header of H.264/MPEG4-AVC was extended by 3 bytes for SVC NAL units [11]. These additional bytes signalize whether the NAL unit is required for decoding a specific spatiotemporal resolution and quality (or bitrate) as illustrated in Figure 1. The simple priority ID "PRID" indicator is used to infer the global priority identifier of the current NAL unit. A lower value of PRID indicates a higher priority. In oder to provide a finer discrimination between SVC NAL units and to facilitate bitstream parsing, the NALU header allows to assign different priorities inside each scalable domain thanks to the values of temporal id, dependency id, and quality id fields. The reserved bit "R" can be ignored and flag "I" specifies whether the current frame is an instantaneous decoding refresh (IDR) frame. The interlayer prediction flag "N" indicates whether another layer (base layer) may be used for decoding the current layer and "U" bit specifies the reference base pictures utility (used or not) during the interprediction process. Then, discardable flag "D" signals that the content of the information in current NAL units is not used as a reference for the higher level of dependency id. At last, "O" gets involved with the decoded picture output process and "RR" are reserved bits for future extension.

## 2.3. Flexible macroblock ordering (FMO)

H.264/AVC provides a syntactical tool: FMO, which allows partitioning video frames into slice groups. Seven different modes, corresponding to seven different ordering methods, exist, allowing to group macroblocks inside slice groups. For each frame of a video sequence, it is possible to transmit a set of information called picture parameter set (PPS), in which the parameter slice\_group\_map\_type specifies the FMO mode of the corresponding frame. According to this parameter, it is also possible to transmit additional information to define the mapping between macroblocks and slice groups. Each slice group corresponds to a network abstraction layer (NAL) unit that will be further used as RTP payload. This mapping will assign each macroblock to a slice group which gives a partitioning (up to eight partitions) of the image. There exist six mapping methods for an H.264 bitstream. In this study, we use the mode 6, called *explicit MB*, to slice group mapping, where each macroblock is associated to a slice group index in the range [0..7]. The relation of macroblock to slice group map amounts to finding a relevant partitioning of an image. Evaluation of partitioning relevance strongly depends on the application and often leads to subjective metrics.

## 3. ROI EXTRACTION AND CODING

### 3.1. ROI definition

In image processing, detection of ROIs is often conducted as a segmentation problem if no other assumptions are formulated about the application context and postprocessing operations that will be applied on the signal.

Concerning the application context of our study, we formulate the basic assumption that in the majority of cases, a video signal represents moving objects in front of almost static background. In other words, we make the assumption that the camera is fixed or that it is moving slower than the objects inside the scene. With this model, moving objects represent the ROI and FMO is restricted to 2 slice groups. According to this definition, motion estimation (ME) that occurs during the encoding process delivers relevant information through motion vector values to detect ROIs. In H.264, the finest spatial granularity to perform ME is a  $4 \times 4$  block of pixels while FMO acts at macroblock level. In our simulations, to detect ROIs we compute the median value of motion vectors in a macroblock. Each vector is weighted by the size of the block it applies to. Next, the macroblock is mapped to ROI if this median value is higher than a threshold value, as depicted in Figure 2.

## 3.2. Mapping information coding

The H.264/AVC standard defines a macroblock coding mode applied when no additional motion and residual information need to be transmitted in the bitstream. This mode, called SKIP mode, occurs when the macroblock can be decoded using information from neighbor macroblocks (in the current frame and in the previous frame). In this case, no information concerning the macroblock will be carried by the bitstream. A syntax element, mb\_skip\_run, specifies the number of consecutive skipped macroblocks before reaching a nonskipped macroblock.

In our macroblock to slice group assignment method, a skipped macroblock belongs to slice group 2 (lowest priority). In fact, this assignment is not really effective because no data will be transmitted for this macroblock. The set of skipped macroblocks in a frame can be seen as



FIGURE 2: Macroblock classification according to the motion vector value.

a third slice group (with null size). In a general manner, mb\_skip\_run syntax element can be considered as a signaling element to indicate a set of macroblocks belonging to a slice group (index incremented by one) as depicted in Figure 3. If slice groups with higher indices are lost, the decoding process will still be maintained with lower indexed slice groups. This method generalizes the use of mb\_skip\_run syntax element and allows to code macroblock to slice group mapping without sending explicit mapping with the frame header, picture parameter set (PPS). Indeed, mb\_skip\_run is included into the H.264 bitstream syntax, coded with an efficient entropy coding method. This coding method does not introduce new syntax elements but as the meaning of mb\_skip\_run is modified (in the case of more than one slice group), the provided bitstream is no longer semantically compliant with regard to the H.264 reference decoder. At the client side, each slice group is received independently through a specific RTP packet. To be able to perform bitrate adaptation, the MANE needs to know the relative importance of each slice group without parsing the scalable bitstream. In the next section, we propose a method using SVC high-level syntax to label each slice group with the appropriate priority.

## 4. ADAPTATION AND PACKET SCHEDULING

In the sequel, we will restrict scalability abilities of SVC to the temporal layering with the well-known hierarchical B pictures structure, and to SNR scalability with MGS slices coding. In fact, we assume that spatial scalabilitybased adaption has already occurred when reaching the bottleneck link. Thanks to the additional bytes in SVC NAL unit headers, the network is able to select a subset of layers from the main scalable bitstream. Moreover, in the previous section, we described a coding method in order to provide a data differentiation at image content or ROI level. In this section, we propose a packetization method that combines SVC native scalability modes and the underlying scalability provided by ROI partitioning with FMO.

## 4.1. Packetization and stream-based priority assignment

In this study, we adopt an adaptation framework in which the streaming server sends scalable layers as multiple RTP substreams that are combined into a single RTP stream, adapted to each client transmission condition in the MANE [11] as described in Figure 4. With SVC extended NAL



FIGURE 3: An example of macroblock to slice group map coded via mb\_skip\_run syntax.



FIGURE 4: Scalable bitstream adaptation in the MANE based on users conditions.

unit header, 6 bits indicate simple priority ID. Then, we use this field to specify the importance of a slice group (SG)determined upon ROI definition in Section 3, and the third byte specifies NAL unit assignment to temporal and quality levels. The higher the importance of the SG, the lower the value of the priority ID. Inside a scalability domain (temporal or SNR), packet prioritization derivation is straightforward according to the appropriate level ID in the third byte of the NAL unit header. For example, temporal level 0 corresponds to the highest priority among temporal level IDs. In the case of combined scalability, priority labeling is more complicated and usually dependent on the application. For example, watching a scene with high motion activities may require high temporal resolution rather than high-quality definition because human vision



FIGURE 5: Scalable scheduling principle with three substreams.

does not have time to focus on moving objects details but privileges display fluidity. Then in this example, if the receiver undergoes bandwidth restrictions, it would be more judicious for the MANE to transmit packets with highesttemporal level and lowest-quality level before packets with lowest-temporal level and highest-quality level. On the contrary, with static video contents, the MANE will favor quality rather than temporal resolution. Finally, adding ROI scalability makes possible to deliver different combinations of quality and temporal scalabilities between regions of the same video frame. In Section 5.2, from simulation results, we discuss how to find the best combination of scalable streams to optimize perceived video quality in function of the considered application and media content. Next, we assume that MANE input data is composed of N substreams indexed from higher to lower importance or priority. Each stream can be a simple scalable layer with a given temporal or quality level or a more sophisticated combination of layers as explained before.

### 4.2. Packet scheduling for SVC bitstream

In the remaining of this study, we consider that the MANE sees RLC layer as the bottleneck link and performs packet scheduling from IP layer to RLC layer. In the case of a 3G network, the MANE is most probably between the radio network controller (RNC) and the gateway GPRS support node (GGSN) and we neglect transmission delay variations between the server and the MANE. Then, each RTP packet whose payload is an NAL unit is received by the MANE at  $t = TS + t_0$ , where TS is the sampling instant of the data and  $t_0$  the constant delay between the MANE and the server. Next, to simplify this we put  $t_0 = 0$  knowing that this time only impacts the initial playout delay. Moreover, inside each scalable stream, packets are received in their decoding order which can be different from the sampling order due to the hierarchical B pictures structure. Hence, the head-ofline (HOL) data unit of a stream queue is different from the minimum sampling instant of queued packets: TS<sub>min</sub>.

Input RTP streams are processed successively. When scheduling RTP packet, the algorithm evaluates the transmission queues of the most important streams and, according to network state, the current packet will be delayed or sent to RLC layer. All streams are next transmitted over the same wireless transport channel and when an RTP packet reaches RLC layer, all necessary time slots are used to send the whole packet. Therefore, the general principle of the algorithm is to allow sending a packet only if packet queues with higher priorities are not congested and if expectable bandwidth is sufficient to transmit the packet before its deadline.

In order to detail the algorithm, we are considering that the bitstream is transmitted through a set of *L* streams and the scheduler is up to send the HOL packets of the *k*th stream at time *t*. Let us denote  $TS_k(t)$  as the sampling instant of this packet,  $S_k(t)$  as its size,  $d_k(t)$  as its transmission time, and  $D_{max}$  as the maximum end-to-end delay for all packets of the streaming session. Scheduling opportunities for this packet will be inspected only if its reception deadline is not past and if a significant ratio  $\epsilon$  of the maximum end-to-end delay is still available before reaching this deadline as follows:

$$t - d_k(t) < (1 - \epsilon)D_{\max}.$$
 (1)

If this condition is not verified, the packet is discarded. Otherwise, to perform the transfer of the packet to the RLC layer (see Figure 5), that is to send or to delay the packet, packet queue of the *l*th stream, where l = k + 1, ..., *L*, is considered as a single packet with time stamp  $TS_{\min l}(t)$ . Then, we define  $D_l(t)$ , the transmission time for this aggregated packet and we fix  $t' = t + d_k(t)$ . The second condition which must be verified before sending the packet is

$$t' - \mathrm{TS}_{\min}(t') < (1 - \epsilon)D_{\max} - D_l(t').$$
(2)

With this condition, the algorithm assures that the network is able to send the packet without causing future packets loss from streams with higher priorities. If this condition is not verified, the packet is put on the top of the *k*th queue and the algorithm examines the (k + 1)th stream.

Moreover, packet dependency can occur between packets from the same stream, in the case of a combined scalabilitybased stream definition, or between packets from different streams. Therefore, in order to provide an efficient transmission of scalable layers, the algorithm delays packet delivering until all packets from lower layers which are necessary to decode the current packet are transmitted.



FIGURE 6: 2-state Markov channel model.

Given these two conditions, the main difficulty is to evaluate the 5 variables that are defined as a function of time and need to be calculated in the future. Firstly, let us note that the RTP streams are processed sequentially and thus between t' and t instants, the sizes of the others packet queues  $(l \neq k)$  will increase and their oldest time stamp will remain unchanged. So, we can write  $TS_{\min l}(t') = TS_{\min l}(t)$ . Next, we calculate the  $d_k(t)$  value which amounts to perform a channel delay estimation. In order to do this, we are considering that the channel state is governed by a 2-state Markov chain. Therefore, thanks to this model, the network is simply considered to be in "GOOD" or "BAD" state as depicted in Figure 6. The transition probabilities,  $\lambda$  and  $\mu$ , are considered as function of time variables in order to take into account possible channel state evolutions. In order to complete the network model, we define tti and rfs as the variables that represent the transmission time interval (TTI) and the radio frame size (RFS) constant values. A radio frame is actually an RLC protocol data unit (RLC-PDU). Before reaching the RLC layer, an RTP packet is segmented into radio frames and an RLC-PDU is sent every TTI. In fact, if tti and rfs are constant, we implicitly assume that we are dealing with a dedicated channel with constant bitrate. Nevertheless, in our simulations tti value can be modified in order to simulate a radio resource management-based decision of the network which can perform bandwidth allocation on the long run. Additionally, channel state transitions occur every TTI, so we can write the current time as a discrete variable:  $t = n \times \text{tti}$ . Finally, the transition probabilities,  $\lambda$  and  $\mu$  are dynamically calculated every TTI performing a state transition count over a sliding time window  $T = N \times tti$ .

Let us define the random process TT(t) (transmission time) which represents the time spent by the network (including RLC retransmissions) to send a radio frame whose first sending instant is *t*. Actually, TT is a discrete time process and we have  $TT(t) = TT(n \times tti) = TT(n)$ . As rfs is constant,  $I = [S_k(t)/rfs]$  is the number of RLC-PDUs involved in the transmission of the current HOL RTP packet of the *k*th stream. With these notations, let us denote tti  $\times$ { $n_0, n_1, \ldots, n_I$ } with  $n_0 = n$ , the sequence of sending instants corresponding to the first transmission of the related RLC-PDUs. So, we can express the overall transmission time of the RTP packet as follows:

$$d_k(t) = \sum_{i=n}^{I} \mathrm{TT}(n_i).$$
(3)

In order to evaluate TT(n), we use past observations thanks to radio link control acknowledged mode (RLC AM) error feedback information sent by the receiver. This information is received by the transmitter after a certain feedback delay,  $r \times$  tti, and r is a fixed integer value which depends on RLC configuration. Moreover, we estimate the average value of TT over the RTP packet transmission duration by the average value of TT(n-r). In other words, we consider that the average channel state is constant through RTP packet transmission duration. So, we have the following estimated parameter:

$$\widehat{d}_k(t) = E\{\mathrm{TT}(n-r)\} \times \left\lceil \frac{S_k(t)}{\mathrm{rfs}} \right\rceil.$$
(4)

When the channel is in "GOOD" state, TT(n) = tti and when the channel state is "BAD," we approximate TT(n) by the average TT value of previously retransmitted RLC-PDU (one time at least) over the previously defined time window *T*. Let us denote  $tt_{bad}$  by this average value. We have

$$tt_{bad}(n) = \frac{\sum_{i=n-N,TT(i)>tti}^{n}TT(i)}{\sum_{i=n-N,TT(i)>tti}^{n}i}.$$
(5)

Then, the mean value of TT(n) can be expressed as

$$E\{TT(n)\} = tt_{bad}(n) \times P(TT(n) = tt_{bad}(n) | TT(n-1))$$
  
+ tti \times P(TT(n) = tti | TT(n-1)).  
(6)

In order to provide the estimation of  $D_l(t')$  involved in the scheduling condition defined by (2), we define  $S_l(t')$  as the size of the aggregated RTP packets of the *l*th stream. In addition, let us define  $r_l(t)$  as the source bitrate of this *l*th stream calculated over the previously defined time window *T*. Thus, in the sequel, we will use the following approximation:

$$S_l(t') = S_l(t) + r_l(t) \times d_k(t).$$
(7)

Next, we estimate the transmission time of this aggregated packet assuming that the previous network estimation (6) will be usable over the time interval  $[t, D_l(t')]$ . Therefore, similar to (4), we can write

$$\widehat{D}_{l}(t') = E\{\mathrm{TT}(n-r)\} \times \left\lceil \frac{S_{l}(t')}{\mathrm{rfs}} \right\rceil.$$
(8)

#### 5. EXPERIMENTAL RESULTS

#### 5.1. Simulation tools

To evaluate the efficiency of the proposed approach, some experiments have been conducted using a network simulator provided by the 3GPP video ad hoc group [12].

This software is an offline simulator for an RTP streaming session over 3GPP networks (GPRS, EDGE, and UMTS). Packet errors are simulated using error masks generated from link-level simulations at various bearer rates and block error rate (BLER) values. Moreover, this simulator offers



FIGURE 7: Simulation model.

the possibility to simulate time events (delays) using the time stamp field of the RTP header. The provided network parameters are nearly constant throughout the session. For simulating radio channel conditions two possible input interfaces are provided: bit-error patterns in binary format, as well as RLC-PDU losses in ASCII format. Error masks are used to inject errors at the physical layer. If the RLC-PDU is corrupted or lost, it is discarded (i.e., not given to the receiver/video decoder) or retransmitted if the RLC protocol is in acknowledged mode (AM). The available biterror patterns determine the bitrates and error ratios that can be simulated. Two bit-error patterns with binary format are used in the experiment. These patterns are characterized by a relatively high BER (BER = 9.3e - 3 and BER = 2.9e - 3) and are suited to be used in streaming applications, where RLC layer retransmissions can correct many of the frame losses. All bearers are configured with persistent mode for RLC retransmissions and their bitrates are adjusted using the RLC block size and the TTI parameters provided by the simulator. An erroneous RLC packet is retransmitted until it is correctly received. If the maximum transfer delay due to retransmission is reached, the corresponding RTP packet is discarded. Therefore, the residual BER is always null, only missing RTP packets may occur, as depicted in Figure 7. In order to validate a strategy, results must be provided over a large set of simulations varying the error mask statistics. Therefore, for a simulation, the error pattern is read with an offset varying from 0 at the first run and incremented by 1 for each run and finally the results are evaluated over a set of 64 runs, as recommended in [13].

In addition, the RTP packetization modality is single network abstraction layer (NAL) unit mode (one NAL unit/RTP payload), the division of original stream into many RTP substreams leads to an increase of the number of RTP headers. To limit the multiplications of header information, the interleaved RTP packetization mode allows multitime aggregation packets (NAL units with different time stamps) in the same RTP payload. In our case, we make the assumption that ROHC mechanisms provide RTP/UDP/IP header compression from 40 to 4 bytes in average, which is negligible compared to RTP packet sizes, and we still packetize one NAL unit per RTP payload.



FIGURE 8: Prediction mode structure and ROI coding scheme.

#### 5.2. Simulation results

To evaluate the proposed approach, we present simulation results obtained with the following three test sequences.

- (i) *Mother and daughter* (15 *fps, QCIF,* 450 *frames*): fixed background with slow moving objects.
- (ii) Paris (15 fps, QCIF, 533 frames): fixed background with fairly bustling objects.
- (iii) Stefan (15 fps, QCIF, 450 frames): moving background with bustling objects (this sequence is actually a concatenation of 3 sequences of 150 frames in order to obtain a significant simulation duration).

The prediction mode scheme for frame sequencing is the classical IPPP... pattern in order to evaluate the robustness of the proposed approach and its capacity to limit distortion due to error propagation. The ROI is periodically redefined after each P frame, as illustrated in Figure 8. Concerning the common scalability features, SVC bitstreams are encoded with a group of pictures (GOP) size of 8 (4 temporal levels) and one MGS refinement layer which corresponds to a quantization factor difference of 6 from the base to the refinement quality layer. Then, each RTP packet can be either the quality base layer of a slice group or its enhanced quality layer at a given temporal level. The constants defined in Section 4.2 are used with the following values:  $D_{\text{max}}$  = 1.5 s, rfs = 80 bytes, tti = 10 ms by default, and r = 2. Finally,  $\epsilon$  is fixed to 25% after a progressive decrease (65% at the beginning) during the first seconds of the transmission.

 Mother and daughter
 Paris
 Stefan

 H.264/AVC
 27.58 dB
 26.43 dB
 18.6 dB

 SVC
 34.2 dB
 29.74 dB
 27.73 dB

TABLE 1: Performance comparison between H.264 (one RTP stream) and SVC (2 RTP streams: base layer and SNR refinement).

In fact, at the beginning of the transmission each RTP queue is empty and the scheduling algorithm could cause network congestion as it would transmit all the refinement layers without discarding before reaching the stationary state. Thus, the progressive decrease of  $\epsilon$  allows us to limit this undesirable behaviour during the transitional period.

### 5.2.1. Adaptation capabilities

Table 1 presents simulation results obtained by configuring each channel with a BLER of 10.8% (BER = 9.3e - 3). For "Paris" and "mother and daughter" sequences, the bitrate provided at RLC layer is 64 Kbps and then by removing 4 bytes/packet of RLC header information, the maximum bitrate available at application level (above RTP layer) is approximately 60.8 Kbps. Moreover, for these two sequences, in the case of H.264 coding, a bitrate constrained algorithm at source coding was used in order to match an average target bitrate of 60 Kbps. Concerning "Stefan" sequence, the motion activity is much more significant and to obtain an acceptable quality, we encode the video with an average target bitrate of 120 Kbps. Thus, the corresponding channel used to transmit this sequence is configured with a TTI of 5 ms, leading to a maximum available bitrate of 121.6 Kbps. In the case of SVC coding, the video is encoded without bitrate control algorithm and streamed through two RTP streams. The first one corresponds to the quality base layer transmitted with the highest priority and the second corresponds to the enhanced quality layer transmitted with lower priority. For this first set of simulations, no other scalability features, temporal or SNR, are used to differentiate the RTP streams. PSNR values are measured over the whole sequence and the proposed method allows to gain from 3.3 dB to 9.13 dB. The capacity of our method to better face error bursts is particularly visible in Figure 9. At the beginning of the session, up to  $t = 150 \,\mathrm{ms}$ , the two coding methods provide a good quality. With SVC coding, the quality is a little bit lower, but more constant, due to the progressive decrease of  $\epsilon$  previously described. At the end of this starting period, an error burst occurs and the quality with the nonscalable coding dramatically decreases. However, as the content of the sequence does not vary a lot from one image to another, the decoder is able to maintain an acceptable quality. Next, at around t = 350 ms, another error burst occurs and also the content of the video is quite more animated. Then, with H.264 coding, the decoder is no longer able to provide an acceptable quality, whereas with SVC we observe only a limited quality decrease. So, our proposed method better faces error bursts, adapting the transmitted bitrate given the estimated capacity of the transport channel.



FIGURE 9: Frame PSNR evolution for "mother and daughter" test sequence (BLER = 3.3%, tti = 10 milliseconds).

Moreover, our algorithm provides an adaptation mechanism that avoids fatal packet congestion when the source bitrate increases. This second aspect is particularly interesting in the case of video which represents bustling objects with a lot of camera effects (zoom, traveling, etc.) like "Stefan" sequence. In this sequence, as illustrated in Figure 10, the bitrate (at MANE input) hugely fluctuates due to the high motion activity. On the one hand, our algorithm allows bitrate variations and achieves a good quality when the available channel bitrate is large enough. On the other hand, when the required bitrate overcomes the channel capacity, the quality refinement layer is discarded, leading to a limited quality decrease (t = 8 s). Next, during a short period, even if the source bitrate decreases under the channel capacity, this enhanced quality layer is still discarded. This localized congestion phenomenon is due to the response time of the algorithm. After this transitory period, the full quality is achieved again.

#### 5.2.2. Adaptation capabilities and bandwidth allocation

In this section, the simulations are conducted in order to study the combined effects of channel errors and bandwidth decrease. Indeed, the implementation of a dedicated channel with a purely constant bitrate is not really efficient in terms of radio resource utilization between all users. Then, a more advanced resource allocation strategy would decrease the available bandwidth of the user when his conditions become too bad, in order to better serve other users with better experienced conditions. This allocation strategy, which aims at maximizing the overall network throughput or the sum of the data rates that are delivered to all users in the network,


FIGURE 10: Bitrate adaptation with highly variable source bitrate (Stefan, BLER = 3.3%, tti = 4 milliseconds).



FIGURE 11: Bitrate adaptation with two RTP streams: quality base layer and SNR refinement layer (Paris).

corresponds to an ideal functioning mode of the system but it is not really compatible with a QoS-based approach.

Actually, with a classical video streaming system, it is not really conceivable to adjust the initially allocated channel bitrate without sending feedbacks to the application server, which is generally the only entity able to adapt the streamed bitrate. Moreover, when these feedbacks are implemented, adaptation capabilities of the server are often quite limited in the case of a nonscalable codec: transcoding, bitstream switching, and so forth. Then in our proposed framework, with the MANE located close to the wireless interface, it is possible to limit the bitrate at the entrance of the RLC layer if a resource management decision (e.g., bandwidth decrease) has been reported. In this case, as illustrated in Figure 11, our adaptive packet transmission method allows to maintain a good level of quality while facing a high error rate and a channel capacity decrease. In the presented simulation results, after 15 ms a quality decrease of 1.7 dB in average and 4 dB in the worst case is measured, whereas the available user bitrate is reduced by more than 30% because of the combined effects of allocated bandwidth decrease (30%) and BLER increase.

#### 5.2.3. Scalability and ROI combined approach

In this section, we evaluate the contribution, in terms of psychovisual perception of the ROI-based differentiation combined with SVC intrinsic scalability features. In order to do this, the simulator is configurated like in the previous section with a bandwidth decrease at the 15th second. At the source coding, an ROI partitioning is performed as described in Section 3 and a quality refinement layer is used, leading to a subset of three RTP streams:

- (i) the quality base layer of the whole image (high priority),
- (ii) the refinement layer of the ROI slice group (medium priority),
- (iii) the refinement layer of the background (low priority).

In Figure 12, we can observe the quality variation per image region through the session. So, at the beginning, when channel conditions are favorable, the two regions are transmitted with quite similar quality levels and we reach the



FIGURE 12: Bitrate adaptation with 3 RTP streams: quality base layer, SNR refinement for ROI, and SNR refinement for background ("Paris" sequence).



FIGURE 13: Visual comparison at t = 17.5 seconds (Paris, BLER = 10.8%, tti = 10 milliseconds). (a) No ROI differentiation, (b) ROI and SNR combined scalability ("Paris" sequence).

maximum achievable quality between t = 8 s and t = 15 s. Next, when the channel error rate increases, the available bandwidth is reduced by 50% and we clearly observe two distinct behaviors, following the concerned image region. The quality of the background deeply falls (4 dB in average) and remains almost constant. On the contrary, the quality of the ROI becomes more variable but the PSNR decrease is contained (less than 2 dB in average).



FIGURE 14: Slice group mapping ("Paris" sequence, *t*=17.5 seconds).

In order to illustrate these PSNR variations, a visual comparison is provided in Figure 13. In fact, the main interest of this method is that quality variations of the background are not really perceptible. So, in order to better illustrate the gain of this method in terms of visual perception, we compared the displayed image in two cases: with and without ROI differentiation, with the channel conditions evolution of the previous simulation. Moreover, Figure 14 represents the slice group partitioning between ROI and background for the concerned video frame. Thus, we can observe that figures and human expressions of the personages are provided with better quality when the ROI-based differentiation is applied. Moreover, some coding artefacts are less perceptible around the arm of the woman.

In addition, our proposed algorithm is designed in order to allow more complex layers combinations with temporal scalability. In our simulations, the utilization of the temporal scalability did not provide a substantial additional perceived quality gain. In theory, it would be possible to perform more sophisticated differentiation between images regions. For example, we can imagine a configuration where the stream with the highest priority contains the following layers:

- (i) quality base layer of the ROI with the full temporal resolution,
- (ii) SNR refinement layer of the ROI with a reduced temporal resolution,
- (iii) quality base layer of the background with a reduced temporal resolution.

In fact, the bitrate of a quality base layer, and more particularly for the background, is often low. Hence, the bitrate saved by removing from the temporal resolution of the background is not high enough to compensate for the additional SNR refinement layer of the ROI. Therefore, the global bitrate of this RTP stream would be high and it would not be surely transmitted, leading to degraded performances.

## 6. CONCLUSION

This study proposes a complete framework for scalable and media aware adaptive video streaming over wireless networks. At the source coding, we developed an efficient coding method to detect ROIs and transmit ROI mapping information. Next, using the SVC high-level syntax, we proposed to combine ROI partitioning with common scalability features. In order to multiplex scalable layers, we adopted the MANE approach. In our system, the MANE is close to the wireless interface and it manages RTP packets transmission to the RLC layer following priority rules. In order to do this, a bitrate adaptation algorithm performs packet scheduling based on a channel state estimation. This algorithm considers the delay at RLC layer and packet deadlines in order to maximize the video quality avoiding network congestion. Our simulations show that the proposed method outperforms classical nonscalable streaming approaches and the adaptation capabilities can be used to optimize the resource utilization. Finally, the ROI approach combined with SNR scalability allows to improve again the visual quality. Future work will aim at generalizing this study in the case of a shared wireless transport channel.

#### REFERENCES

- [1] 3GPP, "High Speed Downlink Packet Access (HSDPA)," 3GPP TS 25.308 V7.3.0, June 2007.
- [2] M. Etoh and T. Yoshimura, "Advances in wireless video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 111–122, 2005.
- [3] N. Tizon and B. Pesquet, "Content based QoS differentiation for video streaming in a wireless environment," in *Proceedings* of 15th European Signal Processing Conference (EUSIPCO '07), Poznan, Poland, September 2007.
- [4] B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," *Wireless Communications* and Mobile Computing, vol. 2, no. 6, pp. 573–584, 2002.
- [5] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, 2006.

- [6] D. Tian, X. Li, G. Al-Regib, Y. Altunbasak, and J. R. Jackson, "Optimal packet scheduling for wireless video streaming with error-prone feedback," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '04)*, vol. 2, pp. 1287–1292, Atlanta, Ga, USA, March 2004.
- [7] E. Setton, Z. Xiaoqing, and B. Girod, "Congestion-optimized scheduling of video over wireless ad hoc networks," in *Proceedings of the IEEE International Symposium on Circuits* and Systems (ISCAS '05), vol. 4, pp. 3531–3534, Kobe, Japan, May 2005.
- [8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/MPEG4-AVC extension," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 161–164, Atlanta, Ga, USA, October 2006.
- [9] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer, "Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG4-AVC," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 625–628, Toronto, Canada, July 2006.
- [10] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner, "Radio link buffer management and scheduling for wireless video streaming," *Telecommunication Systems*, vol. 30, no. 1–3, pp. 255–277, 2005.
- [11] S. Wenger, Y.-K. Wang, and T. Schierl, "RTP payload format for SVC video," *draft, Internet Engineering Task Force (IETF)*, February 2008.
- [12] 3GPP and Siemens, "Software simulator for MBMS streaming over UTRAN and GERAN," *document for proposal, TSG System Aspects Working Group4#36, Tdoc S4-050560*, September 2005.
- [13] 3GPP and BenQmobile, "Coponents for TR on video minimum performance requirements," *document for decision, TSG System Aspects Working Group4#39, Tdoc S4-060265*, May 2006.

# Research Article Fast and Accurate Video PQoS Estimation over Wireless Networks

## **Pasquale Pace and Emanuele Viterbo**

Department of DEIS, University of Calabria, 87036 Rende, Italy

Correspondence should be addressed to Pasquale Pace, ppace@deis.unical.it

Received 29 September 2007; Revised 10 February 2008; Accepted 9 April 2008

Recommended by F. Babich

This paper proposes a curve fitting technique for fast and accurate estimation of the perceived quality of streaming media contents, delivered within a wireless network. The model accounts for the effects of various network parameters such as congestion, radio link power, and video transmission bit rate. The evaluation of the perceived quality of service (PQoS) is based on the well-known VQM objective metric, a powerful technique which is highly correlated to the more expensive and time consuming subjective metrics. Currently, PQoS is used only for offline analysis after delivery of the entire video content. Thanks to the proposed simple model, we can estimate in real time the video PQoS and we can rapidly adapt the content transmission through scalable video coding and bit rates in order to offer the best perceived quality to the end users. The designed model has been validated through many different measurements in realistic wireless environments using an ad hoc WiFi test bed.

Copyright © 2008 P. Pace and E. Viterbo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

It is well known that the goal of any QoS mechanism is to maintain a good level of user-perceived QoS even when the network conditions are changing unpredictably.

Typical QoS provisioning solutions for multimedia video applications have been always based on the idea of trying to reserve or assure certain network guarantees, so that packets coming from delay or bandwidth sensitive applications receive a better treatment in the network. This approach has been demonstrated to work very well in fixed networks. However, in wireless networks it is not always possible to offer any guarantee, due to continuously changing conditions and unpredictable radio link quality.

Increasing bandwidth is a necessary first step for accommodating real-time streaming applications, however it is not sufficient, due to large bandwidth fluctuations experienced in wireless networks. Fluctuations in network resource availability, due to channel fading, variable error rate, mobility, and handoff, make QoS provisioning more complex in wireless networks. Moreover, determining how network congestion manifests itself in degraded stream quality is still an open issue and only some very recent studies are available [1, 2]. Understanding the relationship between stream quality and network congestion is an important step to solving this problem, and can lead to better design of streaming protocols, computer networks, and content delivery systems.

One of the critical issues to keep in mind when dealing with provision of multimedia services is the quality of sound or picture presented to the end user, assuming a high-quality source and an error-free environment. This quality is directly proportional to the bit-rate used in the encoding process, thus more recently, diverse solutions were proposed for scalable multimedia transmissions over wireless networks [3, 4]. Many of these adaptive solutions gradually vary the video streams' characteristics in response to fluctuating network conditions thereby allowing for the perceived quality to be gracefully adapted. Nevertheless, the quality experienced by a user of multimedia service not only depends on network parameters but also on higher layers' characteristics. An alternative way for providing the agreed quality of service is to estimate the perceived quality of service (PQoS) index, with the aim of selecting the best scaling for the video content in order to achieve the "golden selection" between quality of service, bandwidth availability, bit rate, and frame transmission rate.

The objective quality perceived by the nonexpert user can be measured with purely subjective criteria, as opposed to the Network QoS, which relies on objective measurable parameters (throughput, BER, etc.). A complicating factor is the individual nature of how users evaluate the quality that they receive. Any two users who may be sharing a common experience (i.e., identical applications) are likely to have significantly different views of the QoS; thus, the important thing is to understand how such individual views are used for estimating the connection between wireless network parameters and user perception of QoS provided over that network.

This linkage will typically take the form of a numerical mapping (mathematical relation) between some measure of the user-perceived quality (e.g., the *mean opinion score* (MOS) [5]) and a particular set of network parameters (e.g., available bandwidth).

Typically, the five-point scale MOS is used to collect feedback from end users on the subjective quality of a media stream. However, assessments of subjective quality are time consuming and expensive; furthermore, they cannot be easily and routinely performed for real time systems. On the other hand, objective metrics would be of great benefit to applications involving scalable video coding and multidimensional bit rate control used in mobile video broadcasting systems. According to these consideration, there is a need for a quality metric estimator, based on the VQM objective metric [6, 7] that accurately matches the subjective quality and can be easily implemented in real-time video systems.

#### 1.1. Paper contributions

This work presents the following key contributions.

- (i) We setup an ad hoc test bed for evaluating the perceived video quality of multimedia contents transmitted over a wireless network using the VQM objective metric.
- (ii) We examine how network parameters such as congestion, signal power level, and transmission bit rate affect streaming media and video data that are sent on demand over the wireless network from a single server center to one or more users equipped with a handled device.
- (iii) We design an accurate analytical model for real-time estimation of the perceived quality according to the network and video parameters. Finally, we verify the quality of the proposed model in several network conditions.

Thanks to this model we can estimate the PQoS of each video and we can rapidly adapt the transmission of the content through scalable video coding and multidimensional bit rate techniques in order to offer the best quality to the end users. Thus, it could be possible to implement and use "*adaptive applications*" as a complement to the traditional network-layer reservations. So, whenever the network resources become scarce and the QoS guarantees are violated, the applications can self-adapt the internal settings (e.g., frame rates, video sizes, etc.) reducing the data rates to

those that the network can support in that precise moment and always guaranteeing a good PQoS value.

## 2. RELATED WORK AND LITERATURE

Most of the proposed solutions [8–10] for QoS guarantee in wireless networks follow a proxy-based approach, and rely on the underlying network to provide services like bandwidth reservation and priority routing and scheduling. Even if the approach is transparent to the applications, lack of support from any intermediate network or node can render the architecture useless. For example, in case priority routing is not supported by a router on the transmission path, the whole scheme will fail. Moreover, proxy-based solutions have scalability problems [11], especially in case of computation intensive proxy functionality like transcoding. With the aim of overcoming the drawbacks of computing the true quality of service perceived by the end users, some quality metric evaluation have been conducted in the last years. Although Feghali et al. [12] proposed a new quality metric for filling the gap between the classical PSNR and the subjective quality metrics, they do not consider other network-level parameters, such as the wireless link power and the effect produced by other data traffic on the same link. In [13] the authors study the user perception of multimedia quality, when impacted by varying network-level parameters such as delay and jitter, however they use subjective quality metrics that are very expensive and time consuming. Paper [14] presents a method for objective evaluation of the perceived quality of MPEG-4 video content, based on a quantification of subjective assessments. Showing that subjectively derived perceived quality of service (PQoS) versus bit rate curves can be successfully approximated by a group of exponential functions, the authors propose a method for exploiting a simple objective metric, which is obtained from the mean frame rate versus bit rate curves of an encoded clip; even in this work no network-level parameters have been considered. Koumaras et al. [1] presented a generic model for mapping QoS-sensitive network parameters to video quality degradation but they considered only the packet loss during the transmission over the wireless link without taking into account the congestion due to the background traffic over the same link and the video resolution in terms of bit rate. Lotfallah et al. [2] identified a parsimonious set of visual content descriptors that can be added to the existing video traces to form advanced video traces, then they developed quality predictors that, based on advanced video traces, predict the quality of the reconstructed video after lossy network transport. Even in this work no considerations are made on video bit rate adaptation according to the background traffic.

In our work, we evaluate the perceived quality value according to the bit rate of the transmitted video, the signal power level, and the data traffic on the wireless link. We design a model that closely approximates VQM objective metric behavior. Thanks to the proposed model; the estimation of the PQoS is extremely easy and fast making the tool suitable for scalable video coding and multidimensional bit rate in mobile wireless video application.

	Video
	P.910 video quality assessment
	P.930 reference impairment system
Subjective	BT.500-6, BT.601-4, BT.802 TV pictures
	BS.562-3, BS.1116 high quality audio
	G.114 delay
	P.920 interactive test for AV
Ohissting	P.OAV objective audiovisual quality assessment
Objective	G.191 software tool for evaluation test

#### TABLE 1: Video ITU recommendations.

## 3. PERCEIVED QUALITY METER METHODS AND RECOMMENDATIONS

Over the last years, emphasis has been put on developing methods and techniques for evaluating the perceived quality of digital video content. These methods are mainly categorized into two classes: the *subjective* and *objective* ones.

The subjective test methods involve an audience of people, who watch a video sequence and score its quality as perceived by them, under specific and controlled watching conditions.

The following opinion scale used in an absolute category rating (ACR) test is the most frequently used in ITU-T [5]: excellent (5), good (4), fair (3), poor (2), and bad (1). The arithmetic mean of all the opinion scores collected is the MOS. The best known subjective techniques for video are the *single stimulus continue quality evaluation (SSCQE)* and the *double stimulus continue quality evaluation (DSCQE)* [15, 16].

The fact that the preparation and execution of subjective tests is costly and time consuming deprives their use in commercial mobile systems which aim at providing audiovisual services at predefined quality levels.

The objective methods are characterized and categorized into classes, according to the procedure of the quality evaluation.

One of these classes requires the source video sequence as a reference entity in the quality evaluation process, and is based on filtering the encoded and source sequences, using perceptual filters (i.e., Sobel filter). Then, a comparison between these two filtered sequences provides results, which are exploited for the perceived quality evaluation [17, 18].

Another class of objective evaluation methods is based on algorithms, which are capable of evaluating the PQoS level of the encoded test sequences, without requiring any source video clip as reference.

A software implementation, which is representative of this nonreference objective evaluation class, is the *quality meter software (QMS)* [19]. The QMS tool measures objectively the instant PQoS level (in a scale from 1 to 100) of digital video clips. The evaluation algorithm of the QMS is based on vectors, which contain information about the averaged luminance differences of adjacent pixels.

Table 1 summarizes ITU recommendations related to video quality assessment methodologies for video codec.

For all previous reasons, a lot of effort has recently been focused on developing cheaper, faster, and easily applicable objective evaluation methods, which emulate the results that are derived from subjective quality assessments, based on criteria and metrics, which can be measured objectively.

Due to the subjective methods limitations, engineers have turned to simple error measures such as mean-squared error (MSE) or peak signal-to-noise ratio (PSNR), suggesting that they would be equally valid. However, these simple measures operate solely on the basis of pixel-wise differences and neglect the impact of video content and viewing conditions on the actual visibility of videos.

PSNR does not take into account human vision and thus cannot be a reliable predictor of perceived visual quality. Human observers will perceive different kinds of distortions in digital video, for example, *jerkiness* (motion that was originally smooth and continuous is perceived as a series of distinct snapshots), *blockiness* (a form of block distortion where one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks), *blurriness* (a global distortion over the entire image, characterized by reduced sharpness of edges and spatial detail), and noise. These distortions cannot be measured by PSNR. ANSI T1.801.03-1996 standard [20, 21] defines a number of features and objective parameters related to the above-mentioned video distortions. These include the following.

- (i) Spatial information (SI) is computed from the image gradient. It is an indicator of the amount of edges in the image.
- (ii) Edge energy is derived from spatial information. The difference in edge energy between reference and processed frames is an indicator of blurring (resulting in a loss of edge energy), blockiness, or noise (resulting in an increase of edge energy).
- (iii) The difference in the ratios of horizontal/vertical (HV) edge energy to non-HV edge energy quantifies the amount of horizontal and vertical edges (especially blocks) in the frame.
- (iv) Temporal information (TI) is computed from the pixel-wise difference between successive frames. It is an indicator of the amount of motion in the video. Repeated frames become apparent as zero TI, and their percentage can be determined for the sequence.

(v) Motion energy is derived from temporal information.

The difference in motion energy between reference and processed video is an indicator of jerkiness (resulting in a loss of motion energy), blockiness, or noise (resulting in an increase of motion energy).

Motion energy difference, percent repeated frames, and other video parameters can then be combined to a measure of perceived jerkiness.

Starting from all the previous considerations, a considerable amount of recent research has focused on the development of quality metrics that have a strong correlation with subjective data. Three metrics based on models of the

peak signal-to-noise ratio (PSNR) [25]	0.793
structural similarity (SSIM) [26, 27]	0.8301
Blurring measure [28]	0.8963
Blocking measure [29]	0.83
motion sum of absolute differences (MSADs) [30]	0.819
video quality metric (VOM) [7]	0.98 (Always over 0.91)

TABLE 2: Pearson correlation index of the most reliable and famous objective video quality metrics.

human visual system (HVS) are summarized in [22]: the Sarnoff just noticeable difference (JND) model, the perceptual distortion metric (PDM) model developed by Winkler [23], and Watson's digital video quality (DVQ) metric [24]. Finally, a general purpose video quality model (VQM) was standardized by ANSI in July 2003 (ANSI T1.801.03-2003), and has been included in draft recommendations from ITU-T study group 9 and ITU-R working party 6Q.

The general model was designed to be a general purpose video quality model (VQM) for video systems that span a very wide range of quality and bit rates, thus it should work well for many other types of coding and transmission systems (e.g., bit rates from 10 kbits/s to 45 Mbits/s, MPEG-1/2/4, digital transmission systems with errors). Extensive subjective and objective tests were conducted to verify its performances. The VQM metric computes the magnitude of the visible difference between two video sequences, whereby larger visible degradations result in larger VQM values. The metric is based on the discrete cosine transform, and incorporates aspects of early visual processing, spatial and temporal filtering, contrast masking, and probability summation.

This model has been shown by the video quality experts group (VQEG) [6] in their phase II full reference television (FR-TV) test to produce excellent estimates of video quality for video systems obtaining an average pearson correlation coefficient over tests of 0.91 [7]. To the best of our knowledge, VQM is the only model to break the 0.9 threshold according to previous studies summarized in Table 2; for this reason, we chose to use it in our work as reference model for the PQoS evaluation during the training phase.

## 4. SYSTEM ARCHITECTURE AND TEST BED DEPLOYMENT

In this section, we describe the network architecture used for evaluating the perceived quality of the transmitted multimedia contents. We recorded several video clips with different bit rates; we used the digital video encoding formats MPEG-4 [31] because it is mostly preferred in the distribution of interactive multimedia services over IP; furthermore, MPEG-4 is also suitable for 3G networks providing better encoding efficiency at low bit rates, compared to the previous formats (MPEG-1, MPEG-2).

The network architecture is shown in Figure 1; it is composed by both wired and wireless segment. The service center



FIGURE 1: A Simple system architecture.

belongs to the wired segment and has the task of sending multimedia contents to the wireless clients (e.g., Laptop, PDA, smartphone, see-through glasses for augmented reality, generic head mounted displays HMD, etc.). On the wireless segment the transmission of multimedia contents can take place in both directions, from the clients to the access point (AP) and vice versa. This architecture can be used to provide real-time video with augmented reality: a classical example is offered by a client device equipped with a wireless camera that can be used by a visitor inside a museum; the camera can record and send the video of the ambient in which the visitor is walking to the service center that is in charge of locating the client and send him multimedia contents regarding the paintings or the art work recorded in the video previously sent. A similar service can be offered in an archaeological site to supply augmented reality area wireless network.

In order to emulate the previous scenario, we create different multimedia video and we transmit them from the wireless mobile device on the right side of Figure 1, to the service center and vice versa using VLC [32], (*VideoLAN* is a software project, which produces free software for video, released under the GNU general public license. VLC media player is a free cross-platform media player, it supports a large number of multimedia formats, without the need for additional codecs; it can also be used as a streaming server, with extended features (video on demand, on the fly transcoding, etc.).) a powerful software well suited for video streaming transmission.

With the aim of implementing a more realistic scenario, we considered also the data traffic generated from other mobile devices within the AP coverage area; this aggregated data traffic represents a set of different applications such as download of audiovideo contents, text files, or web surfing; it can be considered as "*background traffic*" handled by the access point without stringent delay constraints, nevertheless the amount of this background data traffic has, for sure, a heavy impact on the multimedia video transmission in terms of perceived quality, thus the evaluation of the PQoS metric and the resulting analytical models cannot be designed without considering this kind of traffic.

The background traffic was physically implemented as a download of huge data files, using the classical TCP transport stream. The bursty transmission behavior of TCP [33–35] makes PQoS estimation more challenging due to the variable wireless link occupancy. According to this consideration, the background traffic values used during the simulations have to be considered as mean values computed during the whole test. We did not use any analytical model nor synthetic traffic generator in order to emulate the real world scenario of data traffic coming from other applications.



FIGURE 2: The whole test bed system.

The whole system architecture is shown in Figure 2, we used another laptop for generating the aggregated background traffic; moreover, we gradually increased the amount of generated data traffic in order to study the effect on the perceived quality during the transmission on the wireless channel.

Concerning the background traffic values for the whole simulation campaign, the following remark is appropriate. We implemented a wireless IEEE 802.11g [36] network that can support nominal data rate up to 54 Mbps; yet in practice only half of the advertised bit rate can be achieved because wireless networks are particularly error-prone due to radio channel impairments; thus the data signals are subject to attenuation with distance and signal interference.

We observed, through few simple tests, that the perceived video quality is not degraded if the traffic background is smaller than 11 Mbps; this performance is due to the high link capacity supported by the specific AP. (The AP used for the test bed is a USRobotics Wireless MAXg Router 5461A.)

We experimented that 28 Mbps is the maximum background traffic sustainable by our wireless network.

Table 3 summarizes all the system parameters used for the test bed; the transmitted multimedia video contents were extracted from few minutes of an action movie with a high interactivity level in order to evaluate a worse case scenario in terms of variable bit rate; moreover, the contents were chosen in order to cover a wide range of possible applications such as video streaming conference with a variable bit rate satisfying different applications requirements. Each video clip was transcoded to MPEG-4 format, at various variable bit rates (VBR) according to the mean data rates shown in Table 3. Resolution ( $320 \times 240$ ) and constant frame rate of 25 frames per second (fps) were common parameters for the transcoding process in all test videos. These video parameters are typically supported by hand-held mobile devices.

Finally, we evaluated the system performances varying the wireless link quality in terms of signal power level. Using *network stumbler* [37] we obtained the signal power level values over the wireless link depending on the distance from the access point. For each parameter combination we took several samples repeating the perceived quality measurement 8 times with the aim of considering the natural wireless link and background traffic fluctuations.

In order to measure the video quality over the wireless network we used MSU [38]. This free program has many interesting features to evaluate the video quality according to several metrics (i.e., PSNR, DELTA, MSAD, MSE, SSIM, and VQM). Moreover, the obtained results are collected in a .CVS file, thus they can be easily managed through any spread sheet.

During the transmission over the wireless link, few frames can be lost due to low signal level or to high interference conditions; nevertheless, the software used for the PQoS evaluation needs to compare two videos with exactly the same number of frames, thus we implemented a realignment procedure for replacing the lost frames with the last frame correctly received in order to obtain a consistent analysis.

#### 5. TEST BED RESULTS AND ANALYTICAL MODEL

In this section, we show the results in terms of perceived video quality, obtained from the test bed varying the network parameters and we propose a simple analytical model for estimating the perceived quality.

Our model for PQoS estimation is based on simple parameters that can be easily computed in the first "*training phase*." The implementation of an integrated software for the perceived quality measurement of few video contents and the resulting calibration of the polynomial model coefficients are quite simple. In this way, the analytical model plays a primary role in the PQoS estimation and the consequent real-time video scaling or format adaptation. Finally, the proposed method for PQoS estimation can be integrated in any wireless telecommunication system satisfying the following requirements:

- every client has to periodically provide its received power level to the service center through specific backward signalling;
- (2) the service center needs to periodically monitor the data traffic, managed by the access point, and measure the background traffic in order to perform PQoS estimation and adapt the video format.

We remark that in our work, we evaluated PQoS in the training phase through the generic VQM objective metric for a specific source and channel coding techniques. In other words, once fixed the source coding, the channel coding, and the streaming protocol used during the training phase, these techniques should not be changed without repeating the training phase. That is a realistic situation since all the multimedia contents are provided by one service center.

#### 5.1. Fixing the wireless link quality

First of all we fixed the power level of the wireless signal to the best value (i.e., -15 dBm) in order to study the system performance in a very good condition in which the interference has a negligible effect; in this way the perceived

			, ,			
			Video bit rate			
		450 Kbps	810 Kbps	1470 Kbps	1870 Kbps	2350 Kbps
	0 Mbps	0.6452	0.5405	0.3552	0.3625	0.3246
Background traffic	5 Mbps	0.6452	0.5475	0.3552	0.3741	0.3439
	11 Mbps	0.6452	0.5546	0.3552	0.3857	0.3632
	22 Mbps	0.6927	0.5990	0.8080	0.9838	1.2077
	26 Mbps	0.8197	0.6737	1.1294	1.6501	2.4743
	28 Mbps	0.9212	0.7346	1.5547	2.1475	3.2797

 TABLE 3: System traffic parameters.



FIGURE 3: Perceived quality versus background traffic with different video bit rates.

TABLE 4: Network parameters for the training phase.

Video bit rate	Background traffic	Signal power level
$r_1 = 2350 \text{ Kbps}$	$b_1 = 0$ Mbps	$c_1 = -15  \text{dBm} (\text{excellent})$
$r_2 = 1870 \text{ Kbps}$	$b_2 = 5$ Mbps	$c_2 = -40 \text{ dBm} \text{ (good)}$
$r_3 = 1470 \text{ Kbps}$	$b_3 = 11 \text{ Mbps}$	$c_3 = -66 \text{ dBm} \text{ (fair)}$
$r_4 = 810 \text{ Kbps}$	$b_4 = 22 \text{ Mbps}$	$c_4 = -76 \text{ dBm} \text{ (poor)}$
$r_5 = 450 \text{ Kbps}$	$b_5 = 26 \text{ Mbps}$	
	$b_6 = 28 \text{ Mbps}$	

video quality is strictly linked only to the background traffic and the bit rates; the following analysis is oriented to discover the relationship between those two system parameters. Figures 3 and 4 show how the perceived quality decreases when both the background traffic and the bit rate of the transmitted video increase. Furthermore, background traffic values smaller than 11 Mbps do not influence the perceived quality index. Choosing an objective VQM value for each video, an accurate scaling can be done according to the trend of those curves. Table 4 summarizes all the measured quality values that will be used for the analytical model fitting.



FIGURE 4: Perceived quality versus video bit rates with different background traffic.

#### 5.2. Varying the wireless link quality

The link quality is for sure one of the most important parameters in the evaluation of standard QoS index in wireless networks. Its contribution in terms of PQoS is still an open and challenging issue that we consider in this section. Figure 5 shows the perceived quality index with different values of signal strength over the wireless link. This measurement has been carried out by fixing the background traffic value to 11 Mbps in order to study the signal power level effect in a mean working condition in which the background traffic presence cannot drastically affect the contribution of the signal power level. When the measured power level from the receiver is very low (i.e.,  $-76 \, \text{dBm}$ ), the VQM index does not depend on the video bit rate, in fact that curve fluctuates around 1.5 VQM value; thus in this condition, a video with low bit rate has almost the same quality of a video with high bit rate.

In the other two cases (i.e., -66 dBm and -46 dBm) the slight decrease of the VQM value is more evident on videos with higher bit rate.

Following the previous considerations we can argue that the signal power level over the wireless link is weakly related



FIGURE 5: PQoS varying the quality link and the bit rate.



FIGURE 6: PQoS varying the link power level.

to the video bit rate and the background traffic; for this reason we can treat the weight of the power level over the link as an additive value according to the trend in Figure 6. Thanks to the measurements carried out through the test bed we can approximate the trend of the curve with polynomial equation that will be used for designing the analytical model.

#### 5.3. Analytical model for estimating the PQoS value

The main goal of this section is the design of an analytical model in which all the previous PQoS measurements for our wireless network can be used in order to predict the VQM value in a fast, responsive, and reliable way. According to the curves presented in Figures 3 and 4 we pointed out the relations between the video bit rates and the background traffic; now we need to find a mathematical relation that can represent the trend of those curves.

As we already explained, the perceived quality is considered in our work as a function  $g(\cdot)$  of three parameters: the video bit rate *R*, the background traffic *B*, and signal power level over the wireless link *C*. Thus the PQoS can be expressed through the following relation:

$$PQoS = g(R, B, C).$$
(1)

For sake of simplicity we used the normalized version of those quantities according to the following formula:

$$x = \frac{X - \mu(X)}{\sigma(X)},\tag{2}$$

where  $\mu(X)$  and  $\sigma(X)$  are the mean and the standard deviation of the measured quantities, thus

$$PQoS = h(r, b, c).$$
(3)

As already explained in this section, the signal power over the wireless link is not strictly related with the video bit rate and the background traffic; for this reason, treating the wireless link strength as an additive value, we can rewrite the relation (3) as sum of two different functions

$$h(r, b, c) = f_1(r, b) + f_2(c).$$
(4)

Thanks to the measurements carried out through the test bed, we can fit both  $f_1$  and  $f_2$  functions using two polynomials, that is,

$$P_1(x, y) \cong f_1(r, b),$$
  

$$P_2(z) \cong f_2(c),$$
(5)

where

$$P_{1}(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} a_{ij} x^{i} y^{j}$$

$$= \sum_{k=0}^{m-1} \alpha_{k}(x) y^{k} = \sum_{k=0}^{n-1} \beta_{k}(y) x^{k},$$

$$P_{2}(z) = \sum_{k=0}^{\nu-1} c_{k} z^{k}.$$
(6)
(7)

During the training phase we estimate the  $a_{ij}$  and  $c_k$  coefficients in (6) and (7).

In our study, we used (n = 5) different values for video bit rate and (m = 6) different values for background traffic, thus we implemented a linear system of 30 equations in the unknowns  $a_{ij}$  for the polynomial  $P_1$  while we used (v = 4) values for the power level over the wireless link corresponding to a 4 equations linear system in the unknowns  $c_k$  for the polynomial  $P_2$ .

Table 4 shows the values  $(r_1, r_2 \cdots r_n)$ ,  $(b_1, b_2 \cdots b_m)$ , and  $(c_1, c_2 \cdots c_v)$  used for the video bit rate, the background traffic, and power link level, respectively.

Equation (8) provides the exact values of  $a_{ij}$  and  $c_k$  coefficients obtained through the proposed model:

$$(a_{ij}) = \begin{pmatrix} 0.5550 & -0.0464 & -0.1227 & -0.0472 & 0.0555 \\ 0.5579 & 0.2703 & -0.9139 & -0.1132 & 0.4152 \\ -0.2049 & 0.4568 & 1.3469 & 0.0296 & -0.4997 \\ -0.6381 & 0.2907 & 2.2181 & 0.1468 & -0.8619 \\ 0.4646 & 0.0246 & -1.0567 & -0.0464 & 0.4932 \\ 0.4925 & -0.0240 & -1.2682 & -0.0771 & 0.5484 \end{pmatrix},$$

$$(c_k) = \begin{pmatrix} 0.3541 \\ -0.2235 \\ 0.7584 \\ 0.6865 \end{pmatrix}.$$

(8)

TABLE 5: Network parameters for model validation.

Video bit rate	Background traffic	Signal power level
630 Kbps	14300 Kbps	-15 dBm
1000 Kbps	18560 Kbps	-32 dBm
1560 Kbps	24160 Kbps	-60 dBm

TABLE 6: VQM Values measured through MSU software, signal power level -15 dBm.

	Background traffic [Kbps]			
Bit rate [Kbps]	14300	18560	24160	
630	0.569878	0.61557	0.653899	
1000	0.539471	0.592827	0.694409	
1560	0.4555	0.561323	1.031743	

TABLE 7: PQoS Values estimated with the analytical model, signal power level -15 dBm.

	Ba	Background traffic [Kbps]		
Bit rate [Kbps]	14300	18560	24160	
630	0.609	0.608	0.647	
1000	0.497	0.535	0.676	
1560	0.415	0.552	1.023	

Thanks to the model, we can easily evaluate the performances of different scenarios through a colored scale representing the good mix (green and light green-areas) and the bad mix (red and dark-red) of system parameters in terms of perceived quality values. Many interesting considerations can be made observing Figures 7 and 8 because the relations between all the system parameters involved in the evaluation of the PQoS are mixed together. These figures are two different ways for representing the output of the PQoS estimation model according to the available system parameters; the colored maps can be examined fixing the signal power level (Figure 7) or fixing the video bit rate (Figure 8) and varying the other two parameters; in particular the PQoS index increases at higher video bit rate and background traffic. This causes a degradation in terms of perceived quality (see red and dark-red zone on Figure 7). On the other hand, fixing the video bit rate, the PQoS index increases (i.e., quality degrades) with the background traffic and the signal power level (see red and dark-red zones on Figure 8). Thanks to these maps the reader can appreciate in a visual manner the graceful scaling color of the estimated PQoS.

## 5.4. Testing the effectiveness of the analytical model

In this section we demonstrate the reliability of the proposed model showing the correlation between the measurements executed with the MSU software and the results obtained through the analytical framework.



FIGURE 7: PQoS map obtained from the analytical model, background traffic versus video bit rate.

In support of this analysis we recorded new videos with different parameters according to Table 5. The results summarized in Tables 6 and 7 have been obtained fixing the signal power level at -15 dBm; as we can see, the difference between the two approaches is a negligible quantity. The overall pearson linear correlation coefficient [39] between VQM quality and analytical model for the video sequences is equal to 0.986 making the proposed model very usefull. Finally, the accuracy of the proposed method can be valued looking at Figure 9 where the correlation between the values has been plotted.

In order to discover the possible limitations of our model we repeated the previous analysis taking few measurements with different values for the signal power level (i.e., -32 dBm and -60 dBm).

Figure 10 shows that the model fails only if the effect due to a suboptimal signal power level over the wireless link is coupled with a high background traffic value (i.e., 24160 Kbps). In these conditions the effects of the two phenomena are not predicted by our model (4). In such a case the data traffic over the wireless link is high and makes the network work very close to a congestion zone.

In conclusion, the proposed model is effective and robust up to 18560 Kbps of background data traffic in every tested wireless link conditions; these results make the model very useful and attractive for a wide range of realistic wireless network scenarios and video applications.



1.35 1.25 1.15 1.05 Analytical model 0.95 Bg traffic 24160 Kbps Bit rate 1560 Kbps 0.85 0.75 Bg traffic 18560 Kbps 0.65 Bit rate 1000 Kbps 0.55 Bg traffic 14300 Kbps Bit rate 630 Kbps 0.45 0.35 0.35 0.55 0.75 0.95 1.15 1.35 1.55 1.57 VQM • Signal power-32 dBm

FIGURE 8: PQoS map obtained from the analytical model for a fixed bit rate of 2350 Kbps, background traffic versus signal power level.



FIGURE 9: VQM quality versus analytical model quality, signal power level –15 dBm.

## 6. CONCLUSION

In this paper, we have measured the perceived quality of multimedia video contents transmitted over wireless LAN test bed based on the IEEE 802.11g standard. We studied the effects of network parameters on the PQoS index highlighting the connections between them. Finally, we designed an analytical model based on a simple curve fitting technique, well suited for wireless environment, for estimating the PQoS index in a fast and easy way. The proposed analytical model has an average pearson correlation coefficient of 0.986, as proof of its robustness and reliability in many network

FIGURE 10: VQM quality versus analytical model quality, signal power levels –32 dBm and –60 dBm.

• Signal power-60 dBm

conditions. Nevertheless, when the background traffic is very high and the signal power level is not excellent, the model does not work well because the combination of those two effects generates an unpredictable behavior in terms of PQoS. This analysis highlights few natural limitations of the proposed technique due to the congestion of the wireless network. Future work includes the testing of additional video sequences with different codec formats and resolutions in a multiuser scenario.

### REFERENCES

- H. Koumaras, A. Kourtis, C.-H. Lin, and C.-K. Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences," in *Proceedings of the 3rd International Conference on Networking and Services (ICNS '07)*, p. 62, Athens, Greece, June 2007.
- [2] O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 42083, 21 pages, 2006.
- [3] L. Qiong and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through realtime retry limit adaptation," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 278–290, 2004.
- [4] S. H. Shah, K. Chen, and K. Nahrstedt, "Dynamic bandwidth management for single-hop ad hoc wireless networks," in *Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom '03)*, pp. 195–203, Dallas-Fort Worth, Tex, USA, March 2003.
- [5] ITU-T Recommendation, "Subjective video quality assessment methods for multimedia applications," Tech. Rep. TU-T P.910, International Telecommunication Union, Geneva, Switzerland, September 1999.

- [6] "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," www.vqeg.org/.
- [7] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
- [8] M. Margaritidis and G. C. Polyzos, "MobiWeb: enabling adaptive continuous media applications over 3G wireless links," *IEEE Personal Communications*, vol. 7, no. 6, pp. 36–41, 2000.
- [9] B. D. Noble and M. Satyanarayanan, "Experience with adaptive mobile applications in Odyssey," *Mobile Networks and Applications*, vol. 4, no. 4, pp. 245–254, 1999.
- [10] P. Bahl, "Supporting digital video in a managed wireless network," *IEEE Communications Magazine*, vol. 36, no. 6, pp. 94–102, 1998.
- [11] A. Joshi, "On proxy agents, mobility, and web access," *Mobile Networks and Applications*, vol. 5, no. 4, pp. 233–241, 2000.
- [12] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bit rate control via joint adjustment of quantization and frame rate," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 441–446, 2007.
- [13] S. R. Gulliver and G. Ghinea, "The perceptual and attentive impact of delay and jitter in multimedia delivery," *IEEE Transactions on Broadcasting*, vol. 53, no. 2, pp. 449–458, 2007.
- [14] H. Koumaras, A. Kourtis, and D. Martakos, "Evaluation of video quality based on objectively estimated metric," *Journal* of Communications and Networks, vol. 7, no. 3, pp. 235–242, 2005.
- [15] ITU-R Recommendation, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. ITU-R BT.500-10, International Telecommunication Union, Geneva, Switzerland, 2000.
- [16] T. Alpert and L. Contin, "DSCQE experiment for the evaluation of the MPEG-4 VM on error robustness functionality," ISO/IEC—JTC1/SC29/WG11, MPEG 97/M1604, February 1997.
- [17] S. Wolf and M. H. Pinson, "In-service performance metrics for MPEG-2 video systems," in *Proceedings of the Measurement Techniques of the Digital Age Technical Seminar*, pp. 1–10, Montreux, Switzerland, November 1998.
- [18] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," in *Multimedia Systems and Applications II*, vol. 3845 of *Proceedings of SPIE*, pp. 266–277, Boston, Mass, USA, September 1999.
- [19] J. Lauterjung, "Picture quality measurement," in *Proceedings of the International Broadcasting Convention (IBC '98)*, pp. 413–417, Amsterdam, The Netherlands, September 1998.
- [20] ANSI, "Digital transport of one-way video signals parameters of objective performance assessment," Tech. Rep. ANSI T1.801.03-1996, American National Standards Institute, New York, NY, USA, February 1996.
- [21] ITU-R Recommendation, "Methods for objective measurements of perceived audio quality," Tech. Rep. ITU-R BS.1387-1, International Telecommunication Union, Geneva, Switzerland, January 2001.
- [22] Z. Yu and H. R. Wu, "Human visual system based objective digital video quality metrics," in *Proceedings of the 5th International Conference on Signal Processing (ICSP '00)*, vol. 2, pp. 1088–1095, Beijing, China, August 2000.
- [23] S. Winkler, "Perceptual distortion metric for digital color video," in *Human Vision and Electronic Imaging IV*, vol. 3644

of *Proceedings of SPIE*, pp. 175–184, San Jose, Calif, USA, January 1999.

- [24] A. B. Watson, "Toward a perceptual video-quality metric," in *Human Vision and Electronic Imaging III*, vol. 3299 of *Proceedings SPIE*, pp. 139–147, San Jose, Calif, USA, January 1998.
- [25] E. P. Ong, M. H. Loke, W. Lin, Z. Lu, and S. Yao, "Video quality metrics—an analysis for low bit rate videos," in *Proceedings* of *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 1, pp. 889–892, Honolulu, Hawaii, USA, April 2007.
- [26] Q. Li and Z. Wang, "Video quality assessment by incorporating a motion perception model," in *Proceedings of IEEE International Conference on Image Processing (ICIP '07)*, vol. 2, pp. 173–176, San Antonio, Tex, USA, September-October 2007.
- [27] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.
- [28] R. Barland and A. Saadane, "Reference free quality metric for JPEG-2000 compressed images," in *Proceedings of the 8th International Symposium on Signal Processing and Its Applications* (ISSPA '05), vol. 1, pp. 351–354, Sydney, Australia, August 2005.
- [29] R. Muijs and I. Kirenko, "A no-reference blocking artefact measure for adaptive video processing," in *Proceedings of the European Signal Processing Conference (EUSIPCO '05)*, Antalya, Turkey, September 2005.
- [30] M. Ries, O. Nemethova, and M. Rupp, "Motion based reference-free quality estimation for H.264/AVC video streaming," in *Proceedings of the 2nd International Symposium on Wireless Pervasive Computing (ISWPC '07)*, pp. 355–359, San Juan, Puerto Rico, USA, February 2007.
- [31] "Final Draft, International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC)," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Document JVT-G050, March 2003.
- [32] VideoLAN, "Developerd by students of Ecole Centrale Paris," www.videolan.org/.
- [33] T. Ott, J. H. B. Kemperman, and M. Mathis, "The stationary behavior of ideal TCP congestion avoidance," Tech. Rep., Bell Lab, Holmdel, NJ, USA, August 1996.
- [34] N. X. Liu and J. S. Baras, "Long-run performance analysis of a multi-scale TCP traffic model," *IEE Proceedings: Communications*, vol. 151, no. 3, pp. 251–257, 2004.
- [35] W. Ge, Y. Shu, L. Zhang, L. Hao, and O. W. W. Yang, "Measurement and analysis of TCP performance in IEEE 802.11 wireless network," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE* '06), pp. 1846–1849, Ottawa, Canada, May 2006.
- [36] D. Vassis, G. Kormentzas, A. Rouskas, and I. Maglogiannis, "The IEEE 802.11g standard for high data rate WLANs," *IEEE Network*, vol. 19, no. 3, pp. 21–26, 2005.
- [37] M. Milner, "NetStumbler 0.4.0.," 2004, www.stumbler.net/.
- [38] "MSU Video Quality Measurement Tool," http://www .compression.ru/video/.
- [39] A. W. Pearson, "The use of ranking formulae in R&D projects," *R&D Management*, vol. 2, no. 2, pp. 69–73, 1972.

## Research Article

# **Objectives for New Error Criteria for Mobile Broadcasting of Streaming Audiovisual Services**

## Heidi Himmanen,<sup>1, 2</sup> Miska M. Hannuksela,<sup>3</sup> Teppo Kurki,<sup>1</sup> and Jouni Isoaho<sup>1</sup>

<sup>1</sup> Department of IT, University of Turku, Turku 20014, Finland

<sup>2</sup> Turku Centre for Computer Science (TUCS), Turku 20520, Finland

<sup>3</sup> Media Laboratory, Nokia Research Center, P.O. Box 1000, Tampere 33721, Finland

Correspondence should be addressed to Heidi Himmanen, heidi.himmanen@utu.fi

Received 1 October 2007; Revised 13 April 2008; Accepted 2 June 2008

#### Recommended by David Bull

This paper demonstrates the need of and objectives for new error criteria for mobile broadcasting and the problems related to defining numerical error criteria for video services. The current error criterion used in digital video broadcasting to handheld (DVB-H), namely, multiprotocol encapsulation forward error correction (MPE-FEC) frame error ratio (MFER) 5%, was defined to enable instantaneous measurements but is not accurate enough for detailed simulations or postprocessing of measured data. To enable accurate transmission system design, parameter optimization, and performance evaluation, it is necessary to define new practical criteria for measuring the impact of transmission errors. The ambiguity of the MFER criterion is studied, and results for other conventional error criteria are derived from transmission system simulations and objective video quality measurements. The outcomes are compared to results from studies on subjective audiovisual quality. Guidelines are given on the next steps of developing new objective criteria for wireless and mobile video. It is suggested that subjective tests are performed based on the average length and average amount of errors derived from verified mobile radio channel models.

Copyright © 2008 Heidi Himmanen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Mobile broadcasting is a strong trend in modern telecommunications, and one of the driving forces is real-time television (TV) services to mobile terminals. One of the most popular mobile broadcasting standards is digital video broadcasting-handheld (DVB-H) [1] with two main services defined: broadcasting of streaming video applications and file delivery. These two service categories are of very different nature and have different system requirements. Streaming video services, such as TV programs, are real-time services with hard latency constraints. In video applications, some residual errors can be accepted, without sacrificing the subjective audiovisual quality. File delivery applications, on the other hand, require that the file is received or reconstructed correctly before it can be used, while delays are not as serious a matter as for streaming video.

In this article, we consider streaming video services and their error criteria on the transmission system. We take DVB-H as a case study. What brings more complexity to analyzing audiovisual quality is the lack of good objective measures. Further, subjective quality and the importance of audio or video elements are content-dependent. In DVB-H, the multiprotocol encapsulation-forward error correction (MPE-FEC) frame error ratio (MFER) criterion does not give an unambiguous measure of the quality of an audiovisual stream transmitted over the wireless network. Thus, the transmission system designers lack one sufficient tool for optimizing the system performance, as fair comparisons of different solutions cannot be carried out. Inaccurate error criteria can even lead to wrong conclusions about the optimal solutions and parameters. The baseline for this article is that the technical requirements and criteria for designing and optimizing communication systems should be defined based on the requirements set by the services and applications, but should be easily measurable using common existing tools.

The scope is to demonstrate the shortcomings of the current criterion and show the way forward in designing new criteria. The paper gives the transmission system perspective of streaming audiovisual services, video quality, and objective error criteria. We explain the requirements on the joint effort between transmission system designers, audio, and video codec experts, and researchers of usability and human-centred technology. The development of the new error criteria will require a huge amount of additional tests and measurements on channel and transmission error statistics and subjective tests to find threshold values for subjectively perceived acceptability. The paper explains what information and further testing are required from the application and subjective testing in order to design measures that meet the requirements for the transmission system criteria.

The article is arranged as follows. First, an overview of the audio and video compression for DVB-H is given in Section 2. DVB-H as a transmission system is presented in Section 3, and current obstacles in system optimization are illustrated in Section 4 using DVB-H simulation results. In Section 5, comparisons to available subjective quality test results are made. Section 6 gives some background and proposes objectives and test cases for transmission system testing, video codec parameter selection, and subjective testing. Finally, we conclude the article.

## 2. AUDIO AND VIDEO COMPRESSION FOR DVB-H

The IP data casting specifications of DVB-H recommend the use of the high efficiency advanced audio coding version 2 (HE AAC v2) [2] for audio compression and advanced video coding (H.264/AVC) [3] for video compression. Elementary units for transmission of HE AAC v2 and H.264/AVC bit streams are called an access unit and a network abstraction layer (NAL) unit, respectively. An integer number of access units or NAL units are typically encapsulated into one transmission packet. An access unit of HE AAC v2 contains a coded representation of a frame of audio samples. NAL units can be categorized to video coding layer (VCL) NAL units and non-VCL NAL units. VCL NAL units are typicallycoded slices of a picture, covering a certain spatial area of the decoded picture. Non-VCL NAL units are used to convey information that is only indirectly related to the decoding process of the coded pictures. Primary-coded pictures of H.264/AVC can be categorized to three types: instantaneous decoding refresh (IDR) pictures, other reference pictures, and nonreference pictures. An IDR picture contains only intra-coded slices and causes marking of all previous reference pictures to be no longer used as references for subsequent pictures. An IDR picture can, therefore, be used as a random access point for starting of decoding or joining a session and it also provides a resynchronization point for decoding after transmission errors have occurred. A reference picture is stored and maintained as a prediction reference for interprediction until it is marked no longer used for reference according to the reference picture marking process of H.264/AVC. A nonreference picture is not used for reference in interprediction and can, therefore, be removed from a bit stream without consequences to any other pictures.

There are no widely accepted objective methods for measuring subjective audiovisual quality. Certain methods,

such as the peak signal-to-noise ratio (PSNR), can be used in controlled conditions for pairwise comparison but are not generally suitable for quality measurement, for example, when there are more than one source for quality degradation, such as coding impairments and transmission errors [4]. Moreover, the subjective expectation of the quality, the compression efficiency, and the relative importance of audio and video depend on the type of audiovisual content [5]. Hence, large-scale subjective testing is ultimately the only accurate mean for audiovisual quality measurement.

#### 3. DVB-H AS A TRANSMISSION SYSTEM

#### 3.1. Link layer operations

DVB-H is based on the terrestrial DVB-T standard and was ratified by the European telecommunications standards institute (ETSI) in December 2004. The link layer of DVB-H is an amendment to the physical layer of DVB-T to enable better mobile reception and low-power consumption for handheld devices. A good overview of DVB-H can be found in [6].

The link layer operations are presented in Figure 1. The audiovisual content is passed to the link layer in internet protocol (IP) datagrams. The datagrams are encapsulated columnwise into an MPE-FEC frame, the size of which can be selected flexibly. The number of rows of an MPE-FEC frame can be 256, 512, 768, or 1024. The encoding of the MPE-FEC frame using a Reed-Solomon (RS) (255,191) code [1] is performed rowwise, which results in an interleaving scheme referred to as virtual time-interleaving. By varying the amount of application data columns (1-191) and RS data columns (0-64), different code rates can be achieved. If all application and RS data columns are used, the MPE-FEC code rate is 3/4. MPE-FEC code rates are not fixed by the standard, but commonly considered options are 1/2, 2/3, 3/4, 5/6, 7/8, and 1, which represent uncoded link layer. The Reed-Solomon code can correct as many erasures on each row as there are redundancy columns. Thus, with code rate 3/4 up to 64, erasures can be corrected per row.

For transmission, the MPE-FEC frame is divided into sections. An IP datagram forms the payload of an MPE section, and an RS redundancy column forms the payload of an MPE-FEC section. The MPE sections are transmitted first, followed by the MPE-FEC sections. Both are transmitted in a moving picture experts group-2 (MPEG-2) transport stream (TS) format [7].

*Time-slicing* is applied to enable power saving, so that one MPE-FEC frame is transmitted in one time-slice burst. The TS bitrate during the burst is significantly higher than the service bitrate, and the receiver can turn off its radio parts between the bursts to save power. The frame size, transmission bitrate, and offtime between bursts are parameters that affect the video bitrate, service switching time, and power saving. That is, with an IP bitrate of 384 kilobits per second (Kb/s), one 512-row frame contains 1.8 seconds and a 1024-row frame 3.6 seconds of video.

DVB-H contains a large set of network and service-independent parameters. In addition to the link layer operation



FIGURE 1: The DVB-H link layer operations.

described here, there are a set of physical layer parameters, such as modulation, code rate, guard interval length, and orthogonal frequency-division multiplexing (OFDM) mode. With such a large set of options, simulations are usually the most efficient way to find the optimal parameter combinations.

#### 3.2. Current DVB-H error criteria

The DVB-T standard specifies the *C/N* threshold needed to reach the quasierror-free (QEF) reception criterion, which means one uncorrected error event per hour. Due to the high variations occurring in a mobile channel, the QEF criterion is not suitable for instantaneous measurements for mobile broadcasting. Also, in mobile broadcasting, looser error criteria have been accepted than for fixed reception. The common error criterion for DVB-H has been defined as MPE-FEC frame error ratio (MFER), and the quality of restitution (QoR) limit has been set to MFER 5% [6]. In addition to MFER, the erroneous seconds ratio (ESR) criterion has been occasionally used in some measurements. ESR is defined as seconds with errors over the observation period [6].

The MFER error criterion enables instantaneous laboratory measurements. The length of one measurement has usually been 100 frames, of which 5 can be erroneous. Further, the service bitrate has been increased, that is, the offperiod has been shortened, to enable faster measurements. Still, it is a highly time consuming project to perform extensive DVB-H measurements, including all possible combinations of constellations, fast fourier transform (FFT) sizes, guard intervals, code rates, and burst lengths covering pedestrian and vehicular use cases. According to [6], the observation period for field trials has been reduced to one time interval, corresponding to one time-slice burst, as the QoR assessment should be instantaneous.

The MPE-FEC frame error criterion is too inexact to evaluate the impact of the channel and system parameters on subjective audiovisual quality. Optimizing the system parameters using only the MFER, 5% criterion might even be misleading and result in incorrect conclusions about the system performance. As systems are also designed, optimized, and verified using simulations or postprocessing of recorded traces from laboratory measurements or field trials, particular IP packet or even on byte level information can be received. There is definitely a need for more accurate error criteria than frame error-based measures.

#### 3.3. Selection of DVB-H transmission parameters

The DVB-H implementation guidelines [8] give recommendations for parameter selections in DVB-H networks. For the physical layer modulation and code rates quadrature phaseshift keying (QPSK) or quadrature amplitude modulation (16-QAM) with code rates 1/2 or 2/3 are recommended. The choice is a compromise between robustness to transmission errors and throughput bitrate. QPSK 1/2 gives a bitrate of 5 Mbps, whereas 16-QAM 1/2 gives a bitrate of 10 Mbps, using guard interval 1/4 of the OFDM symbol duration. [8] recommends the use of 16-QAM 1/2 or 16-QAM 2/3 for mobile and portable reception.

The selection of FFT mode is based on the expected maximum velocity of the receiver. The 8K FFT mode, which is used in most DVB-T networks, gives the largest coverage area, but provides the lowest receiver velocities compared to 2 K and 4 K. Based on [8], when MPE-FEC is used and DVB-H physical layer parameters are selected properly, the use of the 8K mode is feasible at speeds up to 120 km/h. Transport

stream





FIGURE 2: Consecutive (a) and parallel (b) transmission of different DVB-H services.

The selection of guard interval is based on network topology. For the 8K mode, guard intervals 1/4 or 1/8 are recommended, of which 1/4 tolerates longer single-frequency network (SFN) delays.

Simulations in [9] used several different channel models for DVB-H and showed that, for networks intended primarily for vehicular use, the preferable combinations of modulation, convolutional code rate, and MPE-FEC code rate would, respectively, be QPSK 1/2 3/4, QPSK 1/2 5/6, QPSK 2/3 5/6, 16-QAM 1/2 3/4, or 16-QAM 1/2 5/6. Based on the recommendations and results in [8, 9], the parameters used for evaluating the performance at IP level in Sections 4 and 5 were chosen to be 16-QAM 1/2 3/4, FFT size 8K, and guard interval 1/4. Additionally, in some presented comparisons MPE-FEC is not used, that is, the MPE-FEC code rate is then 1.

When the transmission network is optimized properly, the transmission parameters do not have a direct impact on the video quality but on the size of the coverage area and the capacity of the network. On the other hand, transmission parameters, multiplexing scheme, environment, and movement of the receiver will affect the length and amount of error bursts. In general, when the receiver moves slowly, that is, the channel changes slowly, the error bursts are longer, as the receiver stays in the area with bad reception for a longer time compared to a fast changing channel.

#### 3.4. Multiplexing of services in DVB-H systems

DVB-H services may be transmitted consecutively or in parallel. Consecutive transmission means that only one MPE-FEC frame carrying one service is on air at a time. [8] does not present parallel transmission of services as the main but suggests that IP encapsulators and receivers should support this mode of transmission. Examples of consecutive and parallel transmission of DVB-H services are depicted in Figure 2, where each fill pattern represents one MPE-FEC frame carrying one service.



FIGURE 3: Measurements for evaluating video quality [10].

Parallel transmission can be useful if the service bitrates are very low. Using consecutive transmission in short bursts leads to degradation in time diversity. In mobile transmission, a good choice of burst length would be more than 100 milliseconds. Consecutive transmission, on the other hard, is the main source for the power saving in receivers achieved in DVB-H when compared to continuous parallel transmission of all services.

A special case of transmission would be to transmit several services in every MPE-FEC frame. This could be preferred, for example, if the services are statistically multiplexed together, so that the total capacity of these services is constant. This scheme was utilized in [5] and thus in the results presented in Section 6. With this transmission format, the MFER error criterion becomes even less accurate. An MPE-FEC frame might contain errors after decoding that do not occur in the MPE-sections carrying the data from the wanted service. Thus, the received data could be error-free even if the errors in the MPE-FEC frame cannot be corrected.

## 4. VISIBILITY OF PACKET LOSS IN MPEG-2 AND H.264/AVC VIDEO

Reibman and Kanumuri et al. have studied the visibility of packet loss in MPEG-2 and H.246/AVC in many papers, for example, in [10–13]. In [10], the need for accurate video quality measures is explained in detail. The approach is similar as in this paper. Figure 3 illustrates three measurement points discussed in [10]. Measurement C corresponds to the transmitted bitstream itself and could be taken either at the input to the decoder or inside the network. Measurements in C assume the use of nonreference methods, as the original video is not available for comparison. The new error criteria for mobile broadcasting of streaming audiovisual services considered in this paper should similarly be nonreference video quality measures in point C. However, assumptions about video coding parameters and used concealment algorithms have a significant impact on the perceived quality.

When measuring network performance and error behavior, it is usually preferred to measure over the whole multiplex, that is, over all service. This is the conventional use of the MFER criterion in laboratory and field measurements. However, the subjectively perceived quality can only be measured over one service. This problem has also been recognized by Reibman. The goal in [11] was to have a method to predict the quality of individual videos with low-enough complexity that it can be easily applied to many different video streams being sent across the network. Similarly, when designing the new criteria for mobile broadcasting, we need to move away from the approach of error measures for the whole multiplex. Measuring service specific quality is especially important in time division multiplex (TDM) systems, such as DVB-H, as the packet loss in mobile channels is strongly time variant. Thus, the different services might experience very different error behavior. This is discussed further in Section 5.

The previous work on visibility of packet loss can partly be used for designing new criteria for mobile broadcasting. Still, the approach in [10–13] has been different from the assumptions that have to be made for mobile broadcasting. In the mobile environment, errors will always exist. More important than finding the limit for visibility of packet loss or errors is to find the limit for acceptability of errors. Further, we must make the assumptions of using the simplest receiver, which is described in the implementation guidelines [8], and the simplest decoder. This also includes the assumption that concealment algorithms are not used, and the length of an error in the video cumulates to the next nonpredicted frame (IDR frame in the case of H.264/AVC).

## 5. DIVERSE ANALYSES OF MFER AND OTHER CONVENTIONAL ERROR CRITERIA

In this section, the MFER criterion is analyzed both from the transmission system and video codec perspectives. The ambiguous character of the MFER measure is demonstrated by analyzing it together with two transmission error criteria, namely, IP packet error ratio and byte error ratio, and two objective video quality metrics, namely, peak signal-to-noise ratio (PSNR) and the national telecommunications and information administration (NTIA) video quality metric. In Section 5.1, the transmission system simulation setup is described, and the results are presented in Section 5.2. In Section 5.3, the IP error statistics are analyzed close to the limit for subjectively acceptable quality. The objective video quality analyses are presented in Section 5.4 and, the shortcomings of the MFER criterion are analyzed in Section 5.5.

# 5.1. Simulations on different MPE-FEC decoding strategies

Different MPE-FEC decoding strategies for DVB-H were presented and analyzed by the author in [14, 15]. The decoding method suggested in the DVB-H standard is referred to as section erasure (SE) decoding. An MPE section or an MPE-FEC section is marked as an erasure, if it contains an error, and discarded in the decoding process. SE decoding provides neither efficient MPE-FEC decoding nor video decoding, as a lot of correct data is dropped at the link layer. However, using SE decoding is optional, and the final decision on the decoding strategy is left to the receiver designer. The most efficient of the suggested decoding methods is hierarchical transport stream decoding (HTS), which uses three levels of erasure information: correctly received TS packets, erroneous TS packets, and lost TS packets. HTS provides very good byte-level error performance.

To evaluate the performance of the different decoding strategies, simulations were carried out in the channel models developed for DVB-H [16] similarly as in [17]. The used models are pedestrian outdoor (PO), vehicular urban



FIGURE 4: MPE-FEC frame error rates (MFER), IP packet error rates (IP PER) and byte error rates (SER) after coded, and uncoded data link layer for the Vehicular Urban channel.

(VU) and motorway mural (MR), corresponding to the velocities of 3 km/h, 30 km/h, and 100 km/h, respectively. The physical layer parameters were 16-QAM modulation with convolutional code rate 1/2, 8K OFDM mode, and guard interval duration 1/4 of the OFDM symbol duration. Error traces from the physical layer were established to allow fast simulations at transport stream packet or byte levels. Error traces are series of binary indicators expressing whether a data block contains errors, in this case after the physical layer error correction decoding. The simulated link layer parameters were as follows: MPE-FEC code rate was 3/4 or 1, 512 rows were present in MPE-FEC frames, and an IP packet of length was 512 bytes. The error rates were measured over all services, that is, over the whole transport stream. The services were multiplexed so that one service always uses the whole bandwidth for transmitting the time-slicing bursts. The results are presented in Section 5.2.

## 5.2. Frame, packet and byte error ratios

Figure 4 illustrates different error ratios using SE decoding or uncoded DVB-H link layer (for which MPE-FEC code rate is equal to 1) in the Vehicular Urban channel, corresponding to a velocity of 30 km/h. The frame error ratio for uncoded link layer data (FER uncoded) is above 30% for all simulated carrier-to-noise ratios (C/N). Yet, when studying IP packet error ratio (IP PER) and byte or symbol error ratio (SER) for uncoded data, it is seen that there is much more correct data than the frame error ratio implies. When comparing IP PER for SE and uncoded, the difference of C/N yielding the same IP PER is only 1.3 dB. When designing the system for the presented C/N values based on frame error ratio, MPE-FEC code rate 1 could have been discarded from list of good parameter options.

However, when defining the system parameters based on another error criterion, uncoded link layer could be

		Pedestrian outdoor	Vehicular urban	Motorway rural
C/N at	SE	13.1 dB	14.3 dB	14.6 dB
MFER 5%	HTS	12.8 dB	13.9 dB	13.2 dB
IP PER & SER	SE	4.0%	2.2%	1.5%
IP PER	HTS	4.8%	2.2%	2.0%
SER	HTS	1.6%	0.6%	0.2%

TABLE 1: Carrier-to-noise ratios, IP packet error ratios, and byte error ratios at MFER 5% in the different channels.

a possible choice, as less redundancy is needed. Previous work has shown that good transmission modes also can be found among those not using MPE-FEC coding. In [17], different modulation and code rates were compared based on the IP PER 1% criterion, using SE decoding for all link layer code rates in the PO, VU, and MR channels. When also considering the different service bitrates achieved using different code rates, uncoded link layer was included in the list of good modes. For the PO channel, the uncoded mode was even recommended. If MFER 5% had been used in this comparison, the conclusions would have been very different.

Table 1 demonstrates the ambiguity of the MFER 5% criterion. The C/N required for achieving the MFER 5% point is given for SE and HTS decoding with MPE-FEC code rate 3/4. Other simulation parameters were similar as for the simulations in Figure 4. The IP packet error ratios and byte error ratios were measured at the MFER 5% point. For SE decoding, IP PER and SER give the same results, as with SE decoding all bytes of an erroneous IP packet are erased, which is not the case with HTS decoding. As HTS decoding provides low-byte error ratios, the SER at MFER 5% is very low compared to SE decoding, especially in the Vehicular Urban and Motorway Rural channels. The error ratios also demonstrate the effect of the receiver velocity. At high velocities, an erroneous frame contains less erroneous data than at low velocities. This is mainly due to the fact that error bursts are shorter at high velocities, as the channel changes faster. At high velocities, the amount of errors at the MFER 5% point is different from the error amounts at low velocities. The same also applies to the length and frequency of the error bursts.

The amounts of erasures occurring in the MPE-FEC frames are illustrated for the different channel models in Figure 5, where the distribution of instantaneous IP PER values for each frame is given. The curves represent the situation, where average IP PER is 10%, when MPE-FEC coding is not utilized (uncoded). The figure shows significant differences in error distributions between the different channel models. The curve of the pedestrian model is very steep, whereas for vehicular speeds, there is a large amount of frames with less than 25% of the IP packets erased. Using MPE-FEC code rate 3/4, all frames with IP PER less than 25% would be corrected. The different distribution of errors leads to different MPE-FEC decoding performance even though the average IP PER over all frames is equal.



FIGURE 5: IP packet error ratio for each MPE-FEC frame in different channel conditions [17].

## 5.3. IP error statistics in three different channels at the limits for subjective quality

Some results for subjective audiovisual assessment in DVB-H are available in [5], aiming to discover the approximate value of MFER that is the threshold between subjectively acceptable and unacceptable audiovisual quality. Extensive subjective testing was carried out with four clips of different content types coded according to the lowest interoperability point specified for IP data casting over DVB-H at timeslice interval of about 1.5 seconds [5]. It was concluded that with the tested clips, the boundary of acceptability and unacceptability lies between 6.9% and 13.8% in terms of MFER.

Let us now compare the IP error statistics for the simulated channels with the results from the subjective tests [5]. In Table 2, the IP PERs for MFER 6.9% and 13.8% are presented for MPE-FEC code rate 3/4. As above, the IP packet length was constant 512 bytes. Compared to the VU channel, the MR channel has only slightly lower IP PERs at these MFERs, whereas the PO channel has double the amount of errors.

When measuring the performance of a transmission system, the measurements are performed over the whole transport stream, whereas in subjective quality measurements, the results are gathered for a single service. To enable comparison to subjective tests results in Section 6, a 60second measurement over the whole multiplex is performed. With the used modulation and coding, this corresponds to transmitting 58 video services of capability class A at 128 Kbps or 29 video services of capability class B at 384 Kbps (see Table 4).

In Table 3, comparisons of the IP packet error characteristics of the channels are presented with MPE-FEC code rate 3/4. It is found that the MR channel has shorter error bursts

MFER	IP PER in pedestrian outdoor	IP PER in vehicular urban	IP PER in motorway rural
6.9%	5.5%	2.5%	2.0%
13.8%	11.5%	5.4%	4.4%

TABLE 2: IP PER at MFER 6.9% and 13.8% in three different channels.

TABLE 3: IP error statistics for three different channels measured over 60 seconds.

	Pedestr	ian outdoor	Vehicular urban	Motorway rural
C/N	12 dB	13 dB	14 dB	14 dB
MFER	19.1%	5.7%	10.9%	13.3%
IP PER	18.5%	5.3%	4.0%	4.4%
Amount of IP errors	21830	6236	4704	5134
Average error burst (AEBL)	23.05	19.01	23.76	15.1
StDev of error bursts	115.76	85.18	29.24	10.94
Max error burst	1262	716	147	71
Amount of error bursts	947	328	198	340
Error bursts > 80 packets	24	11	11	0

TABLE 4: Capability classes for DVB-H [18].

Capability	Frame size [pixel]	Frame rate [Hz]	Max bitrate [Kb/s]
А	QCIF: 176 × 144	15	128
В	QCIF: 176 × 144	30	
	QVGA: 320× 240	15	384
С	QVGA: 320 × 240	30	768

than VU at higher MFERs and IP PERs. This indicates that in the MR channel there are more but shorter error bursts. Also, in the PO channel, the average error burst is shorter for a higher error rate than in the VU channel. However, in the PO channel, there are also longer errors than in the VU channel. The variation in length of the error bursts is much larger in the PO channel, whereas for the VU channel, the error lengths are closer to the average. The comparison shows that the error characteristics are very different in different channels, when studying error rates close to the limit for subjectively accepted video quality.

#### 5.4. Objective video quality measurements

The MFER 5%, as an error criterion, can introduce errors of very different lengths and severity to the video stream. To understand and measure these errors better, a set of simulations and objective measurements was performed. The video used was a 180-second clip, corresponding to 100 MPE-FEC frames, recorded from a TV news broadcast. The content was comparable to a typical news broadcast, including low or no motion scenes showing the newsman or generated graphics and high-motion material from different reporting locations. Resolution, frame rate, and bitrate were chosen to be  $320 \times 240$ , 15 Hz, and 384 Kb/s, respectively. The bitrate for the video stream included header overhead, the actual VCL bitrate being 353 Kb/s. No audio track was used for the content.

Video encoding was performed using Nokia H.264 encoder [19] with default settings, except for resolution, frame rate, and bitrate control. Error concealment was not used, as it is an optional feature for DVB-H services. IDR frames were inserted every 1.8 seconds, corresponding to at least one IDR frame in each MPE-FEC frame. The resulting NAL units were encapsulated to IP packets, achieving an average IP packet length of 512 bytes. These IP packets were then inserted into 100 MPE-FEC frames, using 191 application data columns and 512 rows. Corruption was introduced into 5 of the 100 frames using section erasure with IP PER values of 0.026%, 1.7%, and 5.0%, corresponding to the loss of 1, 65 and 191 IP packets per each erroneous MPE-FEC frame. The MPE-FEC frames were decoded using SE decoding. When using code rate 3/4 and the IP packet lengths being equal to the amount of rows in the frame, these amounts represent some extreme cases of residual errors in the MPE-FEC frame. 191 erased IP packets correspond to one completely corrupted MPE-FEC frame. 65 erased IP packets corresponds to the smallest amount of erasures that cannot be corrected with code rate 3/4, when all erased sections are carrying application data. The loss of one IP packet occurs, if all 64 RS redundancy columns are erased and one application column.

Video quality was assessed using three metrics. Despite its drawbacks, PSNR was used as a primary comparison metric due to its ability to provide results for individual video frames. Secondary metric used was the NTIA VQM [20], which is far more complex than PSNR. NTIA VQM tries to account for, for example, jerky motion, blocking, blurring, and other impairments typical to digital video and has been shown to correlate with subjective measurements very well. The third metric used was erroneous seconds ratio (ESR). A second (15 frames) of video was considered to be erroneous if it contained more than 3 successive visibly erroneous frames, corresponding to 200 milliseconds detection threshold [21]. A PSNR difference of 1 dB was considered as error visibility threshold in error assessment.

Average results obtained from the PSNR metric seem to degrade linearly as the IP PER rises. However, profound conclusions should not be drawn from the PSNR scores due to the drawbacks mentioned in Section 2. The NTIA VQM scores seem to indicate that on average, the video quality is



FIGURE 6: PSNR video quality results in MFER 5% with 1.7% and 5.0% IP packet error ratios.

acceptable in all test cases in Table 2. Acceptability threshold for NTIA VQM is around 0.5 [20], corresponding to the border of "fair" and "poor" quality (lower score is better). Despite the good VQM results, the erroneous seconds ratio (ESR) for both the 1.7% and 5.0% IP PER exactly meet the ESR 5% criterion, which is considered to be the limit for acceptable quality in [6]. This can be explained by the ESR metric not accounting for severity of the errors. Errors are clearly longer than the amounts of dropped frames indicate, mostly due to error propagation and the rather sparse placement of the IDR frames. In any case, it seems evident that MFER 5% does not provide an unambiguous error criterion compared to other metrics.

Detailed PSNR results for the 1.7% and 5.0% IP PER simulations are depicted in Figure 6. In addition, the PSNR curve of the error-free video is provided for comparison. These results are derived from the same simulations as the average values in Table 5. Five error bursts and their corresponding drops in terms of PSNR are clearly visible in the figure. Error bursts that occurred during low or no motion scenes, pointed out with arrows, have a significantly smaller quality drop. The result is logical, since losing frames from relatively static content produces only barely, if at all, visible errors. The remaining three error bursts coincide with a high-motion scene, resulting in extremely low-PSNR values, typically 10-15 dB. Such low values result from the dropped frames and do not provide basis for a meaningful comparison as such. Regardless, it is evident that loss of frames in a high-motion scene is critical for the perceived video quality. Due to the low similarity of successive frames in this type of content, a significant amount of information is lost in each burst. It is also notable that with 1.7% IP PER, the PSNR value has a tendency to rise after the initial drop at the start of each error burst. However, error propagation will continue impairing the video until the next nonerroneous IDR frame is encountered, and the video quality returns to optimal levels.

#### 5.5. The shortcomings of the MFER criterion

MFER fails to express many characteristics that would be important for DVB-H system design, some of which are

TABLE 5: Video quality measurement results at MFER 5% atdifferent IP packet error ratios.

IP PER	PSNR [dB]	NTIA VQM	ESR	Average lost picture frames per error burst	Average error length per error burst
0.0%	36.61	0.206	0.0%	_	_
0.026%	35.60	0.208	2.2%	0	0.8 s
1.7%	29.00	0.224	5.0%	6.8	1.8 s
5.0%	27.34	0.227	5.0%	27.0	1.8 s

described in the following. First, MFER does not indicate the relation between the frequency of the errors and their duration. For example, MFER equal to 5% corresponds to one and six erroneous time-slice bursts per minute in streams with 3-second and half-a-second time-slice intervals, respectively. It is not obvious how the frequency and duration of clearly perceivable audiovisual errors impact the subjective quality. Second, MFER does not indicate the residual error rate affecting the content of the erroneous frames. For example, the same value of MFER can result from two different error conditions of very different symbol error rates due to different code rate in MPE-FEC. Audio and video decoders may be able to conceal a relatively small residual error rate satisfactorily, but when it exceeds a threshold, most viewers consider the audiovisual quality as unacceptable regardless of the residual error rate. Third, the distribution of residual errors may play a role in subjective quality. For example, an error burst may not affect the entire time-slice, but the start or the end of the time-slice may be intact. Moreover, the method for transmission can affect the distribution of residual errors. One example is provided in [22], where unequal error protection has been proposed to protect audio, video IDR pictures, and other reference pictures more strongly compared to nonreference pictures. Fourth, the operation of the protocol stack and source decoders may be optimized differently in receiver operations when it comes to handling of transmission errors. For example, some DVB-H receivers may implement the HTS method, while others use the SE decoding. Furthermore, error concealment algorithms have not been specified in audio and video codec specifications, hence resulting into different implementations in source decoders.

In broadcasting, error criteria have been conventionally defined as accepted error events during a certain time. In DVB-T, the accepted limit for quasierror-free reception is one erroneous event per hour. Due to low-transmission error rate and common structures for groups of pictures in which intra-coded pictures are periodically and frequently included, the measure of error events per time is sufficient enough in DVB-T. In mobile broadcasting, varying reception conditions and wider range of possibilities for error protection code rates, time-slicing intervals, and group of picture structures make the measure of error events during a certain time unsatisfactory. In the third generation partnership project (3GPP), some objective quality of experience metrics have been specified [23]. Burst errors are measured using a corruption duration metric, indicating the amount of successive corrupted pictures and successive loss of IP packets. However, the relation of these metrics to subjective quality has not been quantified. Moreover, no numerical limits for these quality metrics have been defined in 3GPP.

## 6. COMPARISON TO SUBJECTIVE ACCEPTANCE OF AUDIOVISUAL QUALITY

The subjectively perceived audiovisual quality of TV services over DVB-H has been studied in [5, 24]. In these studies, the error patterns used to simulate errors caused by the wireless channel were achieved by using channel characteristics from field measurements in a Gilbert-Elliot model. The results can be compared to the vehicular urban (VU) channel model used in this paper, as the field tests were carried out in a similar environment with a car rooftop antenna. The MPE-FEC code rate was 3/4. QCIF videos were coded with an H.264/AVC encoder at bitrate 128 Kbps and at a frame rate of 12.5 Hz. One IDR picture was encoded per each timeslicing burst. Monaural audio at 32 Kbps and 16 Hz sampling frequency was used. No error concealment was used in the tests. The limit for acceptable and unacceptable audiovisual quality was found to be between MFER 6.9% and 13.8% [24]. There were 30 evaluators in the tests, and each clip was played three times, varying the error locations in the audiovisual stream. The length of the clip was around 60 seconds.

Figures 7 and 8 present the average error length and amount of error bursts in the video and audio streams for the tests in [5] for all tested content types: news, sports, music video, and animation. Each point corresponds to one test case, a combination of the content type, and error trace, rated by all evaluators. The filled (solid) points for MFER 1.7% and 6.9% represent acceptable quality, and the unfilled (hollow) points for MFER 13.8% and 20.8% represent unacceptable quality. It seems that the acceptability is more based on the amount of errors than the duration of these. The limit for acceptability of video is between 4 and 6 errors, and for audio between 5 and 7 errors on the average with the used content and parameters.

As explained in Section 3.4, each service should be carried in its own MPE-FEC frame to achieve maximum power saving in receivers rather than transmitting several services in each MPE-FEC frame as in [5]. This means that the used service specific error traces should not be considered to represent conventional DVB-H services. The used multiplexing has probably also caused the surprising error lengths, where the lowest MFER gives the longest errors. What can be used are the ratings and classification into acceptable and unacceptable quality of the different contents with the different amount and duration of errors, as in Figures 7 and 8. Still, new subjective tests are required to fully understand the acceptability of typical error behavior in mobile and portable channels with different encoding parameters, bitrates, and content types. The requirements for the future subjective tests are described in Section 7.3.



1.7% △ 13.8%
 6.9% ◇ 20.7%

16

14

12

10

8

6

4 2

0

0

Amount of errors

FIGURE 7: Video errors for MFER 1.7%, 6.9%, 13.8%, and 20.7% for the test performed in [5].

Average error duration



FIGURE 8: Audio errors for MFER 1.7%, 6.9%, 13.8%, and 20.7% for the test performed in [5].

#### 7. DESIGNING THE NEW ERROR CRITERIA

As described in the previous sections, the MPE-FEC frame error ratio criterion does not provide sufficient means for system design and optimization of DVB-H. There is a need for more appropriate error criteria that would represent the subjective impact of transmission errors on the services and applications. Many challenges in defining such criteria relate to the difficulty to derive an objective measure reflecting the subjective experience of audiovisual content, as the expectation for the experience and the relative weight of audio and video elements depend on the content. Still, the error criteria should be easy to measure, using tools familiar to transmission system designers.

#### 7.1. Transmission system aspects

The performance of DVB-H in different channel models and use cases measured in the laboratory and in the field were compared in [25]. Five parameters for comparing packet channel characteristics were presented in [26] as follows.

- (1) Packet error ratio (PER).
- (2) Average error burst length (AEBL). The AEBL parameter describes the average length of all error bursts. The error burst length is defined as the amount of consecutive erroneous units, that is, amount of erroneous packets between two correctly received packets.
- (3) Variance of error burst lengths (VEBL).
- (4) *Mean time between errors (MTBE).* The MTBE parameter describes the average length of the time between errors. The time between errors is defined as consecutive correctly received units, that is, amount of correctly received packets between two erroneously received packets.
- (5) Variance of time between errors (VTBE).

These parameters have shown to successfully model packet error behavior in packet channels with constant length packets. For streaming audiovisual services, the IP packets are usually of variable lengths. In the next comparison, a constant IP packet length of 512 bytes has been assumed to enable IP PER comparisons, as in the previously presented simulations.

To illustrate that the error behavior is service specific, the AEBL, MFER, IP PER, and TS PER are shown for a complete multiplex and for 16 services separately. The laboratory and field measurements are the same as used in [25] with 16-QAM modulation and convolutional code rate 1/2. The MPE-FEC code rate was 3/4. The multiplex of 9.95 Mbps was carrying 16 equally multiplexed services, each with a bitrate of 622 Kbps at TS level. The error behavior was measured over a stream corresponding to transmission time of 10 minutes.

In Figure 9, the AEBL at TS level is shown for the whole multiplex "All," the average AEBL over all 16 services "Mean", and for each service separately. In all cases, the TS PER over all services is 4-5%. The simulations show that the error behavior is service specific and varies most in the field. In Figure 10, the MFER, IP PER, and TS PER are shown similarly for the TU6 15 Hz channel at C/N = 15 dB, giving an average MFER closest to the area for acceptability in [5] that is, MFER 6.9–13.8%. Surprisingly, the MFER varies more than the TS PER and IP PER. Also, these measures are service specific, although measuring over the whole multiplex gives a fairly good approximation of the service specific TS PER and IP PER.

It is expected that the new objective criteria from the transmission system point of view should be designed as follows.

(i) The five above mentioned parameters should be used for studying service specific error characteristics at TS level. The values for the parameters should be derived from currently used channel models for mobile broadcasting, such as PO, VU, MR, and TU6.



FIGURE 9: Average error burst length at TS level for the VU channel and TU6 15 Hz channels at C/N = 16 dB and in the field in a vehicular urban use case.



FIGURE 10: MFER, IP PER, and TS PER for TU6 15 Hz channel at C/N = 15 dB for 16-QAM 1/2 3/4.

(ii) The effect of the MPE-FEC code rate in different channels should be studied to understand the error behavior at IP level.

(iii) The parameters should be mapped to results from future subjective test described in Section 7.3.

It was concluded in [25] that both VU and TU6 15 Hz are good choices for channel models, when modeling the vehicular use case in an urban environment. If designing subjective test cases based on the laboratory measurements in [25], the *C*/*N* values of 14 dB and 15 dB in the VU or TU6 15 Hz channels could be good starting points. The error statistics at TS level with IP PER and MFER are given in Table 6. Based on the results from [5], *C*/*N* = 14 dB is expected to give unacceptable quality, and *C*/*N* = 15 dB is expected to give acceptable quality with similar contents as in [5]. *C*/*N* points with similar error ratios in the PO channel should also be tested.

## 7.2. The impact of the decoders

One of the challenges in the task of specifying error criteria is the fact that the same transmission error may be concealed differently by audio and video decoder implementations. In a conservative approach, the simplest error-robust audio and video decoder implementations are considered. It can

Channel model	VU	VU	TU6 15 Hz	TU6 15 Hz
<i>C/N</i> [dB]	14	15	14	15
TS PER	16.69%	9.18%	15.41%	8.31%
AEBL	23.8	18.0	19.4	18.2
VEBL	2722	1409	1943	1395
MTBE	118.5	178.5	106.2	200.4
VTBE	77827	174824	69045	209196
IP PER	11.53%	3.18%	9.33%	2.71%
MFER	27.96%	9.43%	22.68%	7.35%

TABLE 6: Error statistics for VU and TU6 15 Hz channels at C/N 14 dB and 15 dB.

be assumed that these error-robust decoders do not crash or halt under any error conditions and are able to receive information on lost packets or detect lost data themselves. The simplest error-robust audio decoder replaces missing audio frames with silent frames. The simplest error-robust video decoder replaces missing or corrupted pictures with the previous correct decoded picture in presentation order. Furthermore, the simplest error-robust video decoder is capable of detecting whether errors occurred in nonreference or reference pictures. If an error occurred only in nonreference pictures, decoding continues from the next correctly received coded picture. If an error occurred in a reference picture, decoding continues from the next correctly received IDR picture.

Error criteria specified according the simplest errorrobust decoders above might produce too conservative results for sophisticated decoder implementations, which may be able to conceal errors successfully. For example, an audio frame may be successfully interpolated from temporally adjacent audio frames if those frames are well correlated. However, concluding whether error concealment operates sufficiently well is a challenging problem. One approach is to include auxiliary error concealment information into the audiovisual streams indicating the most efficient error concealment methods and the quality they are able to obtain. For example, the spare picture supplemental enhancement information message of H.264/AVC indicates which colocated areas in the indicated set of pictures are essentially unchanged so that any of those decoded areas can be used for concealing the corresponding area in an erroneously received coded picture.

#### 7.3. Subjective tests

The average length and amount of errors presented in Figures 7 and 8 can be divided into groups, where the difference between points for acceptable and unacceptable quality is clearly distinguishable. Finding the limits for acceptability by means of subjective testing and understanding what kind of transmission errors cause such error behavior is in the focus, when designing new objective error criteria. Also, understanding the length and frequency of errors on the perceived quality is necessary, when translating the subjective quality measures into objective numerical measures. This will require subjective tests similar to those performed in [5], where the impact of the average amount of errors and average error length are studied.

For consistency, the error information used in the subjective test should be based on error statistics or traces from current mobile radio channel models, as described in Section 7.1. The choice of content and audio and video coding parameters also play significant roles. The encoding and relation between audio and video will represent typical DVB-H service parameters. Probable IP level bitrates with current network parameters in DVB-H are between 300 Kbps and 768 Kbps, corresponding to about 25 and 10 services, respectively, with 16-QAM 1/2 3/4. These correspond to capability classes B and C in Table 4. The quality of the encoded video should be rated acceptable, preferably without visible errors.

The bitrates and contents should be divided into different groups. For example testing the four content types in [5] news, animation, music video, and sports with three different bitrates, for example, 300 Kbps, 500 Kbps, and 700 Kbps, give us 12 different test streams. The content types used in [5] correspond well to findings from user tests on mobile TV content [27]. Applying two C/N points for the VU channel and two for the PO channel gives four different error traces. Further, the error streams and the test clips should be matched in different ways so that the errors occur in different parts of the content. In [5], three different ways of matching each test clip and error trace were tested. Alternatively, the error traces could be chosen so that they represent different services in plots such as Figures 9 and 10, including the service corresponding to the maximum, average, and minimum error ratio. If the subjective ratings for these three cases are similar, we can use the service with the average error ratio to represent the whole multiplex.

After encoding and matching the test clips with the error traces, before running the subjective tests, it should be ensured that the test cases represent different points in similar plots as in Figures 7 and 8. Video clips with different bitrates should be treated separately.

#### 8. CONCLUSIONS

Currently, there are no error criteria for mobile broadcasting of streaming audiovisual services that would express all characteristics important for system design and have a verified correlation to subjective perceived quality. Known transmission system error criteria and objective video quality criteria were studied, and the results were compared to results on subjectively audiovisual quality.

In order to find measures for new error criteria to overcome the presented issues, we analyzed the characteristics from the perspective of the transmission system and video codec. We suggest that quality criteria based on average amount and average duration of errors should be defined based on subjective tests of audiovisual content. The error statistics used in the subjective test cases should be derived from conventional mobile radio channel models.

Here, DVB-H, H.264/AVC, and HE AAC v2 were used as an example system and codecs. However, we believe that the approach can be generalized to other systems and codec designs. Designing new transmission error criteria would be beneficial for developing further understanding of the constraints and degrees of freedom of wireless communication systems for all players in the field.

## ACKNOWLEDGMENT

The authors would like to thank Satu Jumisko-Pyykkö and Vinod Kumar Malamal Vadakital at Tampere University of Technology for providing the error information of the test cases used in [5].

#### REFERENCES

- [1] ETSI EN 302 304, "Digital video broadcasting (DVB); transmission system for handheld terminals (DVB-H)," European Telecommunication Standard, November 2004.
- [2] ISO/IEC 14496-3:2005/Amd 2, "Audio Lossless Coding (ALS), new audio profiles and BSAC extensions," 2006.
- [3] ITU Rec. H.264/ISO IEC 14996-10 AVC, "Advanced video coding for generic audiovisual services," 2003.
- [4] B. Girod and N. Färber, "Wireless video," in *Compressed Video over Networks*, chapter 12, CRC Press, Boca Raton, Fla, USA, 2000.
- [5] S. Jumisko-Pyykkö, V. K. Malamal Vadakital, and J. Korhonen, "Unacceptability of instantaneous errors in mobile television: from annoying audio to video," in *Proceedings of the* 8th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06), pp. 1–8, Helsinki, Finland, September 2006.
- [6] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.
- [7] ISO/IEC 13818-1, "Information technology—generic coding of moving pictures and associated audio information: systems," second edition, December 2000.
- [8] DVB-H Implementation Guidelines, v.1.3.1, DVB Document A092 Rev.2, May 2007.
- [9] P. Hakala and H. Himmanen, "Evaluation of DVB-H broadcast systems using new radio channel models," Tech. Rep. 817, Turku Center for Computer Science, Turku, Finland, 2007, http://www.tucs.fi/.
- [10] A. R. Reibman, S. Kanumuri, V. A. Vaishampayan, and P. C. Cosman, "Visibility of individual packet losses in MPEG-2 video," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 1, pp. 171–174, Singapore, October 2004.
- [11] A. R. Reibman, V. A. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 327–334, 2004.
- [12] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341– 355, 2006.
- [13] S. Kanumuri, S. G. Subramanian, P. C. Cosman, and A. R. Reibman, "Predicting H.264 packet loss visibility using a generalized linear model," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 2245–2248, Atlanta, Ga, USA, October 2006.
- [14] H. Joki and J. Paavola, "A novel algorithm for decapsulation and decoding of DVB-H link layer forward error correction," in *Proceedings of the IEEE International Conference on Commu-*

nications (ICC '06), vol. 11, pp. 5283–5288, Istanbul, Turkey, June 2006.

- [15] J. Paavola, H. Himmanen, T. Jokela, J. Poikonen, and V. Ipatov, "The performance analysis of MPE-FEC decoding methods at the DVB-H link layer for efficient IP packet retrieval," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, part 2, pp. 263– 275, 2007.
- [16] Wing TV project (EUREKA/Celtic) deliverable D11, "Wing TV Network Issues," August 2006, http://projects.celticinitiative.org/WING-TV/.
- [17] H. Himmanen, T. Jokela, J. Paavola, and V. Ipatov, "Performance analyses of the DVB-H link layer forward error correction," in *Handbook on Mobile Broadcasting*, CRC Press, Boca Raton, Fla, USA, 2008.
- [18] ETSI TS 102 005, "Digital video broadcasting (DVB); specification for the use of video and audio coding in DVB service delivered directly over IP protocols," ETSI Technical Specification, April 2006.
- [19] Nokia H.264 encoder, ftp://standards.polycom.com/IMTC\_ Media\_Coding\_AG/.
- [20] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, 2004.
- [21] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," in *Human Vision and Electronic Imaging IX*, vol. 5292 of *Proceedings of SPIE*, pp. 182–193, San Jose, Calif, USA, January 2004.
- [22] V. K. Malamal Vadakital, M. M. Hannuksela, M. Rezaei, and M. Gabbouj, "Method for unequal error protection in DVB-H for mobile television," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.
- [23] 3GPP TS 26.234 v6.5.0, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and Codecs (Release 6)," September 2005.
- [24] S. Jumisko-Pyykkö, V. K. Malamal Vadakital, M. Liinasuo, and M. M. Hannuksela, "Acceptance of audiovisual quality in erroneous television sequences over a DVB-H channel," in *Proceedings of the 2nd International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM '06)*, pp. 1–5, Scottsdale, Ariz, USA, January 2006.
- [25] H. Himmanen, "Studies on channel models and channel characteristics for mobile broadcasting," in *Proceedings of* the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–9, Las Vegas, Nev, USA, March-April 2008.
- [26] J. Poikonen, "Geometric run length packet channel models applied in DVB-H simulations," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.
- [27] D. Schuurman, P. Veevaete, and L. De Marez, "Mobile TV: killer content for the mobile generation," in *Proceedings of the 5th International Conference on Communication and Mass Media*, Athens, Greece, May 2007.

## Research Article

## A Method to Estimate the Horizontal Handover Decision Effect on Indoor Wireless Conversational Video Quality

## Alfonso Fernandez Duran,<sup>1</sup> Raquel Perez Leal,<sup>2</sup> and Jose I. Alonso<sup>2</sup>

<sup>1</sup> Alcatel-Lucent Spain, Ramirez de Prado 5, 28045 Madrid, Spain

<sup>2</sup> Escuela Tecnica Superior de Ingenieros de Telecomunicacion, Universidad Politecnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

Correspondence should be addressed to Alfonso Fernandez Duran, alfonso.fernandez\_duran@alcatel-lucent.es

Received 1 October 2007; Revised 31 January 2008; Accepted 4 April 2008

Recommended by Jianfei Cai

One of the most interesting and valuable services considered in fixed mobile convergence is video telephony. The success of this conversational video service will depend on the conversational video quality achieved in the multicell wireless indoor scenarios. One of the essential elements in the quality is the effect of the horizontal handovers in the conversational video. This paper analyzes the handover decision based on the probability calculation of handover events in the case of relative signal strength with hysteresis threshold (RSSHT) approach, and it proposes a new handover decision mechanism, variable hysteresis, to avoid unnecessary handovers. The paper presents the impact of the number of handovers and their duration time on the video's effective frame rate. Moreover, the effect of video stream modification during a short handover is also analyzed. Probability and handover duration approaches are combined and a new simple method for video quality evaluation is caused by the handovers in multicell indoor WLAN scenarios. Finally, the model proposed has been applied to a real office scenario.

Copyright © 2008 Alfonso Fernandez Duran et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

In the current context of fixed-mobile convergence, WLAN technology based on 802.11 is becoming available in common portable and mobile user terminals. This has brought about the possibility of using WLAN technology in conversational applications. Although IEEE802.11 was originally intended to transport best-effort data traffic, the incorporation of new standards like IEEE802.11e has brought about the opportunity of deploying delay and bandwidth sensitive services, like real-time voice and video communications. In these circumstances, WLANs combined with IP are being used as technology for limited mobility and nomadic services. The success of this scenario will depend on maintaining the communication's continuity through networks with several wireless access points (APs) by means of horizontal handover. WLANs were not initially designed to support handover between access points, based on the fact that users will most probably remain within the networks in a rather stationary way, using nonreal-time services. The normal use of a WLAN typically supports nonreal-time

handovers between APs, provided that the users have access rights to the destination network. The handover consists of an association to a wireless AP once the network client enters the coverage area of the destination AP. The horizontal handover is therefore a break before make process. User terminals usually incorporate just one WLAN transceiver, but in the case where two transceivers could be used, [1] describes a mechanism to manage handovers based on the voice over IP packet transmission retries. Horizontal handover is addressed in IEEE802.11r, [2]. The first dealing with authorization issues between different networks, and the second with increasing speed in the handover between access points.

The horizontal handover process could be split into two steps. The first one is to decide whether a handover is necessary and then select the destination AP. The second covers the layer 2 and layer 3 processes. The first step could take place in parallel with the communication without affecting it, while the second takes time from the service being conveyed. A description of the message transactions and time requirements is described in detail in [3]. According to [1] layer 2 could take between 50 milliseconds and 400 milliseconds, while layer 3, depending on the network settings could take 300 milliseconds or more.

Several studies have analyzed the handover decision process in cellular communications. Studies of the propagation parameters and criteria followed in handover decisions in cellular networks could be found in [4–7]. These studies are mainly related to outdoor mobile communications. Other studies characterize the performance of the handover in WLANs as based on measurements like [8], that characterize the timings and data transfers between network elements, and [9] measure the effects of handovers in voice communications. The impact of the horizontal and vertical handovers in voice communications is studied in [10] using the E model from ITU G.107.

Regarding video services, as a consequence of the evolution of the technologies and applications, advanced coding techniques have been introduced as a video coding format. In this study, ITU-T Rec. H.264 |ISO/IEC 14496-10 and H.264 have been considered. H.264 "represents an evolution of the existing video coding standards (H.261, H.262, and H.263) and it was developed in response to the growing need for higher compression of moving pictures for various applications such as videoconferencing, digital storage media, television broadcasting, Internet streaming, and communication" [11].

The H.264 defines a limited subset of syntax called "profiles" and "levels" in order to facilitate video data interchange between different applications. A "profile" specifies a set of coding tools or algorithms that can be used in generating a conforming bitstream, whereas a "level" imposes constraints on certain key parameters of the bitstream. The recommendation defines seven profiles (Baseline, Extended, Main and four High-profile types) and fifteen "levels" per "profile." The same set of "levels" is defined for all "profiles."

Current studies show that handovers have an impact on the quality of communications, since handovers produce discontinuities in the communication data streams. A discussion on video, the packet sizes and the implications in the PSNR are presented in [12]. The results shown are based only on simulations, and no model is proposed to predict the system performance. A qualitative study of using an intelligent access point handover mechanism in WLAN to obtain user perception for video conferencing quality, before and after applying the intelligent access point handover mechanism is presented in [13]. This reference shows that some strategies could improve the handover performance when the APs are congested, however no models are proposed to predict the effect on the video communications performances. In the case of conversational video, it is possible to obtain a simple video quality estimator using the effective frame rate resulting from the packet losses due to the handover effect, as introduced in reference [14].

Therefore, it is very appropriate to analyze the impact of handover on video communications quality in order to better understand the process involved. Moreover, it is also necessary to improve the method of planning real handover scenarios while maintaining the communication quality. The conversational video degradation due to the handover processes taking place in a wireless network can be addressed from at least two general perspectives: video processing to minimize whatever effect is taking place in the transmission media, and wireless processing carried out in the wireless part. With regard to the radio part, a new decision handover mechanism has been proposed (called variable hysteresis) to reduce unnecessary handovers. Moreover, in the video part, video stream modification has been introduced to minimize the handover duration impact arising from the interdependency of video frames. The following sections introduce the framework for conversational video application handover, a new and simple method of conversational video quality estimation based on handover time duration is also proposed. Finally, quality evaluation is shown in a real office scenario and planning recommendation provided.

## 2. WLAN HANDOVER PRINCIPLES

WLANs belonging to the IEEE802.11 family were not originally conceived to support a fast handover between access points. This has become a drawback when deploying multi-AP networks that convey real-time conversational services like IP-based video and voice telephony. The issue comes from the fact that the time necessary to associate it to a new AP is neither controlled nor limited to shorttime intervals. The association time could be of several hundreds of milliseconds as shown in [8, 15], while quality of communications such as voice could be severely affected by handover times of more than 50 milliseconds [10]. In addition, the availability of resources at the destination AP is not known until the handover has taken place. To solve these issues, the IEEE802.11r workgroup was set up to define a protocol to enable a fast and reliable handover between access points. By means of this new protocol, the mobile node (MN) can establish security and QoS status before taking a transition decision.

IEEE 802.11u is another IEEE task group that was set up to allow devices to interconnect with external networks, as typically found in hotspots. The main goal of this task group is to produce an amendment to the IEEE 802.11 standard to allow a common approach to interconnecting IEEE 802.11 access networks to external networks in a generic and standardized manner. The main particular issues covered are network selection, emergency call support, authorization from subscriber network, and media independent handover support.

## 2.1. Signal strength

To analyze the handover process in WLANs, it is necessary to understand the behavior of the received signal strength. In complex propagation scenarios, such as indoors, small changes in spatial separation between wireless access points and observation points impact causing dramatic changes in the signal amplitude and phase. In typical cellular communications systems, the signal strength analysis is based on the long distance outdoor or combined scenarios that experience Rayleigh fading. Several handover studies assume that the fading can be averaged to make up a random variable following a lognormal distribution as described in, [4, 5, 16, 17] in the following form:

$$f_i(\hat{s}) = \frac{1}{\hat{s}\sigma_i \sqrt{2\pi}} e^{-(\hat{s}-\mu_i')^2/2\sigma_i^2},$$
 (1)

where  $\hat{s}$  is the received signal amplitude of the envelope,  $\mu'_i$  are the average signal losses received at the mobile node from the wireless access point *i*, and could be expressed as

$$\mu'_{i} = k_{1} + k_{2} \log \left( d_{i} \right), \tag{2}$$

where  $d_i$  represents the distances from the observation point to the wireless access point *i*, AP<sub>i</sub>. Constants  $k_1$  and  $k_2$  represent frequency dependent and fixed attenuation factors, and the propagation constant, respectively. Finally,  $\sigma_i$ represents the shadowing that could be reasonably averaged to express slow power variations.

Although (1) represents the fading probability distribution function for the path losses as described in (2), in complex scenarios, such as indoors, in which many obstacles make up the propagation losses, the signal strength could be expressed as

$$\mu_i = P_{\text{tx}} - \left(k_1 + \sum_k \lambda_k + k_2 \log\left(d_i\right)\right),\tag{3}$$

where  $\lambda_k$  is the attenuation of the *k* passing through walls in the path from the observation point to the wireless access point, and  $P_{tx}$  the transmitted power. Moreover, to take the attenuation into account due to different floors in indoor propagation, one additional term could be added to (3) as stated in [18].

Other possible choices of statistical distributions for modeling the envelope have been described and detailed studies, based on exhaustive measurements, have been carried out to characterize indoor propagation. The Weibull distribution appears to be one of the statistical models that best describes the fading amplitude and fading power indoor scenarios, [18–21], improving the lognormal distribution in many cases. The Weibull distribution could be expressed as

$$f_i(\hat{s}) = ba^{-b}\hat{s}^{b-1}e^{-(\hat{s}/a)b}I_{(0,\infty)}(\hat{s}), \tag{4}$$

where  $\hat{s}$  represents the fading amplitude envelope or the fading power, and *a* and *b* are the position and shape values of the distribution, respectively. When the power distribution is represented in dBm, the extreme value distribution function should be used. In fact, if  $\hat{s}$  has a Weibull distribution with parameters *a* and *b*, log( $\hat{s}$ ) has an extreme value distribution with parameters  $\mu = \log(a)$  and  $\sigma = 1/b$  as shown in [22, 23].

The extreme value function for the power probability distribution function (pdf) has been seen as a good approximation. An example of fitting is shown in Figure 1. As can be seen, the power histogram of an indoor trajectory, modeled by the lognormal pdf function, is sufficiently represented by the extreme value function.

This behavior has already been observed in scenarios with complex propagation conditions such as vegetation obstacles [24].



FIGURE 1: Comparison of lognormal and extreme value pdf fit for an indoor trajectory power log.



FIGURE 2: Comparison of lognormal and extreme-value cumulative distribution approximation for an indoor trajectory power log.

Since most of the analysis will be probabilistic, it is interesting to see how the histogram in Figure 1 is approximated in terms of cumulative distribution function (cdf). The comparison results are shown in Figure 2. Integrating the differences between the sample data and the cdf approximations, an overall error of 2% can be seen for the case of lognormal function, and 5.5% for the case of extreme value function. Although lognormal is better overall in this scenario, local analysis shows that the maximum difference between sample data and lognormal cdf is 0.24, while the maximum difference for the extreme value is 0.16. This allows us to consider extreme value as a reasonable approximation. The use of this approximation will allow us to derive analytical expressions for the handover factors involved.

The extreme value probability distribution function is commonly used in the modeling and analysis of phenomena with low occurrence probabilities, as in risk analysis or the study of meteorology.

The pdf of the extreme value distribution can be expressed as

$$f_i(\hat{s}) = \frac{1}{\sigma_i} e^{(\hat{s}-\mu_i)/\sigma_i} e^{-e^{(\hat{s}-\mu_i)/\sigma_i}},\tag{5}$$

where  $\sigma_i$  and  $\mu_i$  are as defined above.

To analyze the strategies for the handover decision in outdoor cellular communications in the case of microcells and macrocells, [4, 6, 16] base their analysis on the estimation of the probability of unnecessary handovers from the statistical power distributions. For example, in the case of two base stations, what would be the probability of handover from base station 1 to 2 and then from 2 back to 1. In the case of indoor WLAN deployments, the conditions of simultaneous coverage of several access points, in which unnecessary handovers could take place between several of them, for instance from AP1 to AP2, from AP2 to AP3, and the back to AP1 are very frequent. The estimation of handover efficiency based on unnecessary handovers then becomes very complicated, as unnecessary handovers are very difficult to distinguish from necessary ones using the transition logic.

An alternative way of analyzing the handovers in outdoor communications is by means of the residence time, that is, once a handover has taken place, how long the mobile node stays in the new base station. In [25], the residence time statistics for handovers are analyzed. In the case of an indoor WLAN, the size of the cells also makes it very difficult to distinguish handovers based on the residence time, since a normal walking speed could produce a relatively high number of valid handovers with a relatively low residence time.

#### 2.2. Handover decision techniques

As an alternative to the analysis carried out for cellular communications, the present study proposes a simple analysis of the handover probability founded on the metrics used to make the handover decision in indoor WLAN communications. It is assumed that the scenario has been properly planned to provide sufficient coverage (minimum signal strength guaranteed at least in one of the access points). The strategy providing a lower probability of handover will present an overall better performance from the communication quality point of view. Based on these principles, several techniques could be used to implement the handover decision. In the following sections, these techniques are presented in order to introduce progressively the mathematical expressions to be used later in the method proposed.

## 2.2.1. Best server handover

If a mobile node sees the power levels  $s_i$  from each access points AP<sub>i</sub> within the coverage range, if no other criteria are implemented, it will associate itself to the one with higher average power. Assuming that the mobile node is currently at AP<sub>0</sub>, it will hand over to another AP<sub>i</sub> if there is one such as  $s_i > s_0$ . In this case, the mobile node will carry out handovers on best server basis.

The probability of a handover taking place will be

$$P_{\rm HO} = \sum_{i=1}^{n} P_i(s_i > s_0).$$
 (6)

Since the cumulative distribution function of (5) has a closed form

$$F(\hat{s}) = 1 - e^{-e^{(s-\mu)/\sigma}}.$$
(7)

Therefore,

$$\operatorname{Prob}(\hat{s} > s_0) = 1 - \operatorname{Prob}(s < s_0). \tag{8}$$

Consequently,

$$\operatorname{Prob}(\hat{s} > s_0) = 1 - \left(1 - e^{-e^{(s_0 - \mu)/\sigma}}\right) = e^{-e^{(s_0 - \mu)/\sigma}}.$$
 (9)

Combining (6) and (9), the probability of having a handover will be given by

$$P_{\rm HO} = \sum_{i=1}^{n} e^{-e^{(s_0 - \mu_i)/\sigma_i}},$$
 (10)

where the mobile node is associated to  $AP_0$  and simultaneously receives a signal strength above the sensitivity threshold from *n* access points  $\{AP_1, AP_2, \dots AP_n\}$ .

The main advantage of the best server approach is that it always keeps the mobile node associated to the AP giving the best signal quality, and therefore the higher bandwidth. On the other hand, the main disadvantage is that the number of handovers taking place during a mobile node moving trajectory could be very high, and therefore communications quality issues could take place, especially, in real-time conversational services.

Practical handover decision approaches usually require a reduction in the number of handovers taking place, while keeping the signal strength as high as possible.

## 2.2.2. Handover with fixed Hysteresis

A common approach to reduce the total number of handovers is to use a fixed hysteresis. The mechanism consists of carrying out a handover to a new AP, when the signal strength has improved over certain h value. If the current AP has a signal strength  $s_0$ , the handover to the new AP<sub>i</sub> will happen if  $s_i > s_0 + h$ .

Figure 3 shows the basic behavior of the hysteresis handover. While the best server approach will produce a handover decision at point A, hysteresis approach will produce the decision to handover to the destination AP at



FIGURE 3: Signal strength and handover with hysteresis.

point *B*. This approach absorbs any potential unnecessary handover originated with improvements lower than h on the signal quality.

Following the same approach as in the case of the best server, the probability of having a handover will be

$$P_{\rm HO} = \sum_{i=1}^{n} P_i(s_i > s_0 + h).$$
(11)

This could be expressed in a closed form as

$$P_{\rm HO} = \sum_{i=1}^{n} e^{-e^{(s_0 + h - \mu_i)/\sigma_i}}.$$
 (12)

In the particular case that same statistics are assumed for the n + 1 access points in a network, and replacing  $s_0$  with the expected value ( $\mu$ ), h could be expressed in a closed form as

$$h = \sigma \ln\left(\ln\left(\frac{n}{P}\right)\right). \tag{13}$$

As can be seen, h depends on the number of access points (n), the probability of having a handover and the standard deviation of the signal strength distribution, but it does not depend on the average signal strength  $(\mu)$ . From (13), it appears that h depends mainly on the standard deviation of the signal strength, with influence from the number of access points. Notice that the reduction in the probability of having a handover requires an increase of h. Similarly, h has to increase in networks with a larger number of access points.

The plain hysteresis approach described has the implicit drawback of producing a certain amount of unnecessary handovers, when the signal strength from the current access point is high, and the ratio  $s_i > s_0 + h$  is still possible.

#### 2.2.3. RSSHT handover

An improvement over the plain hysteresis handover decision approach is used to reduce the unnecessary handovers, when the signal strength of the current access point is high enough. The improved technique is known as relative signal strength with hysteresis and threshold (RSSHT). This technique is described in [8], and it basically allows a handover of the type  $s_i > s_0 + h$  when  $s_0 > T_{CS}$ , being  $T_{CS}$  the signal threshold. The probability of handover in this case could be expressed as

$$P_{\rm HO} = \begin{cases} \sum_{i=1}^{n} P_i(s_i > s_0 + h) & \forall s_0 \le T_{CS}, \\ 0 & \forall s_0 > T_{CS}, \end{cases}$$
(14)

or in a closed form:

$$P_{\rm HO} = \begin{cases} \sum_{i=1}^{n} e^{-e^{(s_0 - \mu_i)/\sigma_i}} & \forall s_0 \le T_{CS}, \\ 0 & \forall s_0 > T_{CS}. \end{cases}$$
(15)

The selection of the hysteresis margin h could be carried out in the same way as in the plain hysteresis approach.

Using RSSHT, it is possible to reduce the total number of handovers. The probability of having a handover for a given signal strength in the current AP will be influenced by the proper selection of the  $T_{CS}$  value. A possibility of selecting a  $T_{CS}$  value could be to estimate the signal strength for a given probability of crossing  $T_{CS}$  for a given access point. For example, assuming the same power distribution in all the access points, an estimation of the threshold could be

$$T_{CS} = \mu + \sigma \ln \left( -\ln \left( 1 - P_{CS} \right) \right), \tag{16}$$

where  $P_{CS}$  (probability of crossing the threshold) could be taken as the same value used for the probability of handover in the estimation of *h* in (13).

Since a reduction in the number of handovers is possible using RSSHT, this technique is commonly used in commercial products. Nevertheless, there is still a remaining part of handovers that are produced near  $T_{CS}$  and are also unnecessary.

#### 2.2.4. Handover with variable hysteresis

To increase the performance of conversational video in multicell wireless networks in the presence of handovers, we propose an improvement based on the use of hysteresis techniques with a variable margin. In such a way that when signal strength is high, the probability of a handover taking place is reduced (increase in the h value), and when signal strength is lower, the probability of handover increases (decrease in the h value). This approach will minimize the effect of having unnecessary handovers near  $T_{CS}$  using RSSHT.

To obtain a variable hysteresis margin, let us define a lower signal strength reference, called  $s_T$ , this value could be as low as the sensitivity value, but in principle it could be an arbitrary value. If the variable hysteresis margin is defined as



FIGURE 4: Variable Hysteresis behavior.

 $s_0 - s_T$ , where  $s_0$  is the current AP<sub>0</sub> signal strength. Assuming that AP<sub>i</sub> is the handover destination mobile node candidate and  $s_i$  is the power level received from it; handover only will take place if  $s_i > 2s_0 - s_T$ , that is, the higher the signal strength, the higher the hysteresis margin, and therefore the lower the probability of a handover taking place will be.

The behavior of the variable hysteresis is compared to the plain hysteresis in Figure 4, where it is noticeable that the handovers that took place at point C and D are no longer necessary, since the variable hysteresis line does not cross the power level of the target AP. This effect is possible since the minimum signal strength of the current AP is rather high.

In these conditions, the probability of a handover taking place will be

$$P_{\rm HO} = \sum_{i=1}^{n} P_i (s_i > s_0 + s_0 - s_T) = \sum_{i=1}^{n} P_i (s_i > 2s_0 - s_T).$$
(17)

This could be expressed in a closed form as

$$P_{\rm HO} = \sum_{i=1}^{n} e^{-e^{(2s_0 - s_T - \mu_i)/\sigma_i}}.$$
 (18)

Just as in the case of plain hysteresis, assuming that all signal distributions are similar, the lower reference signal strength could be estimated as

$$s_T = \mu - \sigma \ln\left(\ln\left(\frac{n}{P}\right)\right).$$
 (19)

The lower bound value is also a function of the expected value  $\mu$ , therefore the selection of this value could depend on how the actual network has been deployed, and could be taken as  $s_T = \mu - h$ , using *h* as in the plain hysteresis or RSSHT cases.

All of the handover techniques described could produce an abnormal behavior in conditions, where none of the access points available are received with a minimum amount of signal strength. These are typical conditions when outage conditions are produced.

#### 2.3. Signal outage conditions

A critical situation will occur when a mobile node such as a handheld device exits, the WLAN, for instance moving outside the coverage area. In this case, the way to keep the communication continuity is whenever possible to carry out a vertical handover to a cellular network.

If  $s_S$  is the signal level threshold to have acceptable communication in the WLAN, the probability that  $AP_i$  is received at the mobile node with a signal below this level will be

$$P(s_i < s_S) = F(s_S) = 1 - e^{-e^{(s_S - \mu_i)/\sigma_i}}.$$
 (20)

The condition for outage will be met, when all access points are below the threshold value. This condition could be expressed as

$$P_{\rm VH} = \prod_{i=0}^{n} P(s_i < s_S), \qquad (21)$$

or in a closed form:

$$P_{\rm VH} = \prod_{i=0}^{n} \left( 1 - e^{-e^{(s_{\rm S}-\mu_i)/\sigma_i}} \right).$$
(22)

In these conditions, the mobile node should have decided a vertical handover with certain anticipation. Descriptions of several techniques used for anticipating the vertical handover decision are available in [26]. Proper WLAN deployment designs should maintain the probability of experiencing outages and therefore vertical handovers in the coverage area in low values.

## 2.4. Comparison of handover decision approaches

The different handover approaches described could be compared by making some simplifying assumptions, and evaluating the probability of experiencing a handover for a given current access point signal strength value. Provided that the minimum acceptable signal strength is achievable, the lower the probability of having a handover is, the better the performance is, and therefore the better the associated approach.

Assuming that all access points present the same signal strength statistical behavior, and being consistent with the results of Figures 1 and 2 with  $\mu = -70 \text{ dBm}$ ,  $\sigma = 10$ , and using h = 10 dB and  $s_T = -92 \text{ dBm}$  for a scenario of four access points, the results for the different approaches are shown in Figure 5.

As can be seen, the probability of a handover taking place for a given signal strength is higher for the case of the best server approach, while minimum for the case of variable hysteresis. The difference between RSSHT and hysteresis appears within the lower handover probability range, caused by the threshold, and in the others both curves are identical and appear to overlap. In these conditions, it performs better than RSSHT, the latter better than hysteresis and hysteresis performs better than the best server. The differences are maintained in the whole range of the power values.

Taking the same values, the probability of signal outage becomes very low  $(2.6 \cdot 10^{-4})$ .



FIGURE 5: Comparison of the different handover approaches.

#### 3. CONVERSATIONAL VIDEO PERFORMANCE

A usual approach to estimating video quality is the peak signal-to-noise ratio (PSNR), or more recently video quality rating (VRQ) both are usually estimated from the mean square error (MSE) of the video frames after the impairments (e.g., packet loss) with respect to the original video frames [27, 28]. From these values, there is some correlation to the video mean opinion score (MOS), unfortunately, the relationship between packet loss and MSE is not straightforward, since not all packets conveyed through the wireless network have the same significance. Alternatively, a relatively simpler quality indicator is proposed in [14]. This indicator is the effective frame rate, which is introduced and discussed in later sections of this paper. This paper also proposes a model to characterize the impact of packet losses on the effective frame rate of the video sequence.

As packet losses occur in the wireless network, video frames are damaged; making some of them unusable, and therefore the total frame rate is reduced. Video quality will be acceptable, if the expected frame rate of the video conversations is kept above certain value.

The consequence of a packet loss in a generic video sequence depends on the particular location of the erroneous packet in the compressed video sequence. The reason for this is related to how compressed video is transmitted through the IP protocol. The plain video source frames are compressed to form a new sequence of compressed video frames or slices. The new sequence could be, depending on the H.264 service profile applied, made up of three types of frames: I (Intra) that transports the content of a complete frame with lower compression ratio, P (Predictive) that transports basic information on an prediction of the next frame based on movement estimators, and *B* (Bidirectional) that transports the difference between the preceding and the following frame. This sequence of slices is grouped into the so-called group of pictures (GoPs) or groups of video (GoV) objects depending on the standard. The GoV could adopt many forms and structures, but for our analysis, we assume a typical configuration of the form IPBBPBBPBBPBBPBB. This means that every 16 frames there is an Intra followed

by Predictive and Bidirectional frames. IP video packets are built from pieces of the aforementioned frame types and delivered to the network. If a packet error has been produced in a packet belonging to an Intra frame, the result is different from the same error produced in a packet belonging to a Predictive or Bidirectional frame.

There are some characteristics that are applicable to the case of conversational video, and in particular to portable conversational video, that are not necessarily applicable to other video services like IPTV or video streaming. The first important characteristic is the low-speed and low-resolution formats (common intermediate format, CIF, or quarter CIF, QCIF), that in turn produce a very low number of packets per frame, especially, if protocol efficiency is taken into account by increasing the average packet size. In these conditions, a single packet could convey a substantial part of a video frame. The second important characteristic comes from the portability and low consumption requirement at the receiving end that in turn requires a lighter processing load to save battery life. The combination of the two aforementioned characteristics makes packet losses impact greatly on the frame integrity and concealment becomes very restrictive. In conversational video, it could be better for instance to maintain a clear fixed image of the other speaker on the screen, than to try error compensation at the risk of severe image distortions and artifacts. Following these characteristics, every time that a packet is lost in a frame, the complete frame becomes unusable, and some actions could be taken at the decoder end to mitigate the effect, such as freezing or copying frames, but the effective frame rate has been reduced and has to accept some form of video quality degradation.

It is possible to obtain a simple video quality estimator based on the effective frame rate resulting from the packet losses due to the handover effect, as introduced in [14]. Although this could not be generalized for all types of IP video, in the case of conversational video, this indicator presents advantages over the use of PSNR: allows simple relationship between the packet loss and objective quality, and intuitively represents the behavior of conversational video over a wireless network.

#### 3.1. Handover impact on video quality

In the present analysis, it is considered that the handover duration  $T_s$  is longer than a video frame. Typical handover durations can be found in [15]. When the handover duration affects several slices, the video stream can only be displayed once an I frame is received, as shown in Figure 6. If  $T_s$  is the GoV and  $T_h$  is the handover duration, two cases are possible when  $T_h < T_s$  and the handover event involves slices from two GoV or when  $T_h > T_s$ . For the rest of the analysis, it is considered that  $T_h < T_s$ , (e.g.,  $T_h < 1000$  milliseconds) and therefore the number of frames lost are always less than two GoVs.

Following this principle, it is possible to estimate the expected number of frames lost due to a handover. Provided that several frames are lost in a handover, the video stream will need to wait until the next I slice, and therefore the



FIGURE 6: Handover effect on displayed video stream.

expected number of frames lost will depend in the position of the handover initiation in the GoV, which can be expressed as

$$E_1 = \frac{1}{n_s} \sum_{i=0}^{n_s-1} (n_s - i).$$
(23)

Depending on the handover duration, it is possible that the handover takes place at the end of one GoV, taking the first slices of the next, which makes GoV unusable. This effect produces an additional expected number of lost frames, which could be expressed as

$$E_2 = \frac{1}{n_s} \sum_{i=1}^{n_h} n_s,$$
 (24)

where  $n_h$  is the handover duration expressed in number of slices, that could be calculated from

$$n_h = \operatorname{ceil}(T_h f_0), \tag{25}$$

where  $T_h$  is the handover duration, and  $f_0$  is the video frame rate.

The expected number of frames lost in the video sequence due to a handover will be

$$E = E_1 + E_2 = \frac{n_s}{2} + n_h - 1.$$
 (26)

The performance of video in terms of frame rate in the presence of handovers could be expressed as follows:

$$f = f_0 (1 - E \cdot P_{\rm HO}).$$
 (27)

The performance will depend on the handover technique used. If the RSSHT technique is used, the resulting frame rate obtained bycombining (15), (26), and (27) will be

$$f = f_0 \left( 1 - \left( \frac{n_s}{2} + n_h - 1 \right) \sum_{i=1}^n e^{-e^{(s_0 - \mu_i)/\sigma_i}} \right) \quad \forall s_0 \le T_{CS}.$$
(28)

If variable hysteresis is used, by combining (18), (26), and (27) the resulting frame rate will come from

$$f = f_0 \left( 1 - \left( \frac{n_s}{2} + n_h - 1 \right) \sum_{i=1}^n e^{-e^{(2s_0 - s_T - \mu_i)/\sigma_i}} \right).$$
(29)

The previous approach is also applicable to the case of signal outage conditions, that is, none of the access points is received above the sensitivity threshold. In this case, (27) becomes

$$f = f_0 (1 - E \cdot (P_{\rm HO} + P_{\rm VH})). \tag{30}$$

Since signal outage is produced when no handover is possible, in the case of signal outage, (30) becomes

$$f = f_0 \left( 1 - \left( \frac{n_s}{2} + n_h - 1 \right) \prod_{i=0}^n \left( 1 - e^{-e^{(s_s - \mu_i)/\sigma_i}} \right) \right).$$
(31)

Equation (31) is valid for the cases in which the signal outage duration is longer than one video frame, which is the typical case. This effect is mainly related to the radio network design and could have very low impact.

In Figures 7 and 8, the case of four cells is shown. According to the signal levels, there is certain probability of experiencing a handover, and thus undergoing a reduction in the frame rate. The example covers several handover durations. If the acceptability limit is considered to be 5 frames/second [29], quality outage can be evaluated as a function of the signal strength coming from the radio network design.

It must be noted that in Figure 7 outage starts at -76 dBm, while in Figure 8 it appears at -79 dBm, which is a 3 dB improvement.

## 3.2. Solution to improve handover impact on video quality

The conversational video degradation due to the handover processes taking place in a wireless network can be addressed



FIGURE 7: Frame rate as a function of the signal quality according to handover probability in multicell networks in the case of RSSHT.



FIGURE 8: Frame rate as a function of the signal quality according to handover probability in multicell networks in the case of variable hysteresis.

from at least two general perspectives: video processing to mitigate whatever effect is taking place in the transmission media, For example, degradations due to channel, handovers, call drops, and so forth, and wireless processing carried out in the wireless part. In the previous section, the variable hysteresis handover technique has been proposed in order to minimize the handover probability, in this section the solution proposed focuses on the video part.

In fact, a technique that could improve the video quality caused by handover will be to reduce the GoV objects lost due to frame interdependence. This will be possible by forcing a reset of the GoV to generate an I frame, breaking off the natural sequence of frames IPBBPBBPBBPBBPBB and reducing the time between I frames. An example is shown in Figure 9.

There are, at least, two possible techniques to implement the proposed reset. First, to take advantage of channel reciprocity for conversational video, which means that downand uplinks channel behaviors are closely related, so the transmitter end can detect a handover event automatically as well as generating the I frame automatically when required. Another technique will be to introduce a feedback from receiver to transmitter through a real-time control protocol.

In these conditions, the new expected value of frames lost in the process will consist of at least one frame lost to detect the event and an additional frame to reset the GoV. On top of these 2 frames, the total duration of the handover will need to be translated in a frame count.

Taking as example the RSSHT case, (28) will become

$$f = f_0 \left( 1 - (n_h + 2) \sum_{i=1}^n e^{-e^{(s_0 - \mu_i)/\sigma_i}} \right) \quad \forall s_0 \le T_{CS}.$$
(32)

As a result, there will be a considerable statistical improvement in the case of short handovers, while this improvement is smaller in the typical case of longer handovers; in any case, it also depends on the relative distance between the handover end event and the next Intra-frame. Figure 10 compares the different approaches discussed in terms of expected frame rate as a function of the average signal level received at the AP. The expected frame rate represents the effective frame rate resulting from the packet losses due to the handover effect. Handover duration is 1000 milliseconds in Figure 10. The shape of the curves is strongly influenced by the handover technique used. Variable hysteresis produces softer shape, while RSSHT produces a somehow abrupt behavior. As can be seen, the greater improvement is achieved with the variable hysteresis approach although some additional improvement is also possible combining both variable hysteresis and Intra reset. Most of the gain is produced by the use of variable hysteresis, as will be confirmed by simulations.

#### 4. SCENARIO SIMULATION

To illustrate the principles shown in the above sections, an example has been selected as shown in Figure 11. The scenario corresponds to an office environment of  $20.2 \times 35.5$  meters covered with four access points represented with the color shapes inside the layout, each color associated to a frequency channel. The walls have been modeled to introduce attenuation in the signal propagation and edge diffraction. The position of the access points has been found with an automatic process to optimize the coverage and capacity simultaneously. The dashed line represents a trajectory of a mobile node at 4 Km/h in the scenario, covering a total distance of 190 m.

The handover performance could be assessed for the different conditions described in the previous sections. The simulations consist of the use of RSSHT and of variable hysteresis techniques with different handover durations through the trajectory and comparing the results. The power histograms of the four access points through the trajectory are shown in Figure 12. Several iterations are also carried out to analyze the influence and sensitivity of the design parameters on the final performance.



FIGURE 9: Handover effect on displayed video stream with Intra reset.



FIGURE 10: Frame rate performance for 1 second handovers using the different techniques.



FIGURE 11: Simulation scenario.

## 4.1. RSSHT handover parameter simulation

The use of a threshold above, where no handover action takes place, produces some improvement in the total number of handover events, even without using hysteresis, this is due to the simple fact that some of the best server handovers are taking place at relatively high signal strengths. Figure 13 represents handover events as a function of time to move through the trajectory by the mobile node at a pedestrian speed. As shown in Figure 13, the total number of handovers for the 3 dB margin is reduced from 25 to 17, and with 7 dB the reduction is from 17 to 12. An additional margin increases to 16 dB, produces a further reduction in the total handover events down to 6, but at the expense of having the lowest signal strength below the system sensitivity. Just as in the case of plain hysteresis, in this case the distribution of handover events also depends on h, and the increase in h is stripping handovers from the registry, and slightly changing their position.

The impact of the hysteresis margin on the number of handover events using threshold is shown in Figure 14. The behavior is monotonic down to 19 dB where 4 handovers are produced; an additional increase in the margin produces no improvement in the handover count.

Although the number of handovers decrease monotonically as the hysteresis margin increases, it is essential to analyze the impact on the minimum signal level. The effect of the hysteresis margin on the minimum signal strength value is shown in Figure 15. The use of a threshold has less influence on the minimum signal strength observed. Hysteresis greater than 9 dB produces minimum signal strength in the range of the system sensitivity, therefore a value lower than 9 dB is the best choice for the h value. Let



FIGURE 12: Power histograms for the four APs through the trajectory.



FIGURE 13: Handover events using RSSHT.

us take 7 dB, since it is producing a minimum signal in the range of -86 dBm.

Summarizing, the *h* value has to be as small as possible to keep the minimum signal level as high as possible, but at the same time *h* has to be as large as possible to reduce the total handover count. The tradeoff selected has been h = 7 dB.

#### 4.2. Variable hysteresis simulation

One of the advantages of the variable hysteresis techniques is that there is no need for parameter tuning. The only parameter is the lower boundary that can be the system sensitivity threshold that is usually -92 dBm for typical IEEE802.11 WLAN.

Figure 16 shows the results obtained using variable hysteresis in the reference scenario. As can be seen, the number of handovers has been reduced from 12 (optimal RSSHT with hysteresis of 7 dB) to 9 keeping the same value of minimum signal strength in the trajectory.

In summary, variable hysteresis produces additional reduction in the unnecessary handover count with respect to the RSSHT.


FIGURE 14: Handover events as a function of the hysteresis margin using  $T_{CS} = -80$  dBm.



FIGURE 15: Minimum signal strength as a function of the hysteresis margin for  $T_{CS} = -75$  dBm.

#### 5. SCENARIO PERFORMANCE RESULTS

Taking the RSSHT handover technique with the parameter setting from the previous sections, it is possible to analyze the impact of the handovers in the conversational video quality.

The approach followed has been to simulate the video performance for different handover durations in the scenario shown in Figure 11.

Assuming that the total duration of the trajectory is 171 seconds, corresponding to a relative low moving speed, the average frame rate of the conversational video obtained in different handover durations is shown in Figure 17.

The fluctuations of the frame rate are due to the frames lost in the handover process. Depending of the duration of the handovers and the frequency of the occurrence, the frame rate can be reduced down to values that are not acceptable for the user. If the generally accepted value of 5 frames/second [29] is taken as figure of merit for video acceptability, it is possible to see how the frame rate evolves and where it will cross the limit. Figure 17 presents this circumstances starting at 134 seconds for all handover durations. This effect is due to the accumulation of handovers starting at this point of the trajectory, as can be seen in Figure 13, not allowing the video frame rate to recover.

If the frame rate drops to 5 frames/second or below, it is considered as producing an outage, and the outage ends once

Handover events. Variable hysteresis threshold



FIGURE 16: Handover using variable hysteresis.



FIGURE 17: Resulting frame rate as a function of the handover duration using RSSHT.

the frame rate raises again above 5 frames/second. Using this principle, the duration of the quality outages can be estimated.

By repeating the process for different handover durations, it is possible to assess the influence of the handover duration on the quality outages.

If GoV reset is applied, the results are significantly improved as shown in Figure 18.

Figure 19 shows the effect of the handover duration on the outage in the simulation scenario for normal RSSHT and RSSHT with Intra reset. As can be seen, the quality outage grows monotonically with the handover duration as expected.

In the present simulation, it is noticeable that handover durations longer than 600 milliseconds produce a relatively flat quality outage in RSSHT, this means that above 600 milliseconds the outage of 7.5% of time is obtained no matter how long the handover is (up to the simulated value of 1000 milliseconds). To explain this behavior, it is necessary to revisit Figure 13. It is clearly visible that an accumulation of handovers can be seen at the end of the trajectory. These handover events are due to a sequence of abrupt signal fluctuations received from two of the access points (AP3 and AP4) in the scenario. This effect is partially mitigated by using the GoV reset in the low range of the handover duration, as it can be expected.

Using the variable hysteresis technique, the average duration of the frame rate along the trajectory is shown in Figure 20. The frame rate values obtained are above 12



FIGURE 18: Resulting frame rate as a function of the handover duration using RSSHT with GoV reset.



FIGURE 19: Quality outage time as a function of the handover duration.

frames per second in all cases which is way above the acceptability limit. The average frame rate experiences rather low reductions thanks to the positions of the handover events that are rather spread along the trajectory

A final comparison has been performed simulating PSNR as indicator to validate the results obtained. To use PSNR as quality indicator, it is necessary to use a known video sequence and for that purposes the "Foreman" sequence has been used. For the simulations, the following parameters have been considered: QCIF test sequence "Foreman" (25 frames/second, 300 frames), mobile user at 4 Km/h, compressed bit rate at 32 Kbps and average handover duration of 1000 milliseconds. Since "Foreman" is shorter than the duration of the simulated trajectory, the video sequence has been concatenated several times. The technique used for received frame lost concealment is the frame copying, that is, an interval of lost frames is replaced by the last usable frame received.

The results obtained repeating the process for the different techniques, and averaging between handover events are shown in Figure 21. Average PSNR at 32 Kbps bit rate without handover is 31 dB [30] and a 25 dB PSNR acceptability limit is assumed. Figure 21 shows similar results as for the case of frame rate. In fact, the outage appears at



FIGURE 20: Resulting frame rate as a function of the handover duration using variable hysteresis.



FIGURE 21: Simulated PSNR using Foreman video sequence.

the end of the path with both RSSHT techniques, with and without Intra reset, however variable hysteresis solves this problem.

# 6. CONCLUSIONS AND FUTURE WORK

The effect of horizontal handover on the conversational video performance in multicell WLAN has been analyzed, and a simple method to predict the impact on the video performance in terms of frame rate has been proposed. The RSSHT handover decision approach, which is a commonly used technique, has been taken as reference for the video performance, and the parameter settings have been discussed. A new technique called variable hysteresis has been proposed and shows a substantial improvement in conversational video performance. A video transport improvement based on the GoV reset after a handover event has also been considered to enhance video quality. The proposed analysis method has been applied to a real scenario by simulating a range of handover durations. The simulations show how video frame rate fluctuates along the trajectory due to the handovers. The scenario has also shown how an accumulation of handovers could produce outages of video quality using RSSHT. The quality outage periods grow with the handover duration. The improvement coming from the video stream modification has been compared with those coming from variable hysteresis handover decision technique: the new technique proposed at this paper. The results show that improvements in the handover process reducing unnecessary handovers produces higher quality gain than the video transport processing.

Subsequent research steps will be, first, a more profound performance analysis and simulation of the variable hystereris handover decision mechanisms to further reduce the handover count in multicell deployments to increase the quality. Second, to introduce new video quality indicators that could also be applicable to video streaming and mobile TV.

#### ACKNOWLEDGMENTS

The authors are thankful for the support of the Spanish Ministry of Education and Science within the framework of the TEC2005-07010-C02-01/TCM and the TSI2005-07306-C02-01/CASERTEL-NGN projects chair in Madrid Polytechnic University. The authors are also thankful for the support of CELTIC Project Easy Wireless II.

# REFERENCES

- S. Kashihara and Y. Oie, "Handover management based on the number of retries for VoIP on WLANs," in *Proceedings of the* 61st IEEE Vehicular Technology Conference (VTC '05), vol. 4, pp. 2201–2206, Stockholm, Sweden, May-June 2005.
- [2] Draft Standard for Information Technology Telecommunications and informationexchange between systems—Local and metropolitan area networks—Specific requirements part 11: Wireless LAN Medium AccessControl (MAC) and Physical Layer (PHY) specifications—Amendment 2: Fast BSS Transition. Active Unapproved Draft.
- [3] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the IEEE 802.11 MAC layer handoff process," *Computer Communication Review*, vol. 33, no. 2, pp. 93–102, 2003.
- [4] M. Zonoozi and P. Dassanayake, "Handover delay and hysteresis margin in microcells and macrocells," in *Proceedings of the 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '97)*, vol. 2, pp. 396– 400, Helsinki, Finland, September 1997.
- [5] M. Gudmundson, "Analysis of handover algorithms [microcellular radio]," in *Proceedings of the 41st IEEE Vehicular Technology Conference (VTC '91)*, pp. 537–542, St. Louis, Mo, USA, May 1991.
- [6] A. Murase, I. Symington, and E. Green, "Handover criterion for macro and microcellular systems," in *Proceedings of the 41st IEEE Vehicular Technology Conference (VTC '91)*, pp. 524–530, St. Louis, Mo, USA, May 1991.
- [7] G. E. Corazza, D. Giancristofaro, and F. Santucci, "Characterization of handover initialization in cellular mobile radio networks," in *Proceedings of the 44th IEEE Vehicular Technology Conference (VTC '94)*, vol. 3, pp. 1869–1872, Stockholm, Sweden, June 1994.
- [8] F. González, J. A. Pérez, and V. H. Zárate, "HAMS: layer 2 handoff accurate measurement strategy in WLANs 802.11," in *Proceedings of the 1st International Workshop on Wireless Network Measurements (WiNMee '05)*, pp. 1–7, Trentino, Italy, April 2005.

- [9] I. Marsh and B. Grönvall, "Performance evaluation of voice handovers in real 802.11 networks," in *Proceedings of the* 2nd International Workshop on Wireless Network Measurements (WiNMee '06), Boston, Mass, USA, April 2006.
- [10] A. F. Duran, E. C. del Pliego, and J. I. Alonso, "Effects of handover on voice quality in wireless convergent networks," in *Proceedings of the IEEE Radio and Wireless Symposium* (*RWS* '07), pp. 23–26, Long Beach, Calif, USA, January 2007.
- [11] ITU-T Recommendation H.264, "Infrastructure of audiovisual services—Coding of moving video. Advanced video coding for generic audiovisual services," September 2005.
- [12] E. Masala, C. F. Chiasserini, M. Meo, and J. C. De Martin, "Real-time transmission of H.264 video over 802.11-based wireless ad hoc networks," in *Proceedings of the Workshop on DSP in Vehicular and Mobile Systems*, pp. 193–207, Springer, Nagoya, Japan, April 2003.
- [13] D. Singh, S. Hoh, A. L. Y. Low, F. L. Lim, S. L. Ng, and J. L. Tan, "Qualitative study of intelligent access point handover in WLAN systems," in *Proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium (NOMS '06)*, pp. 1– 4, Vancouver, BC, Canada, April 2006.
- [14] N. Feamster and H. Balakrishnan, "Packet loss recovery for streaming video," in *Proceedings of the 12th International Packet Video Workshop*, pp. 1–11, Pittsburgh, Pa, USA, April 2002.
- [15] J. Montavont, N. Montavont, and T. Noel, "Enhanced schemes for L2 handover in IEEE 802.11 networks and their evaluations," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 3, pp. 1429–1434, Berlin, Germany, September 2005.
- [16] P. Dassanayake, "Dynamic adjustment of propagation dependant parameters in handover algorithms," in *Proceedings of the* 44th IEEE Vehicular Technology Conference (VTC '94), vol. 1, pp. 73–76, Stockholm, Sweden, June 1994.
- [17] N. Zhang and J. M. Holtzman, "Analysis of handoff algorithms using both absolute and relative measurements," in *Proceedings of the 44th IEEE Vehicular Technology Conference* (VTC '94), vol. 1, pp. 82–86, Stockholm, Sweden, June 1994.
- [18] T. K. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, pp. 51–82, 2003.
- [19] H. Hashemi, M. McGuire, T. Vlasschaert, and D. Tholl, "Measurements and modeling of temporal variations of the indoor radio propagation channel," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 3, part 1-2, pp. 733–737, 1994.
- [20] F. Babich and G. Lombardi, "Statistical analysis and characterization of the indoor propagation channel," *IEEE Transactions* on Communications, vol. 48, no. 3, pp. 455–464, 2000.
- [21] M. H. Ismail and M. M. Matalgah, "On the use of padé approximation for performance evaluation of maximal ratio combining diversity over weibull fading channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, Article ID 58501, 7 pages, 2006.
- [22] C. Walck, "Handbook on statistical distributions for experimentalists," Internal Report SUF-PFY/96-01, University of Stockholm, Stockholm, Sweden, November 2000.
- [23] K. Krishnamoorthy, Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2006.
- [24] S. Perras and L. Bouchard, "Fading characteristics of RF signals due to foliage in frequency bands from 2 to 60 GHz,"

in Proceedings of the 5th International Symposium on Wireless Personal Multimedia Communications (WPMC '02), vol. 1, pp. 267–271, Honolulu, Hawaii, USA, October 2002.

- [25] S. Kourtis and R. Tafazolli, "Evaluation of handover related statistics and the applicability of mobility modelling in their prediction," in *Proceedings of the 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (*PIMRC '00*), vol. 1, pp. 665–670, London, UK, September 2000.
- [26] A. H. Zahran, B. Liang, and A. Saleh, "Signal threshold adaptation for vertical handoff in heterogeneous wireless networks," *Mobile Networks and Applications*, vol. 11, no. 4, pp. 625–640, 2006.
- [27] J. Hu, S. Choudhury, and J. D. Gibson, "PSNR. r,f: assessment of delivered AVC/H.264 video quality over 802.11a WLANs with multipath fading," in *Proceedings of the 1st Multimedia Communications Workshop (MULTICOMM '06)*, pp. 1–6, Istambul, Turkey, June 2006.
- [28] S. Winkler, Digital Video Quality Vision Models and Metrics, John Wiley & Sons, New York, NY, USA, 2005.
- [29] J. D. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '04)*, pp. 535–542, Vienna, Austria, April 2004.
- [30] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

# Research Article Protection of Video Packets over a Wireless Rayleigh Fading Link: FEC versus ARQ

# Julie Neckebroek, Frederik Vanhaverbeke, Danny De Vleeschauwer, and Marc Moeneclaey

Department of Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium

Correspondence should be addressed to Julie Neckebroek, julie.neckebroek@telin.ugent.be

Received 1 October 2007; Revised 25 March 2008; Accepted 8 May 2008

Recommended by David Bull

Video content can be provided to an end user by transmitting video data as a sequence of internet protocol (IP) packets over the network. When the network contains a wireless link, packet erasures occur because of occasional deep fades. In order to maintain a sufficient video quality at the end user, video packets must be protected against erasures by means of a suitable form of error control. In this contribution, we investigate two types of error control: (1) forward error correction (FEC), which involves the transmission of parity packets that enables recovery of a limited number of erased video packets, and (2) the use of an automatic repeat request (ARQ) protocol, where the receiver requests the retransmission of video packets that have been erased. We point out that FEC and ARQ considerably reduce the probability of unrecoverable packet loss, because both error control techniques provide a diversity gain, as compared to the case where no protection against erasures is applied. We derive a simple analytical expression for the diversity gain resulting from FEC or ARQ, in terms of the channel coherence time, the allowable latency, and (for FEC) the allowable overhead or (for ARQ) the time interval between (re)transmissions of copies of a same packet. In the case of HDTV transmission over a 60 GHz indoor wireless link, ARQ happens to outperform FEC.

Copyright © 2008 Julie Neckebroek et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. INTRODUCTION

The internet protocol (IP) allows the provision of a mix of multimedia services (video, audio, voice, data, gaming, etc.) to an end user, by breaking up the bitstreams generated by the various services into IP packets and sending these packets over the network. In this contribution, we consider the delivery of these multimedia services via a wireless channel, and focus on the reliability of the received video data.

The occurrence of fading on wireless channels makes reliable transmission a difficult task, because occasional deep fades give rise to bursts of bit errors at the receiver. IP packets affected by bit errors are erased at the receiver, yielding lost packets at the destination. These lost packets are likely to cause visual distortions when viewing the video content at the destination. Hence, in order to obtain a sufficient quality of experience (QoE) it is imperative to limit the video packet loss rate.

In addition, the frequency selectivity of the wireless channel distorts the transmitted signal. In order to cope with frequency selectivity, we resort to a multicarrier modulation (orthogonal frequency division multiplexing (OFDM)) [1], which turns the frequency-selective channel into a number of parallel frequency-flat channels.

In order to alleviate the damaging impact of fading, one can reduce the probability of bit errors by means of coding on the physical (PHY) layer. Not only the video, but also the other services that are provided via the same wireless link stand to benefit from this coding. In this contribution, we restrict our attention to orthogonal space-time block codes [2–4], for which the optimum decoding reduces to linear processing and simple symbol-by-symbol detection. When this PHY layer coding is not sufficient to yield a satisfactory QoE related to video, additional protection of the video packets must be envisaged.

In order to provide additional protection of the video packets against erasures, one can resort to forward error correction (FEC) coding [5, 6] or to automatic repeat request (ARQ) protocols [7, 8]; these techniques involve the transmission of redundant packets (in addition to the video information packets) or sending a request for retransmitting erased video packets, respectively. Various proposals have been formulated for protecting packets against erasures by means of FEC [9–12]; in this contribution we select reedsolomon (RS) codes, because they are able to recover the maximum possible number of erasures for a given transmission overhead [5, 13]. As far as ARQ protocols are concerned, we consider selective repeat (SR) ARQ, which yields the minimum transmission overhead [7, 8]. It is important to keep in mind, however, that these techniques come with a cost. First, both FEC and ARQ introduce transmission overhead (usually higher for FEC than for ARQ) and some latency. Second, there is a complexity increase: ARQ requires a retransmission buffer and a return channel from the

receiver to the retransmitting network node, and FEC needs

additional encoding/decoding operations. In this contribution, we investigate to what extent the combination of the RS code or the SR ARQ protocol with the space-time PHY layer code improves the reliability of the video transmission over a wireless channel subject to Rayleigh fading. The paper is organized as follows. In Section 2, we introduce some basic concepts about video compression and transmission over an IP network, and describe the space-time coding on the PHY layer. We detail in Section 3 the RS erasure coding and the SR ARQ protocol that are used as additional protection of the video packets against erasures. We provide in Section 4 the error performance analysis for various scenarios, involving spacetime coding or no coding on the PHY layer, with or without protection (RS coding or SR ARQ) of the video packets. In Section 5, we present numerical results, including a case study pertaining to HDTV transmission over a 60 GHz indoor wireless link. Finally, in Section 6 conclusions are drawn regarding system performance and complexity, and some generalizations of the considered assumptions are briefly discussed. A major conclusion is that RS erasure coding and SR ARQ yield the same maximum possible diversity gain, which is determined by the ratio of the allowed latency and the channel coherence time; however, this maximum cannot be achieved because of practical constraints on the allowed overhead (RS erasure coding) or when the time interval between retransmissions exceeds the channel coherence time (SR ARQ).

# 2. VIDEO SOURCE CODING AND TRANSMISSION

In this section, we describe the video packet transmission from the video server to the end user. First, the video source coding method is considered. Next, the different layers in the protocol stack of the OSI-model, that are relevant to this research, are presented.

# 2.1. Video source coding

The video stream is encoded (compressed) according to the MPEG-2 standard [14, 15], which is commonly used as the format for digital television. The Video section of MPEG-2 (part 2) is designed to compress the video stream through appropriate coding by exploiting the existing redundancy in space and time. Uncompressed video can be seen as a sequence of picture frames (e.g., 25 frames per second). Typically, the scenes in successive pictures are very similar. One can take advantage of this similarity to compress the video into three types of frames: intracoded frames (I-frames), predictive-coded frames (P-frames), and bidirectional-predictive-coded frames (B-frames).

An I-frame is a compressed version of a single uncompressed frame. The compression is achieved by exploiting the spatial redundancy in the image and the insensitivity of the human eye to certain changes in the image. P-frames, on the other hand, achieve a higher compression because they take advantage of the resemblence between the picture in the current frame and the picture in the previous I- or P-frame. B-frames are compressed by exploiting both the picture in the preceding I- or P-frame as well as the picture in the following I- or P-frame. These B-frames achieve an even higher compression rate. A commonly used frame pattern is IBBPBBPBBPBB, called a group of pictures (GOPs), which consists of 12 compressed frames and which is repeated. Such a GOP has a duration of 480 milliseconds (25 frames per second).

As the different types of frames achieve different compression rates, their resulting sizes, measured in bits, are not equal. I-frames are larger than P-frames, which in turn are larger than B-frames. Their exact sizes depend on the video content. Typically, the average sizes of I- and P-frames are about 6 and 2 times the average size of a B-frame.

Because of the interdependence of the compressed frames, error propagation occurs: an erroneous I- or P-frame results in errors (after decoding) in the 2 preceding B-frames and in all following frames up to (but not including) the next I-frame. Hence, when an I- or P-frame in a GOP is affected by unrecoverable transmission errors, a visual distortion is likely to occur when viewing the video content. Errors in a Bframe do not propagate to other frames. Hence, when only a B-frame in a GOP is affected by unrecoverable transmission errors, it is possible that no visual distortion occurs through the use of error concealment techniques that exploit the similarity between the erroneous B-frame and surrounding frames.

### 2.2. Protocol stack

Let us consider the case where video data is sent from the video server to the end user, as shown in Figure 1. A source, the video server, broadcasts the video data. Via an aggregation network, this video data reaches a digital subscriber line access multiplexer (DSLAM). The DSLAM sends the data related to a mix of services (video, audio, voice, data, gaming, etc.), over a digital subscriber line (DSL) [16] to the user home gateway (HG). From the HG, the video data is sent through a wireless LAN to the set top box (STB). Figure 1 also displays the different layers of the protocol stack, that are involved in the operation of each of the network nodes. The network nodes are not able to process information from other layers.

#### 2.3. Application layer

The system section of MPEG-2 (part 1) [15] describes how MPEG-compressed video and audio data streams



FIGURE 1: Concatenation of DSL connection and wireless connection (DSLAM = digital subscriber line access multiplexer, HG = home gateway, STB = set-top box).

(along with other data, such as teletext, elementary stream identifiers) are multiplexed together to form a single data stream. Basically, the resulting transport stream (TS) consists of a sequence of MPEG-TS packets, that consist of 188 bytes each (including a 4-byte header).

#### 2.4. Session layer

The real-time transport protocol (RTP) [17] is used to deliver audio and video over the Internet. The RTP packets are filled with an integer number of TS packets. In commercial equipment, an RTP packet typically contains 7 TS packets, which is the maximum number of TS packets that fits inside an Ethernet frame (data link layer). The header of an RTP packet contains, among other things, a sequence number and a time stamp. This allows the detection of missing or out-of-order delivery of RTP packets and to perform synchronization, respectively. The header inserted by this protocol is 12 bytes long.

### 2.5. Transport layer and network layer

The user datagram protocol (UDP) is used on the transport layer to deliver the RTP packets. UDP is well suited for time-sensitive applications that prefer dropped packets to excessively delayed packets.

The UDP packets are passed to the underlying layer, the network layer. This layer uses the IP protocol to deliver the data from source to destination.

#### 2.6. Data link layer

On the medium access control (MAC) sublayer of the data link layer, a header and trailer are added; the latter contains a cyclic redundancy check (CRC). This CRC allows the detection of packets that are corrupted by transmission errors; corrupted packets are not forwarded to the network layer, but are discarded ("erased"). We assume that no ARQ is applied on the MAC layer; the effect of ARQ on the MAC layer is briefly discussed in Section 6.

The structure of a data-link-layer packet is visualized in Figure 2. The packet contains 7 MPEG-TS packets, and the



FIGURE 2: The video data is nested in a structure of packets, each packet and corresponding header results from a different layer in the protocol stack.

various headers/trailers that have been added by the different layers in the protocol stack.

# 2.7. Physical layer

As far as the physical (PHY) layer is concerned, we only consider the wireless link between the HG and the STB. On the PHY layer of the HG transmitter, the L bits to be sent for every data-link-layer packet are mapped onto an Mpoint signal constellation. The resulting M-ary data symbols are transmitted at a rate  $R_s$  (in symbols per second) over the wireless channel; hence the duration of a packet equals  $L/(R_s \log_2(M))$ . The transmission makes use of orthogonal frequency-division multiplexing (OFDM) [1]. The sequence of data symbols at rate  $R_s$  is demultiplexed into  $N_c$  parallel symbol streams, each of rate  $R_s/N_c$ . These  $N_c$  symbol streams are modulated onto  $N_c$  distinct subcarriers, that have a frequency separation of (slightly more than)  $R_s/N_c$ , and the sum of these modulated subcarriers is transmitted. The transmitted signal can be viewed as a sequence of OFDM blocks. As shown in Figure 3, an OFDM block has a duration of  $N_c/R_s$ , and contains  $N_c$  data symbols (i.e., one symbol on each of the  $N_c$  subcarriers). The bandwidth occupied by the resulting transmitted signal is (slightly more than)  $R_s$ . The transmission of an *L*-bit packet involves  $L/(N_c \log_2(M))$ OFDM blocks. Typically, the number  $N_c$  of carriers is on the order of 100 to 1000. Because of the large number of subcarriers, OFDM turns the wireless fading channel into a set of  $N_c$  flat-fading parallel channels.

For each subcarrier, the fading gain is assumed to be piecewise constant over time; the fading gain does not change over a time interval equal to the channel coherence time  $T_{\rm coh}$ , and is statistically independent of the fading gain in other intervals of duration  $T_{\rm coh}$ . During an interval  $T_{\rm coh}$ , several packets are transmitted, as indicated in Figure 4. Packets from other applications are located in between the packets with video data.

On the PHY layer of the STB receiver, the M-ary data symbols are detected, and demapped to bits. On the MAC sublayer, the recovered bits are grouped into packets of size L, and error detection based on the CRC is performed. When an error is detected, the packet is erased; otherwise, the packet is passed to the higher layers.

Because of fading, the received signal is occasionally strongly attenuated. To alleviate the damaging impact of fading on the detection of the *M*-ary data symbols, we consider the use of multiple transmit and receive antennas. A multiple-input multiple-output (MIMO) system with  $N_t$ transmit and  $N_r$  receive antennas allows the introduction



interval 2*i*: 
$$s_{2i}(t)$$
 (on antenna 1)  
 $s_{2i+1}(t)$  (on antenna 2),  
interval 2*i* + 1:  $-(s_{2i+1}(t))^*$  (on antenna 1)  
 $(s_{2i}(t))^*$  (on antenna 2),  
(1)

where ()\* denotes complex conjugate. Hence, each OFDM block  $s_n(t)$  reaches the receiver via  $2N_r$  wireless links.

#### **ADDITIONAL PROTECTION OF THE VIDEO DATA** 3.

As mentioned before, packets yielding an erroneous checksum are discarded (erased) on the MAC layer, because they have been affected by transmission errors; the other packets are assumed to be received correctly. Because of video packet erasures, visual distortions may occur when viewing the received video content. In order to guarantee a sufficient QoE to the end user, the rate of video packet erasures should be limited. When the packet erasure rate caused by transmission errors on the wireless link is too large, additional measures are needed to recover erased video packets. In this contribution, we consider the combination of a PHY layer with either no coding or Alamouti space-time coding with 1 or 2 receive antennas, and additional packet protection by means of either RS erasure coding or SR ARQ.

# 3.1. RS erasure coding

The RS code is defined over the Galois field  $GF(2^q)$ , which implies that an RS code symbol consists of q bits; typically, q = 8. (The RS code symbols are not to be confused with the transmitted data symbols; the former belong to  $GF(2^q)$ , whereas the latter belong to an *M*-point signal constellation.) In the sequel, a video information packet refers to the MPEG-TS payload (i.e., 7 MPEG-TS packets) of the packet as shown in Figure 2. Per group of K of these video information packets, we transmit N - K parity packets. This results in a packet codeword of N packets. The parity packets are constructed such that taking from each packet the *i*th block of *q* bits yields an RS(*N*, *K*) codeword, for all i = 1, 2, ..., L/q. This construction is illustrated in Figure 5. Hence, when e packets from the packet codeword are erased, each of the L/qRS codewords is affected by exactly *e* symbol erasures.

The RS(N,K) code is known to be maximum distance separable (MDS), that is, the code can recover up to N - Kerasures, which cannot be outperformed by any other code with the same number N - K of parity symbols (Note that a receiver without an RS decoder can still process the packet stream by simply ignoring the parity packets, at the expense of a performance degradation as compared to a receiver with an RS decoder.) [5, 13]. When the number of erasures is larger than N - K, erasure decoding fails and unrecoverable packet loss occurs.

The introduction of erasure coding yields an increase of both overhead and latency.

(i) Using an (N, K) block code gives rise to a transmission overhead ovh given by ovh = (N - K)/K,



Video packets

L bits

←  $\rightarrow$ 



FIGURE 4: Video packet stream and fading gain versus time; in this example, 2 video packets are transmitted during the channel coherence time, in which case a packet group consists of 2 packets.

of space-time coding [2-4]. Whereas an uncoded singleinput single-output (SISO) system, that is,  $N_t = N_r =$ 1, provides only one wireless link between the HG and the STB, the number of wireless links provided by an orthogonal space-time block-coded (OSTBC) MIMO system equals  $N_r N_t$ . As compared to an SISO system, the larger number of links resulting from OSTBC MIMO gives rise to a considerably higher robustness against fading, and a much better error performance. Using an OSTBC MIMO system does not require additional bandwidth as compared to the SISO system, but comes at a substantial hardware cost that increases with the number of antennas. The spacetime coding only marginally increases the latency. Optimum decoding of OSTBC MIMO reduces to linear processing and simple symbol-by-symbol detection at the receiver.

In this paper, we will consider the Alamouti spacetime code [2], which requires 2 transmit antennas (and an arbitrary number  $N_r$  of receive antennas). Denoting by  $s_n(t)$  the signal that corresponds to the *n*th OFDM block, Alamouti space-time coding involves the transmission of two OFDM blocks during two consecutive intervals (each of



because for each *K* information packets, N - K additional packets must be transmitted. Hence, denoting by  $R_{\text{pack}}$  (in packets per second) the rate of information packets, the packet transmission rate equals  $(N/K)R_{\text{pack}}$ . This indicates that because of the coding the fraction of time during which the channel is used for video transmission is increased by a factor N/K, leaving less room for the transmission of packets from other applications.

(ii) When at most N - K packets are erased, they can be recovered by means of the RS(N,K) code. To perform erasure decoding, at least K packets must be received correctly. Hence, the RS decoder might need to wait until all N packets of the codeword are received, before the erasure decoding can start. Hence, using the (N,K) block code introduces a maximum additional latency  $T_{lat}$  which equals the duration  $K/R_{pack}$  of a packet codeword. Increasing the latency gives rise to a larger zapping delay, which might unfavorably affect the user's QoE. (The zapping delay is the time that elapses between giving the command to change the TV channel and the appearance of the new TV channel on the screen [18].)

Considering the above, the code parameters *N* and *K* should be selected such that the overhead and latency are limited to reasonable values.

It is convenient that the parity packets are generated by the video server, as this is the only network node (besides the STB of the end user) that has access to the video data. In principle, parity packets could instead be generated by the DSLAM or the HG. However, this would require that the DSLAM or the HG has access to the higher protocol layers (beyond IP), which would increase their complexity and cost.

#### 3.2. Selective repeat ARQ

As far as ARQ is concerned, we consider an SR retransmission protocol. The STB receiver sends a retransmission request for each of the erased video packets, and only copies of the erased packets are retransmitted. To limit the round-trip delay, we assume that retransmissions occur from either the DSLAM or the HG. Of course, the functionality of the retransmitting network node needs to be extended beyond the IP layer, in order to be capable of recognizing retransmission requests related to specific video packets; in addition, this node must have a retransmission buffer containing video packets that have not yet been correctly received. Augmenting the functionality of the DSLAM or HG increases their complexity and cost. As the HG is a consumer product, the DSLAM appears to be the economically justified choice for operating as the retransmitting node. However, the HG offers the shorter round-trip delay.

Upon receiving a retransmission request, the retransmitting network node sends a copy of the packet involved. Retransmissions are scheduled such that the time interval  $T_{\text{retr}}$  between the (re)transmission instants of copies of the same packet is not less than the channel coherence time  $T_{\rm coh}$ . This way, the different copies experience statistically independent fading. When one would select  $T_{\rm retr} < T_{\rm coh}$ , the retransmission of a packet that has been erased because of a deep fade is experiencing the same deep fade, and therefore is likely to be erased as well. Such retransmissions should be avoided, as they are not useful, but rather contribute to the transmission overhead.

The minimum possible time interval  $T_{\text{retr, min}}$  between (re)transmission instants of the same packet is the sum of the packet duration  $L/(R_s \log_2(M))$  and the round-trip delay  $T_{\text{RT}}$ ; the latter is the sum of the two-way propagation delay, the duration of the acknowledgment message, and the processing delays at the receiver and the transmitter [7, 8]. We select  $T_{\text{retr}} = \max(T_{\text{retr, min}}, T_{\text{coh}})$ . When  $T_{\text{retr, min}} > T_{\text{coh}}$ , this yields  $T_{\text{retr}} = T_{\text{retr, min}}$ : the interval between transmission instants is the shortest possible, and (re)transmitted copies of the same packet experience-independent fading. When  $T_{\text{retr, min}} \le T_{\text{coh}}$ , we get  $T_{\text{retr}} = T_{\text{coh}}$ : the retransmission instant is deliberately delayed by an amount ( $T_{\text{coh}} - T_{\text{retr, min}}$ ) with respect to the earliest possible retransmission instant, in order that the (re)transmitted copies of the same packet are affected by independent fading gains.

Since each retransmission gives rise to a latency of  $T_{\text{retr}}$ , the maximum number  $N_{\text{retr}}$  of allowed retransmissions per packet is given by  $N_{\text{retr}} = \lfloor T_{\text{lat}}/T_{\text{retr}} \rfloor$ , in order that the total latency caused by the SR ARQ protocol does not exceed  $T_{\text{lat}}$ .

#### 4. SYSTEM ANALYSIS

In this section, we present the analysis of the system under study. We first investigate the PHY layer, followed by the additional packet protection by means of RS erasure coding or SR ARQ. As a performance measure, we consider the average number of GOPs that are affected by irrecoverable packet loss, over a reference time interval of 12 hours. Finally, analytical results regarding RS erasure coding and SR ARQ are compared.

### 4.1. PHY layer

We consider the cases of uncoded SISO transmission, and Alamouti orthogonal space-time coding (2 transmit antennas) with 1 or 2 receive antennas. The probability  $P_{bit}(x)$ , that a bit is received in error, depends on the instantaneous channel state *x*. The channel state *x* is the sum of the squared fading gains that are involved in the transmission of the considered bit (1 fading gain for SISO, and 2 or 4 fading gains for Alamouti with 1 or 2 receive antennas). Limiting our attention to QPSK transmission,  $P_{bit}(x)$  is given by [2, 6]

$$P_{\text{bit}}(x) = \begin{cases} Q\left(\sqrt{\frac{2E_b x}{N_0}}\right) & \text{uncoded SISO,} \\ Q\left(\sqrt{\frac{E_b x}{N_0}}\right) & \text{Alamouti,} \end{cases}$$
(2)

where

$$Q(\nu) = \frac{1}{\sqrt{2\pi}} \int_{\nu}^{+\infty} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u \tag{3}$$

is the complement of the cumulative distribution function of a zero-mean unit-variance Gaussian random variable. In (2),  $E_b$  denotes the transmitted energy per bit of the video packet, and  $N_0$  is the one-sided power spectral density of the noise at the receiver.  $P_{\text{bit}}(x)$  equals 1/2 for x = 0, and converges to 0 when  $x \rightarrow \infty$ ; the larger  $E_b/N_0$  is, the faster this convergence occurs. When the fading gains are normalized such that the average energy per bit at each receive antenna also equals  $E_b$ , the probability density function p(x) of the channel state is given by [6]

$$p(x) = \frac{x^{D-1} \exp(-x)}{(D-1)!},$$
(4)

with D = 1 for uncoded SISO and D = 2 or D = 4 for Alamouti with  $N_r = 1$  or  $N_r = 2$ . The quantity D is the *diversity* provided by the PHY layer; basically, D equals the number of physical links between the transmitter and the receiver that are exploited by the transmission scheme. As we will shortly demonstrate, the error performance improves with increasing D; this is intuitively clear, because all D links must fail for a packet erasure to occur.

From (2), the packet erasure probability  $P_{\text{pack}}(x)$  conditioned on *x* equals

$$P_{\text{pack}}(x) = 1 - (1 - P_{\text{bit}}(x))^{L}.$$
 (5)

To obtain (5), we have assumed that all  $N_c$  subcarriers of the OFDM signal experience the same value of the channel state x, and have taken into account that the packet duration is less than the channel coherence time, so that the channel state is the same for all L bits of a packet. The effect of relaxing this assumption is briefly discussed in Section 6. For x = 0,  $P_{\text{pack}}(x)$  and  $1 - P_{\text{pack}}(x)$  equal  $1 - 2^{-L}$  and  $2^{-L}$ , respectively. For  $x \to \infty$ ,  $P_{\text{pack}}(x)$  and  $1 - P_{\text{pack}}(x)$  converge to zero and to one, respectively; the speed of convergence increases with increasing  $E_b/N_0$ . Finally, note from (2) that  $P_{\text{bit}}(x)$  and  $P_{\text{pack}}(x)$  depend on x and  $E_b/N_0$  only through the variable  $y = xE_b/N_0$ .

Before we consider in the next subsections the cases where RS erasure coding or SR ARQ is used in order to recover erased packets, we now investigate the system performance under the assumption that no such error control measures are taken.

We define a *packet group* as the set of packets that are transmitted consecutively in time during an interval of duration  $T_{\rm coh}$  over which the fading is constant. We denote by  $N_{\rm coh}$  the number of packets transmitted during the interval  $T_{\rm coh}$ . For the example shown in Figure 4, we have  $N_{\rm coh} = 2$ . As we consider the case where only information packets and no parity packets are transmitted, we have  $N_{\rm coh} = \lceil T_{\rm coh} R_{\rm pack} \rceil$ . The probability  $P_{\rm group}(e)$  that *e* packets are erased within a packet group of size  $N_{\rm coh}$ , irrespective of the channel state, is given by

$$P_{\text{group}}(e) = \frac{N_{\text{coh}}!}{e! (N_{\text{coh}} - e)!} \times \int_{0}^{+\infty} P_{\text{pack}}^{e}(x) (1 - P_{\text{pack}}(x))^{N_{\text{coh}} - e} p(x) dx, \quad (6)$$
$$e = 0, \dots, N_{\text{coh}}.$$

Considering the behavior of  $1 - P_{\text{pack}}(x)$ ,  $P_{\text{group}}(0)$  converges to 1 for large  $E_b/N_0$ . For large  $E_b/N_0$  and e > 0,  $P_{\text{pack}}^e(x)$  goes to zero much faster than p(x) for increasing x, so that the factor  $\exp(-x)$  in (4) can be approximated as  $\exp(-x) \approx 1$ . Using the approximation in (6) along with the substitution

$$F\left(\frac{E_b x}{N_0}\right) = \frac{N_{\text{coh}}!}{e! \left(N_{\text{coh}} - e\right)!} P_{\text{pack}}^e(x) \left(1 - P_{\text{pack}}(x)\right)^{N_{\text{coh}} - e},$$
(7)

we obtain, for high  $E_b/N_0$ ,

$$P_{\text{group}}(e) \approx \int_{0}^{+\infty} F\left(\frac{E_{b} x}{N_{0}}\right) \frac{x^{D-1}}{(D-1)!} dx$$
$$= \left(\frac{E_{b}}{N_{0}}\right)^{-D} \int_{0}^{+\infty} F(y) \frac{y^{D-1}}{(D-1)!} dy, \quad e = 1, \dots, N_{\text{coh}}.$$
(8)

Taking into account that F(y) is not a function of  $E_b/N_0$ , we have  $P_{\text{group}}(e) \propto (E_b/N_0)^{-D}$  for e > 0.

Let us now compute the probability  $P_{\text{GOP}}$  that a GOP is affected by unrecoverable packet loss. As no measures are taken to recover erased packets, each erased packet is lost. Denoting by  $T_{\text{GOP}}$  and  $N_{\text{GOP}}$  the duration of one GOP and the number of packet groups that fit within the duration of one GOP, respectively, we have  $T_{\text{GOP}} = N_{\text{GOP}}N_{\text{coh}}/R_{\text{pack}}$ , and

$$P_{\text{GOP}} = 1 - \left(P_{\text{group}}(0)\right)^{N_{\text{GOP}}} \\ = 1 - \left(1 - \sum_{e=1}^{N_{\text{coh}}} P_{\text{group}}(e)\right)^{N_{\text{GOP}}} \\ = \sum_{i=1}^{N_{\text{GOP}}} \frac{N_{\text{GOP}}! (-1)^{i-1}}{i! (N_{\text{GOP}} - i)!} \left(\sum_{e=1}^{N_{\text{coh}}} P_{\text{group}}(e)\right)^{i}$$
(9)  
$$\approx N_{\text{GOP}} \sum_{e=1}^{N_{\text{coh}}} P_{\text{group}}(e) \\ = N_{\text{GOP}} (1 - P_{\text{group}}(0)).$$

The approximation in (9) corresponds to keeping only the term with i = 1, which is the dominating term at high  $E_b/N_0$ . Hence, for large  $E_b/N_0$ , we obtain  $P_{\text{GOP}} \propto (E_b/N_0)^{-D}$ . This illustrates the impact of the PHY layer diversity *D*: the larger *D*, the smaller the probability that a GOP is affected by packet erasures.

From (9), we compute the average number  $E[\#\text{GOP}_{unrec}]$  of GOPs that are affected by unrecoverable packet loss in a reference interval  $T_{ref}$  of 12 hours. Denoting by  $N_{ref}$  the number of GOP intervals in  $T_{ref}$ , we have  $T_{ref} = N_{ref}T_{GOP} = N_{ref}N_{GOP}T_{coh}$ . Hence,

$$E[\#\text{GOP}_{\text{unrec}}] = N_{\text{ref}} P_{\text{GOP}}$$

$$\approx N_{\text{ref}} N_{\text{GOP}} (1 - P_{\text{group}}(0)) \qquad (10)$$

$$= \frac{T_{\text{ref}}}{T_{\text{coh}}} (1 - P_{\text{group}}(0)).$$

The approximation in (10) holds for large  $E_b/N_0$ . Note that, at high  $E_b/N_0$ ,  $E[\#\text{GOP}_{unrec}]$  is independent of the GOP duration, and proportional to  $(E_b/N_0)^{-D}$ .

### 4.2. Packet protection by means of RS erasure coding

Now we consider the case where (N - K) parity packets are added to K information packets, yielding a (N, K) RS packet codeword. The number N<sub>coh</sub> of packets transmitted during the interval  $T_{\rm coh}$  is now given by  $N_{\rm coh} = \lceil (N/K)T_{\rm coh}R_{\rm pack} \rceil$ , which denotes the size of a packet group. We assume that the N packets of the packet codeword are distributed over  $N_{\text{group}}$ packet groups, to which we associate the indices 1, 2, ... and  $N_{\text{group}}$ . We denote by  $e_n$  the number of erased packets in the packet group with index n ( $n = 1, ..., N_{\text{group}}$ ), and introduce the vector  $\mathbf{e} = (e_1, \dots, e_{N_{\text{group}}})$ . We define by  $Pr(\mathbf{e})$ the probability that the number of erased packets in the groups with indices 1, 2, ... and  $N_{\text{group}}$  equals  $e_1, e_2, ...$  and  $e_{N_{\text{group}}}$ , respectively. Assume for simplicity that N is an integer multiple of  $N_{\rm coh}$  and that the first packet of the codeword is also the first packet of a packet group; in this case, we have  $N_{\text{group}} = N/N_{\text{coh}}$ , and each of the packet groups contains exactly  $N_{\rm coh}$  packets from the considered codeword. Taking into account that erasures in different packet groups are statistically independent, we obtain

$$\Pr(\mathbf{e}) = \prod_{n=1}^{N_{\text{group}}} P_{\text{group}}(e_n), \qquad (11)$$

where  $P_{\text{group}}(e)$  is given by (6), but with  $N_{\text{coh}} = [(N/K)T_{\text{coh}}R_{\text{pack}}]$ . When N is not an integer multiple of  $N_{\text{coh}}$  and/or the first packet of the codeword is not the first packet of a group, an edge effect occurs: we get  $N_{\text{group}} = [N/N_{\text{coh}}]$  or  $N_{\text{group}} = [N/N_{\text{coh}}] + 1$ , depending on the position of the first packet of the codeword within its packet group; for example, Figure 6 shows a situation with N = 5,  $N_{\text{coh}} = 3$ , and  $N_{\text{group}} = 3$ . Then (11) must be slightly modified by taking into account that the packet groups with indices 1 and  $N_{\text{group}}$  might contain fewer than  $N_{\text{coh}}$  packets from the considered codeword. Recalling that, for high  $E_b/N_0$ ,  $P_{\text{group}}(e) \propto (E_b/N_0)^{-D}$  for e > 0 and  $P_{\text{group}}(0) \approx 1$ ; it follows from (11) that  $\Pr(\mathbf{e}) \propto (E_b/N_0)^{-nD}$  with *n* denoting the number of nonzero entries of  $\mathbf{e}$ .

From (11), the probability  $P_{\text{RS}}(e_{\text{tot}})$  that  $e_{\text{tot}}$  erasures occur in the packet codeword is given by

$$P_{\rm RS}(e_{\rm tot}) = \sum_{e_1+e_2+\cdots+e_{N_{\rm group}}=e_{\rm tot}} \Pr(\mathbf{e}).$$
(12)

Finally, the probability Pr(decoding failure) that the erasures cannot be recovered by the RS decoder (because  $e_{\text{tot}}$  is larger than N - K) becomes

$$Pr[decoding failure] = \sum_{e_{tot}=N-K+1}^{N} P_{RS}(e_{tot})$$

$$= 1 - \sum_{e_{tot}=0}^{N-K} P_{RS}(e_{tot}).$$
(13)

In order to obtain at least (N - K + 1) erasures in the codeword, at least  $\gamma_{RS} = \lceil (N - K + 1)/N_{coh} \rceil$  packet groups must contain erased packets; this implies that the vectors  $\mathbf{e}$  in (12) must have at least  $\gamma_{RS}$  nonzero entries. Hence, for large  $E_b/N_0$ , Pr(decoding failure) is proportional to  $(E_b/N_0)^{-\gamma_{RS}D}$ . Taking into account that ovh = (N - K)/K,  $T_{lat} = K/R_{pack}$  and  $N_{coh} = \lceil (N/K)T_{coh}R_{pack} \rceil = \lceil NT_{coh}/T_{lat} \rceil \approx NT_{coh}/T_{lat}$ ,  $\gamma_{RS}$  can be expressed as

$$\gamma_{\rm RS} = \left\lceil \frac{N - K + 1}{N_{\rm coh}} \right\rceil \approx \left\lceil \frac{N - K}{N_{\rm coh}} \right\rceil \approx \left\lceil \frac{\text{ovh}}{1 + \text{ovh}} \cdot \frac{T_{\rm lat}}{T_{\rm coh}} \right\rceil.$$
(14)

Note that  $y_{RS}$  is an increasing function of both ovh and  $T_{lat}$ .

Now we consider the probability  $P_{\text{GOP}}$  that a GOP is affected by an unrecoverable packet loss. Denoting by  $N_{\text{RS}}$ the number of packet codewords in one GOP interval  $T_{\text{GOP}}$ , we have  $T_{\text{GOP}} = N_{\text{RS}}K/R_{\text{pack}}$ , and

$$P_{\text{GOP}} = 1 - (1 - \Pr[\text{decoding failure}])^{N_{\text{RS}}}$$
  

$$\approx N_{\text{RS}} \Pr[\text{decoding failure}].$$
(15)

Similary, the average number of GOPs that are affected by unrecoverable packet loss during a reference period  $T_{ref}$  of 12 hours is given by

$$E[\#\text{GOP}_{\text{unrec}}] = N_{\text{ref}}P_{\text{GOP}}$$

$$\approx N_{\text{ref}}N_{\text{RS}}Pr[\text{decoding failure}] \qquad (16)$$

$$= \frac{T_{\text{ref}}}{T_{\text{lat}}}Pr[\text{decoding failure}],$$

where  $T_{\text{ref}} = N_{\text{ref}}T_{\text{GOP}} = N_{\text{ref}}N_{\text{RS}}T_{\text{lat}}$ . The approximations in (15) and (16) are valid for large  $E_b/N_0$ . We deduce from (15) and (16) that both  $P_{\text{GOP}}$  and  $E[\#\text{GOP}_{\text{unrec}}]$  are proportional to  $(E_b/N_0)^{-\gamma_{\text{RS}}D}$ . Hence, as compared to the case where no erasure coding is used, the effect of the RS(N,K) code is to increase the diversity order from D to  $\gamma_{\text{RS}}D$ : erasure coding introduces a diversity gain of  $\gamma_{\text{RS}}$ . According to (14), a tradeoff exists between the achievable diversity gain and the allowable overhead and latency: the smaller the allowable overhead and latency, the smaller the achievable diversity gain.

# 4.3. Packet protection by means of selective repeat ARQ

With the proposed retransmission strategy, a packet will be lost definitively when it has been erased during the first transmission *and* during  $N_{\text{retr}}$  successive retransmissions. The probability  $P_{\text{ARO, unrec}}(\mathbf{x})$  of this event is given by

$$P_{\text{ARQ, unrec}}(\mathbf{x}) = \prod_{i=0}^{N_{\text{retr}}} P_{\text{pack}}(x_i), \qquad (17)$$

where  $P_{\text{pack}}(x)$  is the packet erasure probability corresponding to a channel state x (see (5)), and  $\mathbf{x} = (x_0, \dots, x_{N_{\text{retr,max}}})$ contains the values of the channel state at the first transmission and the subsequent  $N_{\text{retr}}$  retransmissions of the considered packet. The probability  $P_{\text{group, unrec}}(\mathbf{x})$  that at least



FIGURE 5: Construction of a packet codeword.

one packet from a packet group of  $N_{\text{coh}} = [T_{\text{coh}}R_{\text{pack}}]$  packets (which all experience the same channel state) is erased definitively is given by

$$P_{\text{group, unrec}}(\mathbf{x}) = 1 - (1 - P_{\text{ARQ, unrec}}(\mathbf{x}))^{N_{\text{coh}}}$$
$$= \sum_{j=1}^{N_{\text{coh}}} \frac{N_{\text{coh}}!}{j! (N_{\text{coh}} - j)!} (-1)^{j-1} P_{\text{ARQ, unrec}}^{j}(\mathbf{x}).$$
(18)

Averaging  $P_{\text{group, unrec}}(\mathbf{x})$  over the channel gain statistics yields the probability  $P_{\text{group, unrec}}$  that at least one packet in a packet group is definitively lost, irrespective of the channel state values:

$$P_{\text{group, unrec}} = \sum_{j=1}^{N_{\text{coh}}} \frac{N_{\text{coh}}!}{j! (N_{\text{coh}} - j)!} (-1)^{j-1} E[P_{\text{ARQ, eras}}^{j}(\mathbf{x})]$$
$$= \sum_{j=1}^{N_{\text{coh}}} \frac{N_{\text{coh}}!}{j! (N_{\text{coh}} - j)!} (-1)^{j-1} E\left[\prod_{i=0}^{N_{\text{retr}}} P_{\text{pack}}^{j}(x_{i})\right]$$
$$= \sum_{j=1}^{N_{\text{coh}}} \frac{N_{\text{coh}}!}{j! (N_{\text{coh}} - j)!} (-1)^{j-1} (E[P_{\text{pack}}^{j}(x)])^{N_{\text{retr}}+1}$$
(19)

with

$$E[P_{\text{pack}}^{j}(x)] = \int_{0}^{+\infty} P_{\text{pack}}^{j}(x)p(x)\mathrm{d}x \qquad (20)$$

and where p(x) is given by (4). For large  $E_b/N_0$ , we have  $E[P_{\text{pack}}^j(x)] \propto (E_b/N_0)^{-D}$ , so that  $P_{\text{group, unrec}}$  is proportional to  $(E_b/N_0)^{-(1+N_{\text{retr}})D}$ .

Following the same reasoning as in Section 4.1, the quantities  $P_{\text{GOP}}$  and  $E[\#\text{GOP}_{\text{unrec}}]$  are given by

$$P_{\text{GOP}} = 1 - (1 - P_{\text{group, unrec}})^{N_{\text{GOP}}}$$

$$\approx N_{\text{GOP}}P_{\text{group, unrec}},$$

$$E[\#\text{GOP}_{\text{unrec}}] = N_{\text{ref}}P_{\text{GOP}}$$

$$\approx N_{\text{ref}}N_{\text{GOP}}P_{\text{group, unrec}}$$

$$= \frac{T_{\text{ref}}}{T_{\text{coh}}}P_{\text{group, unrec}}.$$
(21)

For large  $E_b/N_0$ , both  $P_{\text{GOP}}$  and  $E[\#\text{GOP}_{\text{unrec}}]$  are proportional to  $(E_b/N_0)^{-(1+N_{\text{retr}})D}$ . Hence, as compared to the case of no retransmissions, the use of SR ARQ provides a diversity gain  $\gamma_{\text{ARQ}}$  which is given by  $\gamma_{\text{ARQ}} = 1 + N_{\text{retr}} = 1 + \lfloor T_{\text{lat}}/T_{\text{retr}} \rfloor$ .

Let us compute the average overhead E[ovh] related to the retransmission protocol. The average number E[#transm] of transmissions per packet is related to the average overhead by E[#transm] = 1 + E[ovh]. It is easily verified that

$$\Pr[\#\text{transm} = i] = \begin{cases} (1 - P_{\text{pack}})P_{\text{pack}}^{i-1} & i = 1, \dots, N_{\text{retr}}, \\ P_{\text{pack}}^{N_{\text{retr}}} & i = 1 + N_{\text{retr}}, \end{cases}$$
(22)



FIGURE 6: Situation where a packet codeword is distributed over 3 packet groups (N = 5,  $N_{coh} = 3$ ,  $N_{group} = 3$ ).

where  $P_{\text{pack}}$  is the probability that a packet is erased and irrespective of the channel condition

$$P_{\text{pack}} = \int_{0}^{+\infty} P_{\text{pack}}(x) p(x) dx.$$
 (23)

For large  $E_b/N_0$ ,  $P_{\text{pack}} \propto (E_b/N_0)^{-D}$ . From (22) we obtain

$$E[\text{ovh}] = P_{\text{pack}} \frac{1 - P_{\text{pack}}^{N_{\text{retr}}}}{1 - P_{\text{pack}}}.$$
 (24)

For large  $E_b/N_0$ , we have  $E[\text{ovh}] \approx P_{\text{pack}} \propto (E_b/N_0)^{-D}$ . This indicates that the average overhead resulting from SR ARQ decreases with increasing  $E_b/N_0$  and increasing PHY layer diversity *D*.

# 4.4. Comparison of RS erasure coding and selective repeat ARQ

For high  $E_b/N_0$ , given packet transmission rate  $R_{\text{pack}}$  and a given PHY layer diversity D, the system yielding the largest diversity gain gives rise to the smallest  $E[\#\text{GOP}_{\text{unrec}}]$ . In the case of RS erasure coding, the highest possible diversity gain  $\gamma_{\text{RS, max}}$  equals  $[T_{\text{lat}}/T_{\text{coh}}]$ , which is achieved for  $\text{ovh} \rightarrow \infty$ . For SR ARQ, the maximum diversity gain is  $\gamma_{\text{ARQ, max}} = 1 + [T_{\text{lat}}/T_{\text{coh}}]$ ; this gain is obtained when  $T_{\text{retr}} = T_{\text{coh}}$ , which is the smallest value of  $T_{\text{retr}}$  that yields statistically independent (re)transmissions of the same packet. Unless  $T_{\text{lat}}$  is an integer multiple of  $T_{\text{coh}}$ , we get  $\gamma_{\text{RS, max}} = \gamma_{\text{ARQ, max}}$ , which indicates that RS erasure coding and SR ARQ yield the same potential diversity gain. However, the achievable diversity gain is limited by practical constraints.

- (i) In the case of RS erasure coding, the allowable overhead *ovh* is limited by bandwidth constraints. In most practical systems, one imposes the constraint ovh < 1, so that (14) yields  $\gamma_{\rm RS} < [T_{\rm lat}/(2T_{\rm coh})] \approx \gamma_{\rm RS,max}/2$ : under this constaint on the overhead, at most half of the maximum possible diversity gain is achievable.
- (ii) In the case of SR ARQ,  $\gamma_{ARQ} = 1 + [T_{lat}/max(T_{coh}, T_{retr,min})]$  so that the maximum diversity gain  $\gamma_{ARQ, max}$  cannot be achieved when  $T_{retr, min} > T_{coh}$ .

Hence, the diversity gain resulting from RS erasure coding is limited by the allowed overhead, whereas in the case of SR ARQ the diversity gain is limited by the ratio  $T_{\text{retr, min}}/T_{\text{coh}}$ . When  $T_{\text{retr, min}} < T_{\text{coh}}$ , the system with SR

ARQ yields the largest possible diversity gain  $\gamma_{ARQ, max}$ , and outperforms the system with RS erasure decoding. When  $T_{retr, min} > T_{coh}$ , neither RS erasure coding nor SR ARQ achieves the maximum possible diversity gain; when

$$\operatorname{ovh} < \left(\frac{T_{\operatorname{retr,min}}}{T_{\operatorname{coh}}} - 1\right)^{-1},\tag{25}$$

the system with SR ARQ outperforms the system with RS erasure coding; otherwise, the system with RS erasure coding yields the better performance. For example, it follows from (25) that RS erasure decoding needs an overhead larger than 50% in order to beat SR ARQ with  $T_{\text{retr}, \min} = 3T_{\text{coh}}$ .

The RS erasure coding introduces a fixed overhead and latency, which are determined by the parameters (N,K) of the RS code. In the case of SR ARQ, the number of retransmissions of a packet is a random number between 0 and  $N_{\rm tr}$ . Therefore, the latency and overhead resulting from SR ARQ are also random, with a maximum value determined by  $N_{\rm tr}$ , and an average value that decreases with increasing  $E_b/N_0$  and increasing PHY layer diversity *D*; typically, these averages are considerably smaller than the fixed overhead and latency resulting from RS erasure coding.

Further, from the complexity point of view, one should take into account that the system with SR ARQ requires the presence of a return channel and an increase of the functionality (beyond the IP layer) of the retransmitting network node (DSLAM or HG). The system with RS erasure coding requires additional complexity for the construction (at the video server) and the decoding (at the STB) of the RS packet codeword.

Finally, we mention that the achieved diversity gain depends neither on the packet size *L* nor on the packet transmission rate  $R_{\text{pack}}$ , but solely on the parameters  $T_{\text{lat}}/T_{\text{coh}}$  and (for RS erasure coding) ovh or (for SR ARQ)  $T_{\text{retr,min}}/T_{\text{coh}}$ .

#### 5. NUMERICAL RESULTS

#### 5.1. General numerical results

Assuming that a packet consists of  $L = 10^4$  bits and a packet group contains  $N_{\rm coh} = 5$  packets, we have displayed in Figures 7–11 several quantities as a function of  $E_b/N_0$ , for SISO (D = 1) and Alamouti with 1 or 2 receive antennas (D = 2 or D = 4). The presented curves confirm the high  $E_b/N_0$  behavior that we established in Section 4, and illustrate the impact of the PHY layer diversity D on the performance.

- (i) Figure 7 shows the probability  $P_{\text{pack}}$  from (23) that a packet is erased after transmission over the wireless link. We observe that  $P_{\text{pack}} \propto (E_b/N_0)^{-D}$  at high  $E_b/N_0$ .
- (ii) The average number of erased packets in a packet group, conditioned on the event that at least 1 packet from the group has been erased, is shown in Figure 8. Note that even at large  $E_b/N_0$ , packet erasures tend to occur in bursts: as the channel state is constant over the channel coherence time, a small value of the channel state (deep fade) is likely to give rise to multiple erasures within a packet group.



FIGURE 7: Probability  $P_{\text{pack}}$  that a packet is erased.



FIGURE 8: Average number of erased packets in a packet group, conditioned on the event that at least one packet in the packet group is erased.



FIGURE 9: Probability of a decoding failure.

- (iii) Figure 9 shows Pr(decoding failure) (see (13)), for N = 100 and N K = 10. As a decoding failure occurs when at least 11 packets in the codeword are erased, a minimum of 3 packet groups is involved in a decoding failure. Hence, according to Section 4, Pr[decoding failure]  $\propto (E_b/N_0)^{-3D}$  at high  $E_b/N_0$ , which is confirmed by Figure 9.
- (iv) Figure 10 shows the average transmission overhead E[ovh] from (24), that results from SR ARQ with a maximum of 3 retransmissions. Comparison with Figure 7 reveals that  $E[\text{ovh}] \propto P_{\text{pack}}$  at high  $E_b/N_0$ , which confirms our results from Section 4. At small  $E_b/N_0$ , E[ovh] converges to  $N_r = 3$ , which corresponds to the case where each packet is retransmitted  $N_r$  times.
- (v) Figure 11 shows the probability  $P_{\text{group, unrec}}$  (see (19)) that at least one packet from a packet group is definitively lost after 3 retransmissions. Note that  $P_{\text{group, unrec}} \propto (E_b/N_0)^{-4D}$  at high  $E_b/N_0$ .

# 5.2. Results applied to HDTV transmission over a 60 GHz indoor wireless link

Now we consider the transmission of compressed HDTV [19] according to the configuration shown in Figure 1. The compressed video bitrate equals 7.5 Mbps. The link between the HG and the STB is a 60 GHz indoor wireless connection; assuming nonline-of-sight (NLOS) conditions, this connection is modeled as a Rayleigh fading channel, with a coherence time  $T_{\rm coh} = 20$  milliseconds (corresponding to slow motion of about 0.4 m/s) [20]. In order to limit the zapping delay, the latency  $T_{\rm lat}$  caused by protecting the video packets against erasures should not exceed 150 milliseconds [21]. The HDTV performance target is a maximum of 1 GOP with unrecoverable packets in 12 hours.

When protecting the video packets by means of an RS packet codeword, we consider transmission overheads of 10%, 20%, and 40%.

When using SR ARQ, we consider two distinct scenarios as far as the location of the retransmission buffer is concerned.

- (i) When the retransmission buffer is located at the HG,  $T_{\text{retr,min}}$  is limited to about 5 milliseconds. As 5 milliseconds is less than the 20 milliseconds channel coherence time, the transmitter will defer the retransmission of a packet until 20 milliseconds have elapsed since the previous (re)transmission of the considered packet; hence, this yields  $T_{\text{retr}} = 20$  milliseconds.
- (ii) In the case of a low-cost HG, the retransmission buffer is not located at the HG but further upstream, at the DSLAM. The resulting  $T_{\text{retr, min}}$  is on the order of 45 milliseconds [22, 23], which exceeds the 20 milliseconds channel coherence time. In this case, we have  $T_{\text{retr}} = 45$  milliseconds.

Assuming that the average sizes of an I-frame and a P-frame are 6 times and 2 times the average size of a



FIGURE 10: Average transmission overhead *E*[ovh] from ARQ with maximum 3 retransmissions.



FIGURE 11: Probability  $P_{\text{group, eras}}$  that at least one packet from a packet group is definitively erased (ARQ with maximum 3 retransmissions).



FIGURE 12: Average number of GOPs affected by unrecoverable packet loss in 12 hours (SISO, ARQ).

B-frame, Table 1 shows the average sizes of the different types of frames and of the GOP consisting of the frame sequence IBBPBBPBBPBBP. Note that each type of frame gives rise to multiple IP packets. As the IP packet rate is about 700 packets/s and the channel coherence time is 20 milliseconds, about 14 IP packets fit within the channel coherence time (assuming that IP packets are transmitted at constant regular intervals). Taking into account the propagation of errors from an I- or P-frame to other frames in the GOP, unrecoverable packet loss in an I- or P-frame is very likely to give rise to a visual distortion. Considering that I- and Pframes in a GOP constitute on average 60% of the IP video packets, and packet losses tend to occur in bursts with sizes comparable to the channel coherence time (14 IP packets in our scenario), it follows that when a GOP is affected by an unrecoverable packet loss, the probability that the packet losses occur in I- or P-frame is about 60%. Assuming that packet losses in B-frames are unnoticed but losses in I- or P-frames yield visible distortions, the probability that a GOP affected by unrecoverable packet loss yields a visual distortion is about 60%. (In [20], an experiment is reported which indicates that there is a probability of about 20% that a lost packet yields a visual distortion. However, in [20] the packet losses do not occur in bursts. In the case of bursty packet losses, the probability that a burst of packet losses yields a visual distorition is expected to be larger than 20%.) Moreover, some of the IP packets contain other information (audio, data) related to the HDTV program, that is multiplexed with the video information. The loss of packets containing a multiplex of B-frame information and other HDTV-related information reduces the QoE (because of audible clicks), although the errors in the B-frame do not propagate and could be concealed. Therefore, the average number of GOPs that is affected by unrecoverable packet loss in 12 hours is a meaningful indicator of the QoE.

When conducting the performance analysis, we assumed that the erasure probability on the DSL link is negligibly small as compared to that on the wireless link between the HG and the STB.

Figures 12–18 show the average number of GOPs with unrecoverable packet loss in 12 hours as a function of  $E_b/N_0$ , for the different combinations of PHY layer strategies (SISO and Alamouti with 1 or 2 receive antennas) and packet protection strategies (SR ARQ, RS erasure coding, none). When using SR ARQ, the cases  $T_{retr} = 45$  milliseconds and  $T_{retr} = 20$  milliseconds correspond to diversity gains  $\gamma_{ARQ}$  of 4 (max. 3 retransmission) and 8 (max. 7 retransmissions), respectively. In the case of RS erasure coding, overheads of 10%, 20%, and 40% yield diversity gains  $\gamma_{RS}$  of 1 (i.e., no diversity gain), 2, and 3, respectively. Considering as a performance figure the value of  $E_b/N_0$  that corresponds to  $E(no. of GOP_{unrec} in 12 hours) = 1$ , Table 2 collects the performance figure for the different cases. The following observations can be made.

(i) The highest possible diversity gain is [*T*<sub>lat</sub>/*T*<sub>coh</sub>] =
 8. This diversity gain is achieved for SR ARQ with *T*<sub>retr</sub> = *T*<sub>coh</sub>, that is, when the retransmission buffer is at the HG.

	GOP = {IBBPBBPBBPBB}, 25 frames/s, 7.5 Mbit/s video bitrate		
	size (kbit)	# MPEG-2 TS packets	# IP packets
one I-frame	1080	714	102
one P-frame	360	238	34
one B-frame	180	119	17
one GOP	3600	2380	340

TABLE 1: Average sizes of I-frame, P-frame, B-frame, and GOP.

TABLE 2: Value of  $E_b/N_0$  yielding 1 GOP with unrecoverable packet loss per 12 hours.

		$E_b/N_0 @ E[\#GOP_{unrec} in 12 hours] = 1$				
	no RS,	AI	RQ		RS erasure decoding	5
	no ARQ	$T_{\rm del} = 20 \ {\rm ms}$	$T_{\rm del} = 45 \ {\rm ms}$	ovh = 10%	ovh = 20%	ovh = 40%
SISO	73 dB	17 dB	25.5 dB	71 dB	43 dB	31 dB
Alamouti, $N_r = 1$	43 dB	14 dB	18.5 dB	41 dB	27.5 dB	20.5 dB
Alamouti, $N_r = 2$	25.5 dB	9 dB	12 dB	23.5 dB	16.5 dB	12.5 dB

- (ii) Because of their larger diversity gain, the systems with SR ARQ outperform the systems with RS coding. In order to achieve a diversity gain of 4, the transmission overhead of systems with RS coding should be increased to about 70%. A diversity gain of 2 is obtained for the systems with SR ARQ when  $T_{\text{retr}}$  is between 50 milliseconds and 75 milliseconds.
- (iii) Figure 18 compares RS coding and SR ARQ in terms of  $E(no. of GOP_{unrec} in 12 hours)$  for Alamouti with 1 receive antenna, where the system parameters have been selected such that RS coding and SR ARQ yield the same diversity (see Table 3). We observe that the RS code performs worse than SR ARQ. This is because for the RS code the number of dominant erasure patterns yielding irrecoverable packet loss is larger than for SR ARQ.
- (iv) The performance of the SISO system without any packet protection is very poor. The performance is improved by space-time coding on the PHY layer (which increases the PHY layer diversity *D*) and/or packet protection by means of RS coding or SR ARQ (which provides additional diversity gain). To some extent, less packet protection can be compensated by using more receive antennas, and vice versa.

# 6. CONCLUSIONS AND REMARKS

In this paper, we have considered a generic system for video transmission over a wireless link, with space-time coding on the PHY layer and additional video packet protection by means of SR ARQ or RS erasure coding. We have pointed out that SR ARQ and RS erasure coding give rise to a diversity gain yielding improved error performance, and have presented simple analytical expressions for this gain. For both SR ARQ and RS erasure coding, the maximum possible diversity gain equals  $[T_{\text{lat}}/T_{\text{coh}}]$ . However, when

	TABLE 3	
	RS	SR ARQ
$\gamma_{\rm RS} = \gamma_{\rm ARQ} = 2$	ovh = 20%	$N_{\rm retr} = 1$
$\gamma_{\rm RS} = \gamma_{\rm ARQ} = 3$	ovh = 40%	$N_{\rm retr} = 2$



FIGURE 13: Average number of GOPs affected by unrecoverable packet loss in 12 hours (Alamouti,  $N_r = 1$ , ARQ).

using RS erasure coding this maximum diversity gain cannot be achieved because of practical limitations on the allowed transmission overhead. SR ARQ yields the maximum diversity gain provided that  $T_{\text{retr, min}} < T_{\text{coh}}$ ; otherwise, the actual diversity gain is less. Our theoretical findings have been illustrated in a case study involving HDTV transmission over a 60 GHz indoor wireless link.

The RS erasure coding gives rise to a fixed overhead and latency that are determined by the parameters of the RS code. In the case of SR ARQ, the instantaneous overhead and latency are random; their maximum values are determined



FIGURE 14: Average number of GOPs affected by unrecoverable packet loss in 12 hours (Alamouti,  $N_r = 2$ , ARQ).

by the maximum number of retransmissions, while their averages decrease with increasing  $E_b/N_0$  are considerably less than the corresponding values for RS erasure coding.

The application of RS erasure coding does not require any modifications of the functionality of the intermediate network nodes, as the construction and the decoding of the RS packet codewords are carried out by the video server and the end user, respectively. Application of SR ARQ involves increasing the functionality (and cost) of the network node where the retransmission buffer is located. From an error performance point of view, the HG should be selected as the retransmitting node, as it provides the smallest round-trip delay and, hence, the largest diversity gain; however, in order to keep the HG a low-cost consumer product, the DSLAM can be selected as the retransmitting node, with the penalty of a larger round-trip delay and a smaller resulting diversity gain. Further, application of ARQ requires the presence of a return channel.

Our performance analysis assumes that the channel state is the same for all OFDM subcarriers. This assumption is valid when the signal bandwidth  $(R_s)$  does not exceed the 90% coherence bandwidth of the channel. For the 60 GHz indoor radio channel under NLOS conditions, the 90% coherence bandwidth is about 6 MHz [24], so that our analysis is valid for bitrates up to 12 Mbps (assuming QPSK transmission). When the signal bandwidth is larger than the 90% coherence bandwidth, different subcarriers experience different channel states (which could be exploited to increase the PHY layer diversity by means of frequency-interleaving and coding across the subcarriers of an OFDM block). The detailed analysis of this case is beyond the scope of this paper, but we have been able to verify that the diversity gains  $y_{\rm RS}$  and  $\gamma_{ARQ}$  from Section 4 still apply, so that the main conclusions from this paper remain valid.

WLANs often make use of stop-and-wait (S&W) ARQ on the MAC layer. This form of ARQ has not been included in our performance analysis. We briefly explain how the presence of S&W ARQ on the MAC layer affects the performance. Denoting by  $N_{\text{retr, S&W}}$ , and  $T_{\text{retr, S&W}}$  the maximum



FIGURE 15: Average number of GOPs affected by unrecoverable packet loss in 12 hours (SISO, RS).



FIGURE 16: Average number of GOPs affected by unrecoverable packet loss in 12 hours (Alamouti,  $N_r = 1$ , RS).

number of retransmissions and the time interval between (re)transmissions of a same packet, S&W ARQ introduces a maximum latency of  $T_{\text{lat, S&W}} = N_{\text{retr, S&W}} T_{\text{retr, S&W}}$ . When combined with RS erasure coding, the resulting maximum latency equals  $T_{\text{lat}} = T_{\text{lat, S&W}} + K/R_{\text{pack}}$ . When combined with SR ARQ, the resulting maximum latency equals  $T_{\text{lat}} =$  $N_{\text{retr, SR}}T_{\text{retr, SR}} + T_{\text{lat, S&W}}$  with  $N_{\text{retr, SR}}$  and  $T_{\text{retr, SR}}$  denoting the maximum number of retransmissions and the time between (re)transmissions of the same packet for the SR ARQ protocol; because of the restriction  $T_{\text{retr, SR}} > T_{\text{lat, S&W}}$ , we get  $T_{\text{lat}} > (N_{\text{retr, SR}} + 1)T_{\text{lat, S&W}}$ . The resulting diversity order is given by  $\gamma_{S\&W}\gamma_{RS}D$  (RS erasure coding) or  $\gamma_{S\&W}\gamma_{SR}D$ (SR ARQ), where  $\gamma_{RS} = [(N - K + 1)/N_{coh}], \gamma_{RS} = 1 +$  $N_{\text{retr, SR}}$ , and  $\gamma_{\text{S\&W}}$  is the diversity gain resulting from the S&W ARQ protocol on the MAC layer. As the diversity order does not increase when retransmitted packets experience the same channel state as the packet originally transmitted, the diversity gain from S&W ARQ is evaluated as  $\gamma_{S\&W}$  =  $[T_{\text{lat, S&W}}/T_{\text{coh}}].$ 



FIGURE 17: Average number of GOPs affected by unrecoverable packet loss in 12 hours (Alamouti,  $N_r = 2$ , RS).



FIGURE 18: Average number of GOPs affected by unrecoverable packet loss in 12 hours, RS versus ARQ (Alamouti,  $N_r = 1$ ).

### ACKNOWLEDGMENTS

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (Contract no. 216715). The second author is a Postdoctoral Fellow with the Fund for Scientific Research, Flanders (FWO-Vlaanderen), Belgium.

# REFERENCES

- J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.
- [2] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [3] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*, Cambridge University Press, Cambridge, UK, 2007.

- [4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [5] G. C. Clark Jr. and J. B. Cain, Error-Correction Coding for Digital Communications, Springer, New York, NY, USA, 1981.
- [6] J. G. Proakis, *Digital Communications*, McGraw Hill, New York, NY, USA, 2000.
- [7] H. O. Burton and D. D. Sullivan, "Errors and error control," *Proceedings of the IEEE*, vol. 60, no. 11, pp. 1293–1301, 1972.
- [8] R. A. Comroe and D. J. Costello Jr., "ARQ schemes for data transmission in mobile radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 472–481, 1984.
- [9] M. Luby, L. Vicisano, J. Gemmell, L. Rizzo, M. Handley, and J. Crowcroft, "The use of forward error correction (FEC) in reliable multicast," IETF RFC 3453, December 2002.
- [10] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor codes for reliable download delivery in wireless broadcast systems," in *Proceedings of the 3rd IEEE Consumer Communications and Networking Conference (CCNC '06)*, vol. 1, pp. 192–197, Las Vegas, Nev, USA, January 2006.
- [11] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," IETF RFC 2733, December 1999.
- [12] F. Vanhaverbeke, F. Simoens, M. Moeneclaey, and D. De Vleeschauwer, "Binary erasure codes for packet transmission subject to correlated erasures," in *Proceedings of the 7th Pacific Rim Conference on Multimedia (PCM '06)*, pp. 48–55, Hangzhou, China, November 2006.
- [13] S. B. Wicker and V. K. Bhargava, Reed-Solomon Codes and Their Applications, IEEE Press, New York, NY, USA, 1994.
- [14] S. R. Ely and C. Eng, "MPEG video coding: a basic tutorial introduction," Research and Development Report BBC RD 1996/3, British Broadcasting Corporation, London, UK, 1996, http://downloads.bbc.co.uk/rd/pubs/reports/1996-03.pdf.
- [15] P. A. Sarginson, "MPEG-2: overview of the systems layer," Research and Development Report BBC RD 1996/2, British Broadcasting Corporation, London, UK, 1996, http://downloads.bbc.co.uk/rd/pubs/reports/1996-02.pdf.
- [16] Overview of digital subscriber line (DSL) recommendations, ITU-R recommendation G.995.1, February 2001.
- [17] D. Hoffman, G. Fernando, V. Goyal, and M. Civanlar, "RTP payload format for MPEG1/MPEG2 video," IETF RFC 2250, January 1998.
- [18] N. Degrande, K. Laevens, D. De Vleeschauwer, and R. Sharpe, "Increasing the user perceived quality for IPTV services," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 94–100, 2008.
- [19] "Parameter values for the HDTV+ standards for production and international programme ex-change," ITU-R recommendation BT.709-5, 2002.
- [20] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341– 355, 2006.
- [21] M. Watson, "Proposal for evaluation process for forward error correction codes for DVB-IPI," DVB IPI document TM-IPI2084 edition, September 2005.
- [22] A. Gurtov and S. Floyd, "Modelling wireless links for transport protocols," ACM Computer Communications Review, vol. 34, no. 2, pp. 85–96, 2004.

- [23] C. Hoene, A. Gunther, and A. Wolisz, "Measuring the impact of slow user motion on packet loss and delay over IEEE 802.11b wireless links," in *Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks* (*LCN* '03), pp. 652–662, Bonn, Germany, October 2003.
- [24] H. Yang, P. F. M. Smulders, and M. H. A. J. Herben, "Channel characteristics and transmission performance for various channel configurations at 60 GHz," *EURASIP Journal* on Wireless Communications and Networking, vol. 2007, Article ID 19613, 15 pages, 2007.

# Research Article

# A Transparent Loss Recovery Scheme Using Packet Redirection for Wireless Video Transmissions

# Chi-Huang Shih,<sup>1</sup> Ce-Kuen Shieh,<sup>1</sup> and Wen-Shyang Hwang<sup>2</sup>

<sup>1</sup> Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan <sup>2</sup> Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung 807, Taiwan

Correspondence should be addressed to Chi-Huang Shih, chshih@hpds.ee.ncku.edu.tw

Received 1 October 2007; Revised 14 February 2008; Accepted 17 March 2008

Recommended by F. Babich

With the wide deployment of wireless networks and the rapid integration of various emerging networking technologies nowadays, Internet video applications must be updated on a sufficiently timely basis to support high end-to-end quality of service (QoS) levels over heterogeneous infrastructures. However, updating the legacy applications to provide QoS support is both complex and expensive since the video applications must communicate with underlying architectures when carrying out QoS provisioning, and furthermore, should be both aware of and adaptive to variations in the network conditions. Accordingly, this paper presents a transparent loss recovery scheme to transparently support the robust video transmission on behalf of real-time streaming video applications. The proposed scheme includes the following two modules: (i) a transparent QoS mechanism which enables the QoS setup of video applications without the requirement for any modification of the existing legacy applications through its use of an efficient packet redirection scheme; and (ii) an instant frame-level FEC technique which performs online FEC bandwidth allocation within TCP-friendly rate constraints in a frame-by-frame basis to minimize the additional FEC processing delay. The experimental results show that the proposed scheme achieves nearly the same video quality that can be obtained by the optimal frame-level FEC under varying network conditions while maintaining low end-to-end delay.

Copyright © 2008 Chi-Huang Shih et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. INTRODUCTION

As different types of wireless networks are converging into the wired Internet, providing end-to-end quality of service (QoS) is essential for video transmission over the wireless Internet. Enabling QoS involves multidisciplinary solutions and the related areas could range from end-users applications to the underlying network architectures. Generally, the QoS support in wireless Internet can be achieved through the network-centric and the end-system centric approaches [1].In the network-centric approach, the integrated service (IntServ) [2] and the differentiated service (DiffServ) [3] are two well-known architectures to support QoS provisioning for the Internet. IntServ and signaling protocols, such as reservation protocol (RSVP), provide the per-flow QoS guarantee. DiffServ provides both the guaranteed and relative QoS by dividing packets into different service classes and forwarding them as different priorities. For wireless networks, there have been many studies related to the QoS provisioning such as the third generation partnership

project (3GPP) for UMTS networks [4], IEEE 802.11e for wireless local area networks [5], and IEEE 802.16 for wireless local and metropolitan area networks [6]. The core wireless infrastructures need to provide prioritized QoS services to support various applications with different requirements.

On the other hand, the end-system centric approach is considered as a QoS enhancement solution without the underlying QoS architectures from the network. For the time-varying network conditions, the applications can employ the adaptive control mechanisms to minimize the impairment of video delivery caused by channel errors and congestions during transmission. Congestion control and error control are two main mechanisms to support the robust video transmission. Congestion control mechanism aims at reducing packet loss and delay due to network congestion. Error-control mechanism intends to combat the transmission errors by recovering lost data; for example, two popular error-control schemes are forward error correction (FEC) and automatic repeat request (ARQ).

To support end-to-end QoS over the wireless Internet, the applications are expected to have two following QoS enhancements: (1) the communication between end users and the underlying QoS architecture for QoS negotiation; and (2) the employment of network adaptive control mechanisms to achieve the effective adaptation of network conditions. The QoS-aware middleware has been introduced to render the application design independent from the underlying network QoS architecture and further achieve the integration of different QoS solutions [7]. It is noted that the major changes to both the end system and legacy video application are necessary, and therefore the deployment of QoS-enabled wireless video services could be slow and difficult. This problem can be solved by improving the degree of transparency to minimize the modifications of the legacy applications. In [8], by adopting so-called QoS library redirection (QLR), applications can set up QoS without source-code modification but the application-dependant library is required for all target applications individually. In [9],a link-layer performance enhancing proxy (PEP) is proposed to cope with the impairment introduced by wireless links over the Internet stack. However, the link-layer approach is inflexible to enable QoS for diverse applications with different requirements.

In this paper, a transparent QoS mechanism (TQM) is proposed to provide a flexible platform to transparently support QoS services. It is noted that this paper focuses on the provision of video services in IP-based wireless networks. Therefore, the proposed TQM transparently supports IP services by employing a packet redirection technique that operates based on the TCP/IP stack. Based on the technique of packet redirection, TQM performs the packet control operations to provide various QoS enhancements for the legacy applications. The proposed TQM aims at facilitating OoS deployment over the wireless Internet. The main features of TQM are: (1) end users can specify which target application receives what types of QoS enhancements and the target application is transparently to be QoS-aware without any modification; (2) no modifications are required to the existing Internet transport protocols; and (3) TQM supports diverse QoS enhancement modules (EM) and thus one can integrate various EMs to maximize the perceived video quality. In this paper, we primarily focus on designing an adaptive error-control scheme for TQM to cope with the varying wireless channel errors, and particularly its packetlevel FEC enhancement module.

FEC introduces the redundancy that trades the additional bandwidth cost to protect video streams from wireless losses. Unfortunately, the effectiveness of FEC decreases since the redundancy cost could lead the self-induced congestion to cause the adverse effects on video quality such as congestion losses due to buffer overflow and the longer end-to-end latency due to queueing delay [10, 11]. This is significant to the compressed VBR video source since it usually exhibits long-range dependence (LRD) with larger losses and/or delay within a concentrated period [12]. Congestion losses impede the successful loss recovery since the amount of packet losses induced by both wireless error and congestion might exceed the error correction capacity of FEC. The longer end-to-end latency makes packet arrival useless to video decoder with the timing constraint. In addition, the failed redundancy and also the useless video data lead the unnecessary bandwidth waste, studied as the congestion collapse issues by recent works, would degrade the network utility [13]. To enhance the effectiveness of FEC, it is therefore necessary to consider both the loss recovery aspects of FEC and the level of network congestion. Park and Wang [10] consider the optimal problem of designing an adaptive FEC protocol for real-time MPEG video transmission over the Internet without regard to TCP-friendly transmission rate constraints. Their proposed FEC-control method adapts the redundancy degree to perceived packet loss on the network. When increased redundancy results in a nonincreasing recovery performance due to the fact that selfcongestion impedes the timely recovery of video information, the adaptive FEC protocol exponentially decreases the redundancy degree to avoid adverse network effects on video quality. On the other hand, Wu et al. [14] derive the analytical FEC model for a TCP-friendly MPEG video stream with temporal scaling to obtain the optimal reconstruction quality in the presence of packet losses. In their model, FEC is applied to different types of video frames while the temporal scaling technique is used to adjust the stream data rate by discarding frames based on the frame dependency of MPEG video. Yuan et al. [15] present an FEC model, which applies FEC at the group of picture (GOP) level, to increase the error correction capacity of FEC for MPEG video streams within TCPfriendly constraints. Compared with the optimal frame-level FEC technique proposed in [14] by Wu et al., this GOP-level FEC technique requires more computational complexity in average to process a larger amount of video data. However, both techniques rely on the presence of a large buffer to collect an entire GOP for optimally calculating the FEC coding rate in the video sender. This results in a coding buffering delay on the order of GOP duration in the video receiver before the smooth video presentation begins. The coding buffering delay generally contributes to the overall end-to-end delay of video. Usually, the acceptable delay depends on the video applications. For interactive services, such as video conferencing, the end-to-end delay should not exceed 100 milliseconds to ensure good quality [16].

In this paper, the design of FEC-on-TQM integrates several EMs to transparently support robust video transmission on behalf of real-time streaming video applications. FECon-TQM utilizes an instant frame-level FEC technique to minimize the coding buffering delay experienced by the user, while still maintaining near-highest video quality that the optimal frame-level FEC can obtain within the TCPfriendly rate constraints. In order to maintain high video quality with low delay, we first derive a model of video frame priorities based on the temporal dependency of MPEG video to distribute the available TCP-friendly bandwidth budget among video frames. Then the decision of applying FEC to video frames or discarding frames to match the TCP-friendly transmission rate is done on a frame-by-frame basis. To evaluate the performances of the proposed scheme, we constructed the experiments in a controlled network environment. The experimental results show that the instant frame-level FEC technique achieves nearly the same video quality that can be obtained by the optimal frame-level FEC while maintaining low end-to-end delay of video. Based on the above mentioned techniques of packet redirection and instant frame-level FEC, FEC-on-TQM carries out the transparent loss recovery without any modification of legacy applications and obtains a high delivered video quality for low-delay video streaming services.

The remainder of this paper is organized as follows. Section 2 describes the basic operating principles of TQM, while Section 3 describes the design of FEC-on-TQM and its implementation issues. Section 4 reviews the instant frame-level FEC control scheme. Section 5 presents and discusses the experimental performance evaluation results. Finally, Section 6 provides some brief concluding remarks and indicates the intended direction of future research.

# 2. TRANSPARENT QoS MECHANISM (TQM)

# 2.1. TQM architecture

The TQM mechanism proposed in this study provides a transparent QoS enhancement for Internet multimedia applications. In order to improve its flexibility, TQM is designed for implementation at the application layer, and therefore enables existing Internet transport protocols to operate in their usual way. As shown in Figure 1, TQM uses two modules, that is, "Flow State" and "QoS Manager", to accommodate the diverse characteristics of existing applications and their underlying transport protocols. When an application is launched, some records are created in the kernel space for system communication and maintenance purposes. Without modifying either the kernel or the application, the virtual flow state module in TQM collects flow information from these records and presents this information to the QoS Manager. The user can then interact with the QoS Manager to specify directly those applications which require enhanced QoS support. Subsequently, the QoS Manager applies the flow information supplied by the flow state module to transparently carry out EM operations on behalf of the applications.

TQM achieves transparent QoS enhancement by means of a packet redirection scheme. Specifically, the implementation of QoS-manager module is based on the functionality of the IP firewall [17] as well as the divert socket [18]. IP firewall filters packets traveling up or down the IP stack; and it defines the target action on these filtered packets, according to firewall rules. Instead of specifying typical target actions such as ACCEPT or DENY, target DIVERT can redirect filtered packets to a divert socket. Divert socket is one element of general BSD socket and can be bound to a specific port of the host for IP packet interception and injection. Since the IP firewall is located at the bottom of the IP stack, it redirects the IP flows which are specified in divert messages received from the QoS manager to a specific system port. The QoS-manager module employs the divert socket to bind this specific port and then receives the IP data packets from it. Using the same port, the EM-



FIGURE 1: Basic components and operations of TQM host. Note that sender data flow is marked as "Out" (solid line) and receiver data flow is marked as "In" (dotted line).

processed IP packets are then injected back into the general network protocol stack and are subsequently processed using regular system routines. Accordingly, the packet redirection of TQM operates in a two-way manner. In the TQM sender, the QoS Manager redirects data flows generated by the applications and injects them into the protocol stack for network transmission. Conversely, in the TQM receiver, the QoS Manager redirects the data flows received from the underlying network infrastructure and then injects them into upper-layer applications.

Figure 2 presents a detailed illustration of the major components in the TQM QoS Manager. TQM provides a transparent QoS enhancement capability through the use of data planes and control planes. In the data plane, userspecific flows are identified and the related flow information is passed to the underlying EM. Executing the flow information management function in the data plane involves three separate components, namely the application filter, the user interface, and the application list. Briefly, the application filter collects the flow information relating to launched applications from the flow state module, and users monitor their applications on the application list through the user interface. By accessing the user interface, a user can view those flows which have been launched. Subsequently, he or she can specify the particular flow (or flows) for which enhanced QoS support is required. By querying the application list, the application filter detects and discards any flow information relating to applications which have not been specified by a user.

# Application list

Some applications, for example, DNS query, time services, http services, and so forth, do not have strict QoS requirements. Therefore, TQM uses the application list to indicate the multimedia applications for which the user specifies that QoS enhancement support may be required. The user selects



FIGURE 2: TQM architecture.

these specific applications on the application list via the user interface, and can arbitrarily add or delete selection records on demand. By accessing this list, the application filter can indicate to the flow state module the specific applications for which it should collect flow information.

### Application filter

Using the flow state module, the application filter retrieves the flow information required to transparently start QoS sessions on behalf of the applications. Generally, this information is related to five tuples (i.e., the transport protocol, the source IP address, the source port, the destination port, and the destination address).When a user requests support for a specific flow, this information is passed to the control plane, which then establishes the QoS session.

In the control plane, the EMs set up QoS sessions on behalf of the legacy applications. Importantly, the control plane in the proposed TQM mechanism contains many QoS EMs of different types. These EMs may be used either separately or cooperatively in order to carry out various functions. Consequently, TQM provides a flexible and efficient mechanism for the QoS enhancement of diverse applications. Based on the flow information received from the data plane, the EM identifies the user-specified flow and then intercepts the corresponding IP packets to carry out the QoS enhancement process. The EM-processed packets are then returned to the network protocol stack for further processing. Note that IP flow packets may be added, deleted or modified during QoS enhancement depending on the particular EM type.

### 2.2. EM types

TQM comprises a collection of different EMs to support video applications with diverse QoS requirements. Individual EMs may function independently for a specific IP flow or may cooperate with other EMs to improve the overall QoS. The TQM architecture is sufficiently flexible to support both the addition of new EMs and updates of the algorithms upon which the EMs are based. According to the general QoS solutions ranging from the underlying network architectures to upper applications, there are three basic examples of EM to support the QoS requirements of wireless video transmissions, namely, packet mapping control, TCPfriendly congestion control, and adaptive error control. These three EMs are discussed briefly in the following subsections.

#### 2.2.1. Packet mapping control EM

In general, QoS architectures employ some form of traffic category concept to support the implementation of scalable and manageable wireless networks with service differentiation capabilities. For example, the 3GPP working group has defined four different QoS classes, that is, conversational, streaming, interactive and background, based on a consideration of the delay sensitivity of different applications. Meanwhile, the 802.11e standard prescribes eight different traffic categories for wireless local area networks. Similarly, the 802.16 standard defines four different types of service flow in wireless metropolitan area networks, namely, unsolicited grant service, real-time polling service, nonreal-time polling service, and best effort service. Through application-specific QoS mapping mechanisms, data flows are assigned to the appropriate traffic category and are then transmitted with the corresponding priority. Furthermore, individual video packets may also be categorized in accordance with their loss and delay properties; and then assigned to different prioritized transmission classes in order to optimize the video quality under given rate or cost constraints.

To accommodate these various strategies, it is necessary to provide differentiation both among multiple flows and within single flow. The general differentiation parameters include the IP source/destination address, the protocol, the source/destination port number, and the type of service (TOS) value. In the proposed TQM mechanism, the packet mapping control EM maps the flows identified by the Flow State module to user-specified traffic categories. Meanwhile, differentiation within a single flow is achieved without modifying the applications by transparently diverting the IP packets to the packet mapping control EM. The packet mapping control EM first identifies the packet using appropriate classification criteria (e.g., video frame/layer type) and then maps the packet to the appropriate prioritized class. In mapping the packet to the prioritized class, the EM either directly forwards the identified packets to the underlying network architecture or marks packets with the proper TOS value according to the criteria specified by the underlying architecture before packet forwarding.

# 2.2.2. TCP-friendly congestion control EM

Since the capacity of a wireless channel is scarce and timevarying, bursty losses and excessive delays caused by network congestion can significantly degrade the perceived video quality. Accordingly, congestion control mechanisms aim to minimize the impairment of the delivered video quality caused by network congestion by reducing packet losses and delays. Additionally, video streams must share the available bandwidth equally with other TCP-based flows. Based on the QoS requirements of multimedia transport, TCPfriendly congestion control mechanisms smoothly adjust the transmission rate and avoid increasing the latency by not retransmitting lost packets. Two types of TCP-congestion control mechanism are generally used for multimedia applications, namely sender-based rate adjustment and modelbased rate adjustment. The former mechanism is similar to TCP in that it performs additive increase and multiplicative decrease (AIMD) control at the sender end [19]. Conversely, the latter scheme uses a throughput equation based on a stochastic TCP model to adjust the transmission rate as a function of the loss event rate and the round trip time (RTT) [20, 21].

To transport video in a TCP-friendly manner, video applications must match the output rate to the available network bandwidth, as estimated by a TCP-friendly congestion control protocol. One approach for achieving TCPfriendly video transmission is for the video application to use a rate control scheme to regulate the coded bit stream under the constraint of certain given conditions while simultaneously maximizing the user's perception of the media stream quality [22, 23]. However, this approach generally requires modification of the original legacy applications. An alternative approach is to control the transmission rate of the encoded video stream packets so as to obtain a tradeoff between the degree of TCP-friendliness and the perceived media quality [24]. The TQM mechanism proposed in this study constructs a TCP-friendly congestion control protocol which operates totally transparently to the legacy applications and supports these two approaches to achieve the rate matching. In the first approach, the

video transcoding can be used to adapt the output video stream to the available TFRC bandwidth [25]. The general video transcoding parameters consist of frame size, frame rate, and quantization value. In the second approach, the TCP-friendly congestion control EM uses a simple packet control mechanism to either discard packets according to the priority classes they belong to or, if the output rate of the video streams exceeds the TCP-friendly sending rate, to postpone packet transmission until the next transmission period.

#### 2.2.3. Adaptive error control EM

Error-control schemes aim at coping with the wireless error problems for high-bit-rate video transmissions over error-prone wireless networks. Typical examples of error control schemes presented in the literature include ARQ and FEC. The objective of both schemes is to obtain a higher data throughput by recovering corrupted packets. However, the two schemes adopt different strategies to achieve this objective. Specifically, ARQ retransmits lost packets, whereas FEC deliberately generates redundant data to enable the reconstruction of any video data which is lost during transmission. Although it has been shown that ARQ is more effective than FEC, its end-to-end retransmissions may impede the timely presentation of video content. Therefore, the FEC scheme is generally preferred for realtime video applications. Since the loss condition changes dynamically in wireless environments, and furthermore, the self-induced congestion caused by the generation of excessive redundant data has an adverse effect on the video quality, it is desirable for FEC-control schemes to have the ability to adapt dynamically to varying network conditions, that is, to changes in the packet loss rate or the level of network congestion. Therefore, the adaptive error control EM in the proposed TQM mechanism installs the FEC encoder in the data sender and the FEC decoder in the data receiver to support the transparent FEC coding of video applications.

The current study focuses primarily on the integration of these three EMs on the proposed TQM to carry out the robust video transmission and the testing of an instant frame-level FEC technique to adjust the number of redundant packets according to the network conditions.

#### 3. FEC-ON-TQM

#### 3.1. Overall structure

The aim of FEC-on-TQM is to enhance the perceived quality of the delivered video by transparently recovering packet losses on behalf of legacy video applications with no FEC capability. In this study, FEC-on-TQM integrates three EMs mentioned in Section 2.2 to achieve a tradeoff between robust video transmissions and efficient bandwidth utilization. Figure 3 illustrates the end-to-end communication between two TQM hosts. In the video sender, the packet mapping EM identifies the frame type of video packets redirected by the QoS manager and forwards a stream of video frames to the FEC EM. The FEC EM applies



FIGURE 3: FEC-on-TQM transmission scheme.

forward error correction on each frame using the TCPfriendly transmission rate passed on from the congestion control EM, producing a data packet stream including source video packets and redundant packets for network transmission. In the video receiver, the FEC EM collects a sequence of data packets belonging to the same frame and performs a loss recovery operation if required, that is, if a loss condition exists.

To improve the FEC efficiency, the QoS Manager in the video sender requires knowledge of the current network conditions such that the redundancy control function in the FEC mechanism can be adapted accordingly. In FECon-TQM, this is achieved by periodically sending feedback packets from the QoS manager in the video receiver to the QoS manager in the video sender. The feedback packet is useful for the TQM sender in providing packet loss rate and estimating the network round trip time. Since TQM redirects the IP packets of specific applications, packet loss information can be inspected by TQM. In addition to monitoring the packet loss rate, the QoS manager in the video receiver employs a loss differentiation algorithm (LDA) to indicate the congestion signal by discriminating congestion losses from wireless losses for all of the incoming packets, that is, both the source packets and the redundant packets. Two different types of LDA algorithm are commonly employed in receivers nowadays, namely split-connection LDA and end-to-end LDA. The former algorithms, for example, agent-based LDA, differentiate losses using some form of network assistance [26]. Conversely, end-to-end LDA schemes employ a time-based approach and measure queuing delays in an end-to-end manner. The measurement metric used in end-to-end LDA algorithms may be either the packet interarrival time [27] or the relative one-way trip time (ROTT) [28]. Since split-connection LDA algorithms generally require modification of the core network, the current FEC-on-TQM implementation adopts an end-toend LDA scheme. If the LDA regards the loss as a congestion loss, the congestion control EM includes it in the calculation of TCP-friendly transmission rate. Note that in the current deployment, the two end TQM hosts create a new connection to form a feedback channel to carry the packet loss information computed in the receiver through the common socket API. For RTP streams, these feedback messages could be merged into RTCP packets to reduce the control packet overhead [29].

# 3.2. FEC-on-TQM design

In this study, we use systematic Reed-Solomon erasure codes RS (n, k) to protect video data from channel losses. The RS encoder chooses k video data items as an FEC block and generates (n-k) redundant data items for the block. Every data item has its own sequence number used to indicate the corresponding position within the block. With this position information, the RS decoder can locate the position of the lost items and then correct up to (n-k) lost items. Furthermore, the FEC-on-TQM applies a packet-level RS code as FEC since it enables full transparency at the application layer and has a high efficiency over error-prone wireless channels [30]. Packet-level FEC schemes group the source data packets into blocks of a predetermined size k, and then encode n = k + h packets for network transmission, where  $h \ge 0$  is the number of redundant packets. Provided that k or more packets are successively received, the block can be completely reconstructed. In FEC-on-TQM, one feedback packet is sent from the video receiver to the video sender for every FEC block. Note that packet-level FEC extends the media stream simply by inserting redundant packets into the stream. Therefore, the method requires only minor modification to the source packets and supports a higher degree of transparency.

Figure 4 illustrates the data format of RTP packets through FEC-on-TQM. Other than the IP/UDP/RTP header, FEC is applied only to the source data (i.e., the payload) of the redirected IP packets. Following the FEC coding process, redundant data is generated and grouped into redundant packets in accordance with the header information on the source packets. Furthermore, an additional FEC header is added before the data portion to aid the decoding process. The RTP structures of both the source packets and the redundant packets are defined in accordance with the recommendations of RFC 2733 (In RFC 2733, the length of the RTP header is 12 bytes or more and the length of the FEC header is 12 bytes. In the later IETF Internet draft, the length of the FEC header is 10 bytes [31]) [32]. It is noted that changing the size of the packets or modifying the packet header requires a subsequent adjustment of the length and checksum fields in the IP header for IP validation purposes.Since the additional FEC header may cause the packet length to exceed the path maximum transit unit (MTU) size, the fragmentation may split a long packet into



FIGURE 4: IP datagram format of RTP stream through FEC-on-TQM.

two separate packets, resulting in a dependency between the two packets. However, in the receiver end, one of these packets may result in the corruption of the second packet, and hence the efficiency of the FEC recovery process is reduced. Based on the RTP header, FEC-on-TQM identifies the video frame type that the redirected packets belong to by extracting the payload type field and the video-specific header attached to the RTP fixed header. In addition, the timestamp field in the RTP header also helps the frame type extraction since the RTP packets belonging to the same video frame usually tagged with the similar timestamp value.

# 4. INSTANT FRAME-LEVEL FEC

This section develops the instant frame-level FEC technique for real-time streaming video flows within TCP-friendly constraints. To achieve a tradeoff between error correction capacity and low-delay video transmission, the proposed FEC technique determines the amount of FEC required by every video frame in the presence of packet loss when video frames are just dispatched by the streaming video application or the video encoder. A detailed discussion of the proposed instant frame-level FEC is presented in the paragraphs below.

#### 4.1. TCP-friendly video flows with frame-level FEC

A TCP-friendly video flow needs to regulate its output rate to match the TCP-friendly transmission rate. Based on a bitrate response function of Reno TCP [33], the TCP-friendly bandwidth T (in bytes/s) is given by

$$T = \frac{S}{r\sqrt{2p/3} + t_{\rm RTO}(3\sqrt{3p/8})p(1+32p^2)},$$
(1)

where *S* is the packet size in bytes, *r* is the round-trip time in seconds, *p* is the current packet loss probability, and  $t_{\text{RTO}}$  is the TCP retransmit timeout value in seconds.

In MPEG video, the raw video data are encoded as intracoded (I), predictive (P), and bidirectional (B) video frames. An I frame is encoded without dependence on any past frames. A P frame is encoded based on motion differences from the previous I frame or P frame. B frames are encoded based on the motion differences from the immediate past and future I or P frames. Due to the coding dependency, these three frame types (I, P, and B) have a descending order of importance. After coding, the I, P, and B frames are arranged in a periodic sequence which is called group of picture (GOP). For instance, a typical GOP pattern is IBBPBBPBBPBB and the GOP size is accordingly 12. Each frame has to be converted into packets for network transmission and the packet size should not be larger than the path MTU along the traversing links from the sender to the receiver. For a GOP of size L, the video sender transmits a series of packet blocks at a frame rate  $R_f$  frames per second:

$$k_1, k_2, \dots, k_i, \dots, k_L, \tag{2}$$

where  $k_i$  is the number of packets of *i*th video frame in the GOP. The frame interval is  $1/R_f$ . Due to the different spatial and temporal activities in the video encoding process, video frames are of variable lengths. According to (1), the available bandwidth allocated to the GOP is given by

$$T_{\rm GOP} = T \times \frac{L}{R_f}.$$
 (3)

In frame-level FEC, the systematic RS erasure code RS  $(n_i, k_i)$  is applied as FEC to protect video frames. Given the target loss probability  $P_{\text{target}}$ , the estimated packet loss rate P and fixed  $k_i$ , the lower bound on  $n_i$  can be computed using

$$P_{\text{target}} = \sum_{j=n_i-k_i+1}^{n_i} {\binom{n_i}{j}} p^j (1-p)^{n_i-j}.$$
 (4)

According to (4), the FEC encoder generates  $h_i = n_i - k_i$  redundant packets for the *i*th video frame and the number of packets transmitted to the network is  $n_i$ . In the FEC decoder, the *i*th video frame can be reconstructed when any  $k_i$  or more packets out of  $n_i$  transport packets are successfully received. In [34], the picture quality of MPEG-4 is showed to be acceptable at a loss rate of  $10^{-5}$  and good at a loss rate of  $10^{-6}$ . In this paper, the value of  $P_{\text{target}}$  in (4) was set to  $10^{-6}$  to obtain high-visual-quality video experience.

Figure 5(a) illustrates the processing sequence of source and redundant packets for the instant frame-level FEC. Following the frame *i* of  $k_i$  source packets, the desired redundant packets  $h_i$  are generated and transmitted along with the source packets to combat packet loss in the network. The processing sequences of optimal frame-level FEC and GOPlevel FEC are showed in Figures 5(b) and 5(c), respectively. Both techniques need to defer the FEC processing until an entire GOP arrives. The difference between them is that the optimal frame-level FEC adjusts the amount of FEC per frame while the GOP-level FEC determines an appropriate amount of FEC for a GOP.

For frame *i*, the video data bandwidth  $B_{\text{data}}(i)$  and the required FEC bandwidth  $B_{\text{req}}(i)$  to achieve the target packet loss probability can be easily computed as follows:

$$B_{\text{data}}(i) = k_i \times S, \tag{5}$$

$$B_{\rm req}(i) = n_i \times S. \tag{6}$$

Therefore, the total bandwidth that the frame-level FEC can allocate to the GOP must satisfy the following constraint:

$$\sum_{i=1}^{L} B_{\text{req}}(i) \le T_{\text{GOP}}.$$
(7)



FIGURE 5: Sequence of source and FEC redundant packets: (a) instant frame-level FEC; (b) optimal frame-level FEC; (c) GOP-level FEC.

To cater for this rate constraint problem, the temporal scaling approach can be used to adjust the amount of video data while preserving the real-time requirement of streaming video applications. In the temporal scaling approach, the video frames with less importance level are discarded before transmission to match the available TCP-friendly transmission rate. For instance, the frame discarding order can be B, P, and I frame.

#### 4.2. Classification of video frames

The GOP pattern in MPEG video is typically arranged as follows:

$$IB_{0,0} \cdots B_{0,N_{BP}-1}P_1 \cdots P_m B_{m,0} \cdots B_{m,N_{BP}-1}P_{m+1} \\ \cdots P_{N_P} B_{N_P,0} \cdots B_{N_P,N_{BP}-1},$$
(8)

where  $N_{\rm P}$  is the number of P frames in the GOP and  $N_{\rm BP}$ is the number of B frames in between an I and a P frame or two P frames. Therefore, the number of B frames  $N_{\rm B}$ in the GOP is given by  $N_{\rm B} = (1 + N_{\rm P}) \times N_{\rm BP}$ . Generally, I frame is encoded with high-spatial quality in the GOP and the subsequent P frames have a gradually degraded spatial quality. Accordingly, the frame type and the frame distance from the reference I frame are two basic criteria to classify video frames in the GOP. Since losing I frame can cause a significant impact on video quality for the entire GOP, I frame has a highest priority. Due to the temporal dependency, P frames which are closer to the reference I frame has higher priority. As to B frames, which are not used as references of other frames in the GOP, the temporal quality degradation caused by continuous B-frame loss should be considered in classifying video frames. According to the distance of B frames from the reference I frame, B frames are evenly chosen to form  $N_{\rm BP}$  frame groups of size  $N_{\rm P}$  as follows:

$$\{B_{0,0}B_{1,0}\cdots B_{N_{P},0}\}$$

$$\{B_{0,1}B_{1,1}\cdots B_{N_{P},1}\}$$

$$\cdots$$

$$\{B_{0,N_{BP}-1}B_{1,N_{BP}-1}\cdots B_{N_{P},N_{BP}-1}\}.$$

$$(9)$$

In each B-frame group, B frames which are closer to the reference I frame have higher priority. Figure 6 shows the



FIGURE 6: Video frame classification within a GOP.

classification of video frames in this study. Therefore, the frame priority sequence in the GOP can be given by

$$\frac{IP_1P_2\cdots P_{N_P}B_{0,0}B_{1,0}\cdots B_{N_P,0}B_{0,1}\cdots B_{N_P,1}}{\cdots B_{0,N_{RP}-1}\cdots B_{N_P,N_{RP}-1}}.$$
(10)

Based on this prioritized frame sequence, each frame in the GOP pattern is associated with a priority distance from the leading I frame. For instance, with the GOP pattern of " $IB_{00}B_{01}P_1B_{10}B_{11}P_2B_{20}B_{21}P_3B_{30}B_{31}$ ", its corresponding priority sequence is " $I_1P_1P_2P_3B_{00}B_{10}B_{20}B_{30}B_{01}B_{11}B_{21}B_{31}$ " and thus the associated distance sequence of the GOP pattern can be represented as "0 4 8 1 5 9 2 6 10 3 7 11".

#### 4.3. Adaptive frame rate allocation

The proposed FEC technique aims at minimizing the additional FEC processing delay while achieving the near-highest video quality obtained by the optimal frame-level FEC within TCP-friendly rate constraints. To achieve this goal, the FEC performs an online rate allocation for video frames and allows frame discarding to adjust the output data rate without impairing the timing constraints of streaming video. Upon receiving the frame *i*, the required bandwidth  $n_i$  of the frame is calculated using (4) to achieve the target packet loss probability. Then, the FEC allocates the bandwidth of the frame *i*,  $B_{\text{frame}}(i)$ , subject to the transmission rate budget and the frame priorities. After the successful frame bandwidth allocation, the frame *i* is delivered to the network with its redundancy generated by the FEC. Based on the video frame classification described in the previous section, the bandwidth allocated to the video frame should be adaptive to its priority distance from the reference I frame. This adaptive frame rate allocation can be done using a linear weighting factor related to the priority distance of the video frame. Figure 7 illustrates the weighted rate allocation of video frames. Denoting N as the maximum priority distance in the GOP pattern and  $d_i$  as the priority distance between the frame i and the reference I frame, the corresponding weight  $F_i$  of the frame i is defined as follows:

$$F_i = \frac{N - d_i}{N}.\tag{11}$$

Using (11) and summing up the weights of all frames, the allocation weight for the GOP can be obtained by

$$F_{\rm GOP} = \sum_{i=1}^{N} F_{i.}$$
 (12)

Since I frame is the first encoded and delivered frame in the GOP, the rate allocation starts by allocating enough FEC bandwidth ( $n_1 \times S$ ), to I frame using (4). After I-frame rate allocation, the available bandwidth for the subsequent frames can be obtained:

$$T_{\text{avail}} = T_{\text{GOP}} - n_1 \times S. \tag{13}$$

Then, the weighted frame bandwidth of the frame i in the GOP is calculated by

$$B_{\text{weight}}(i) = \frac{F_i}{F_{\text{GOP}}} \times T_{\text{avail}}.$$
 (14)

Noted that the frame can be contiguously discarded when the available TCP-friendly transmission rate is low or when fast motion change occurs to induce larger frame size. For the case of contiguous frame discarding, the available bandwidth budget for the next frame accumulates as the number of contiguous discarded frames increases. Let *m* be the number of frames discarded contiguously before the frame *i*, and let  $B_{\text{weight}}(i, m)$  be the weighted frame bandwidth accumulated from the frame (*i*-*m*) to the frame *i*, (14) can be modified as follows:

$$B_{\text{weight}}(i,m) = \sum_{j=i-m}^{i} \frac{F_j}{F_{\text{GOP}}} \times T_{\text{avail}}.$$
 (15)

Furthermore, to better utilize the transmission rate budget, the remaining frame bandwidth of the frame  $i, B_{rem}(i)$ , is also added to the bandwidth budget of the frame (i + 1). Therefore, the available bandwidth for the frame *i* is given by

$$B_{\text{avail}}(i) = B_{\text{weight}}(i, m) + B_{\text{rem}}(i-1).$$
(16)

In allocating the frame bandwidth to video frames, we employ the following set of strategies with respect to the type of the frame *i*.





- (i) For the B frame, if the available bandwidth  $B_{\text{avail}}(i)$  is less than the required bandwidth  $B_{\text{req}}(i)$ , the frame is discarded and accordingly, its available bandwidth will be added to the bandwidth budget of the next frame.
- (ii) For the P frame, if the available bandwidth  $B_{\text{avail}}(i)$  is less than the required bandwidth  $B_{\text{req}}(i)$ , the frame will borrow from the remaining transmission rate budget,  $T_{\text{rem}}$ , until either the required bandwidth is met or the transmission rate budget is exhausted since the loss of one P frame can cause the severe quality degradation of other P and B frames. The P frame is discarded when the remaining transmission rate budget is less than the amount of source data  $B_{\text{data}}(i)$ .
- (iii) For the I frame, the I frame is never discarded since the loss of I frame causes the coding failure of an entire GOP.

The adaptive frame rate allocation algorithm used in the instant frame-level FEC is described in Algorithm 1.

# 5. PERFORMANCE RESULTS

#### 5.1. Experimental setup

To evaluate the performance of the proposed FEC-on-TQM, a prototype implementation of TQM was developed on the FreeBSD and Linux platforms.As shown in Figure 8, the experimental setup consisted of a video sender, a video receiver and a network bridge, and these three hosts were running on Linux-based x86 PCs. The network bridge bridged packets between networks and produced packet losses and transfer delay for a specified data flow. In the video server, MPEG video sequences were transmitted to the video receiver using the VideoLAN Server (VLS) [35]. FEC-on-TQM EMs were constructed in the video sender and the video receiver, respectively, and an omniscient scheme was built to provide an end-to-end LDA function [36]. The wireless network employed an 802.11b access point i = 1; //begin calculation from the first frame in the video stream While (not end of stream){ Calculate  $B_{req}(i)$  using (4) and (6); **Switch** (type of frame *i*){ Case "I": Calculate  $T_{\text{GOP}}$  using (1) and (3);  $T_{\rm rem} = T_{\rm GOP}; B_{\rm rem}(i) = m = 0;$  $B_{\text{frame}}(i) = \min\{T_{\text{rem}}, B_{\text{reg}}(i)\};$ Break; Case "P": Calculate  $B_{\text{avail}}(i)$  using (16); If  $(T_{\text{rem}} < B_{\text{data}}(i))$ {  $B_{\text{frame}}(i) = 0$ ; // Frame *i* is discarded m = m + 1;} Else {  $B_{\text{frame}}(i) = \min\{T_{\text{rem}}, B_{\text{req}}(i)\};$ m = 0; $B_{\text{rem}}(i) = B_{\text{rem}}(i-1) + B_{\text{avail}}(i) - B_{\text{frame}}(i);$ Break; Case "B": Calculate  $B_{\text{avail}}(i)$  using (16); If  $(B_{req}(i) > B_{avail}(i))$ {  $B_{\text{frame}}(i) = 0$ ; // Frame *i* is discarded m = m + 1;} Else {  $B_{\text{frame}}(i) = B_{\text{req}}(i)$ m = 0;}  $B_{\text{rem}}(i) = B_{\text{rem}}(i-1) + B_{\text{avail}}(i) - B_{\text{frame}}(i);$ Break; If  $(B_{\text{frame}}(i) > B_{\text{data}}(i))$ FEC is deployed for frame *i*, the FEC bandwidth allocated to frame *i* is  $B_{\text{frame}}(i)$ ;  $T_{\rm rem} = T_{\rm rem} - B_{\rm frame}(i);$ i = i + 1;}

ALGORITHM 1: Adaptive rate allocation algorithm.



FIGURE 8: Experimental setup.

(AP) operating in distributed coordination function (DCF) mode to connect the video receiver. The video receiver was arranged in clear line of sight (LoS) of the AP to ensure the channel quality and generated packet losses caused by wireless bit errors.

# 5.2. TQM overhead

As described in Section 2, TQM adopts a flow redirection approach to support FEC functionality for legacy applications. In this approach, packets are copied from the kernel space to the user space and are then sent back to the kernel space following processing. Clearly, this processing overhead introduces the additional delays at the sender and the receiver end. In multimedia applications, delay is a major factor influencing the perceived media quality. To evaluate the overhead incurred by the TQM scheme, an experiment was performed to determine the average delay time incurred by the TQM flow redirection process for packets of various sizes. In the experiment, the video sender sent ICMP ping packets to the video receiver, and then waited for the echo packets to be returned. Each echo packet was redirected to the TQM module, processed, and then injected to protocol stack for transmission. The RTT was then measured at the video sender end in order to calculate the difference between the TQM-redirected RTT and the non-TQM-redirected RTT. The corresponding results are presented in Table 1 for ICMP packets of four different sizes. It can be seen that the difference between the directed and nondirected RTTs increases with an increasing packet size. However, the difference between the two times is very small. For a 1024-byte ICMP packet, which is around the typical packet size in most video applications, the RTT increases by only 0.276% when the flow is redirected to TQM. In other words, compared to the total transmission and processing delays incurred along the path between the sender and the receiver, the TQM overhead is relatively minor.

A second experiment was then performed to investigate whether the additional delays caused by TQM and FEC-on-TQM would accumulate to such an extent that they affected the overall playout time of video applications. For FECon-TQM, the frame type extraction time, the FEC encoding/decoding time, and the redirection delay are considered to measure the cumulative delay. Using the same setup as that shown in Figure 8, long-lived video sequences were sent from the video sender to the video receiver. The number of FEC redundant packets was set to a static value of 1 for each video frame. Since FEC decoding is not triggered in the TQM receiver under loss-free conditions, an assumption was made that one transported packet per FEC block was lost during transmission. The total playout time was estimated at the receiver and was then compared with the playout time with no flow redirection or cumulative delays. Table 2 presents the results obtained for eight different video sequences. It is observed that the TQM redirection delay does not exceed 90 milliseconds for any of the sequences. Furthermore, for the case of FEC-on-TQM, it can be seen that the FEC coding operations result in an additional cumulative delay of approximately 3-4 milliseconds. In practice, however, this additional delay can be easily absorbed by the application buffer and will have no effect on the user's perception of the delivered video quality. In the current experimental setup, both the video sender and the video receiver are fitted with Pentium-4 1.8 G processors and 256 MB of memory. The experimental results suggest that this configuration is sufficient to support the proposed TQM.

#### 5.3. The FEC-on-TQM performance

To facilitate a controlled environment for performance measurement purpose, the video application at the sender is replaced by an emulator that feeds the frame of stored MPEG-4 video clips at the real-time frame rate to the video receiver. The rest of experimental set-up keeps unchanged. The video sequence "Foreman" of CIF format is used in this experiment. The videoclip is transmitted at a frame rate 25 frames per second and the GOP size is 9. In general, video frames are segmented into packets for transmission. Therefore, the application-level quality depends not only on the successful recovery of lost packets, but also on the dependency relations and delay constraint between the individual video frames. A frame is considered to be decodable at the video receiver when at least a fixed portion DT (decodable threshold) of the data in each frame is received in time and all of the frames it depends on are also decodable. In the experiment, the stream is transmitted in packets of 1Kbytes and the value of DT is set to 1 for all video frames to observe error performance by dismissing the effect of error resilience features of video transport. The



FIGURE 9: Gilbert model.

video receiver processes the packet arrival in accordance with the hard real-time constraint model presented in [10]. Also, we implement three FEC schemes including the instant frame-level FEC, the optimal frame-level FEC and the GOPlevel FEC scheme on FEC-on-TQM. For all three schemes, TCP-friendly transmission rate is calculated for every GOP. As in [20],  $t_{\text{RTO}}$  is typically set to four times of the round trip time *r*. In order to match the TCP-friendly sending rate with real-time constraints, both the optimal frame-level scheme and the GOP-level scheme employ the temporal scaling technique to discard frames based on the video frame priority described in Section 4.2.

In this evaluation, either congestion loss or wireless loss causes the end-to-end packet loss. A Gilbert model or a twostate Markov model is used to generate burst loss patterns over the transmission channel (Figure 9). The two states of the model are Good-state and Bad-state. In the network bridge, we employ a packet-level Gilbert model to drop packets with the parameters of average packet loss rate ( $P_B$ ) and average burst packet loss length ( $L_B$ ). In Good-state, a packet is dropped with the probability of 0, while in Bad-state, a packet is dropped with the probability of 1. The transition probabilities of  $P_{gb}$  and  $P_{bg}$  can be derived by the values of  $P_B$  and  $L_B$  as follows:

$$P_{\rm bg} = \frac{1}{L_B}, \qquad P_{\rm gb} = P_{\rm bg} \times \frac{P_B}{1 - P_B}.$$
 (17)

In the video receiver, on the other hand, a bit-level Gilbert model is used to model packet loss due to bit errors in the wireless channel. This bit-level model is described by an average bit error rate ( $P_b$ ) and an average burst bit error length ( $L_b$ ). A packet of size  $\tau$  in bits is dropped when at least one bit within  $\tau$  bits is lost. Table 3 presents the system parameter settings for the network.

To verify the performance of FEC schemes designed for delay-sensitive video service, we compare the end-to-end frame delay among three FEC schemes built on TQM in different network conditions. The end-to-end frame delay is defined as the time difference between the frame release time in the video sender application and the arrival time of the similar frame in the video receiver application. When the frames are discarded by FEC schemes or are lost during transmission, they are excluded from the calculation of end-to-end delay. Figure 10 shows the end-to-end frame delay analysis of three FEC schemes with  $P_B = 0.01$  and  $P_b = 10^{-6}$ . It is noted that the instant frame-level scheme performs the FEC encoding process on a frame-by-frame basis, while both the optimal frame-level and the GOP-level scheme

ICMP packet size (bytes)	Non-TQM- redirected RTT (ms)	TQM-redirected RTT (ms)	RTT difference	RTT difference/ non-TQM-redirected RTT
56	0.975	0.981	0.006	0.615%
512	2.579	2.587	0.008	0.310%
1024	4.344	4.356	0.012	0.276%
2048	6.849	6.864	0.015	0.219%

TABLE 1: Flow redirection overhead.

TABLE 2: Cumulative delay of FEC-on-TQM scheme.

Video sequence	Total length	Redirection delay (ms)	Redirection + FEC coding delay (ms)
Cars	56 min	86.39	89.68
X-Men	55 min, 25 s	89.25	92.16
Mission impossible	63 min, 6 s	85.55	88.85
Brokeback mountain	65 min, 39 s	83.66	86.63
Final destination	45 min, 50 s	88.21	91.85
The Da Vinci code	76 min, 46 s	85.36	88.06
Monster house	45 min, 6 s	87.85	90.98
World Trade Center	53 min, 14 s	86.68	90.16

TABLE 3: Network settings.

Parameters	Scenario 1	Scenario 2
Delay	12.5 ms	12.5 ms
$P_B$	0.01	0.01, 0.02,, 0.1
$L_B$	3	3
$P_b$	$10^{-6}, 10^{-5}, \ldots, 10^{-1}$	$10^{-3}$
L <sub>b</sub>	80	80

defers the encoding process until all video frames within the GOP are received. From Figure 10, we can observe that: (1) the instant frame-level scheme has much smaller endto-end frame delay than other two schemes and the delay values are below 40 milliseconds (i.e., frame interval); (2) for the optimal frame-level or the GOP-level scheme, the GOP buffering results in a sawtooth delay curve and peaks in the curve are correlated with the periodic insertion of I frames; and (3) the end-to-end frame delays of the GOP-level scheme are slightly smaller than that of the optimal framelevel scheme, since the GOP-level scheme performs the FEC encoding on a GOP and this requires less redundancies for network transmission to protect the entire GOP against the fixed network packet loss.

In the first experimental scenario, the bit error rate is varied from  $10^{-6}$  to  $10^{-1}$  as the packet loss rate is fixed to 0.01. Figures 11 and 12 show the average end-to-end frame delay and the average PSNR, respectively. It is noted that for the bit error rate =  $10^{-1}$ , all packets are lost and none of video frames is available to measure the end-to-end frame delay. From Figures 11 and 12, it can be seen that for the instant frame-level scheme: (1) the average end-to-end frame delay is below 40 milliseconds as the bit error rate varies, and (2) the average PSNR curve is close to that of the optimal frame-level scheme and the largest PSNR



FIGURE 10: End-to-end timing between frames 1–100 with parameters  $P_B = 0.01$ ,  $P_b = 10^{-6}$ .

difference of 0.52 dB occurs as the bit error rate is  $10^{-2}$ . We also observe that compared to the GOP-level scheme, the optimal frame-level scheme has lower average PSNR value with the larger end-to-end frame delay. Figures 13 and 14 show the performance results for another scenario. As the packet loss rate is increased, the TCP-friendly sending rate decreases accordingly. Besides the similar observations as in the first scenario, the difference in the average PSNR value between the instant frame-level scheme and the optimal frame-level scheme is ranged from 0.01 dB to 0.71 dB. In Figure 13, the delay gap between the optimal frame-level scheme and the GOP-level scheme becomes large as the packet loss rate exceeds 0.01. This is because that the amount of discarded frames increases in the optimal frame-level scheme to match the low TCP-friendly sending rate and the frames with high priority, such as I frames, usually



FIGURE 11: Comparison of end-to-end frame delays with varied bit error rate.



FIGURE 12: PSNR comparison with varied bit error rate.

gains transmission opportunities to cause a large end-toend frame delay. To conclude, the experimental results show that the instant frame-level scheme better preserves the timing aspects of real-time streaming video while achieving the near-highest video quality that the optimal frame-level scheme can obtain within the TCP-friendly rate constraints. Therefore, the proposed scheme transparently supports the robust video transmission on behalf of video applications and is suitable for the provision of high quality real-time video streaming with low delay.

# 6. CONCLUSIONS

This paper has developed a transparent QoS mechanism designated as TQM to transparently establish QoS sessions on behalf of multimedia applications over the wireless



FIGURE 13: Comparison of end-to-end frame delays with varied packet loss rate.



FIGURE 14: PSNR comparison with varied packet loss rate.

Internet. With no modification required to the existing legacy applications, TQM uses an efficient packet redirection scheme to divert user-specified flows for QoS enhancement processing and provides a suitable platform for the implementation of different QoS EMs. In TQM, EMs may function either independently or with other EMs to improve the overall QoS enhancement. As a first step, this study has discussed the implementation involved in integrating available EMs to support the transparent loss recovery based on FEC. The FEC-on-TQM module uses an instant framelevel FEC technique to achieve the near-highest video quality that the optimal frame-level scheme can obtain within the TCP-friendly rate constraints for real-time streaming video applications. The experimental results have shown that: (1) the overheads incurred by the proposed TQM scheme are minor compared to the total transmission and processing

delays incurred along the path between the sender and the receiver, and (2) FEC-on-TQM successfully integrates several EMs to carry out the proposed instant frame-level FEC technique, resulting in a low-delay, good-visual-quality video experience. Future studies include the testing of network-adaptive video applications, such as Helix video server, on FEC-on-TQM, and the integration of EMs to further improve the overall QoS of diverse multimedia applications over the wireless Internet.

### REFERENCES

- J. Wroclawski, The use of RSVP with IETF Integrated Services, RFC 2210, September 1997.
- [2] S. Black, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," RFC 2475, December 1998.
- [3] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless Internet," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 123–133, 2005.
- [4] Quality of service (QoS) concept and architecture, 3GPP TS 23.107, September 2003.
- [5] IEEE Standard, "Local and metropolitan area networks specific requirements part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 8: medium access control (MAC) quality of Service enhancements," pp.1–189, 2005.
- [6] IEEE Standard, "802.16-2004 part 16: air interface for fixed broadband wireless access systems," October 2004.
- [7] K. Nahrstedt, D. Xu, D. Wichadakul, and B. Li, "QoS-aware middleware for ubiquitous and heterogeneous environments," *IEEE Communications Magazine*, vol. 39, no. 11, pp. 140–148, 2001.
- [8] Y.-B. Miao, W.-S. Hwang, and C.-K. Shieh, "A transparent deployment method of RSVP-aware applications on UNIX," *Computer Networks*, vol. 40, no. 1, pp. 45–56, 2002.
- [9] L. Munoz, M. Garcia, J. Choque, R. Aguero, and P. Mahonen, "Optimizing Internet flows over IEEE 802.11b wireless local area networks: a performance-enhancing proxy based on forward error correction," *IEEE Communications Magazine*, vol. 39, no. 12, pp. 60–67, 2001.
- [10] K. Park and W. Wang, "QoS-sensitive transport of real-time MPEG video using adaptive redundancy control," *Computer Communications*, vol. 24, no. 1, pp. 78–92, 2001.
- [11] O. Ait-Hellal, E. Altman, A. Jean-Marie, and I. A. Kurkova, "On loss probabilities in presence of redundant packets and several traffic sources," *Performance Evaluation*, vol. 36-37, pp. 485–518, 1999.
- [12] K. Park and W. Willinger, Self-Similar Network Traffic and Performance Evaluation, Wiley-Interscience, New York, NY, USA, 2000.
- [13] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the Internet," *IEEE/ACM Transactions on Networking*, vol. 7, no. 4, pp. 458–472, 1999.
- [14] H. Wu, M. Claypool, and R. Kinicki, "Adjusting forward error correction with temporal scaling for TCP-friendly streaming MPEG," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 1, no. 4, pp. 315–337, 2005.
- [15] Y. Yuan, B. F. Cockburn, T. Sikora, and M. Mandal, "Efficient allocation of packet-level forward error correction in video streaming over the Internet," *Journal of Electronic Imaging*, vol. 16, no. 2, Article ID 023012, 12 pages, 2007.

- [16] M. Baldi and Y. Ofek, "End-to-end delay analysis of videoconferencing over packet-switched networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 4, pp. 479–492, 2000.
- [17] Linux IP firewall HOWTO, http://www.linuxfaq.com /LDP/ HOWTO/Firewall-HOWTO.html.
- [18] Divert sockets, http://tldp.org/HOWTO/Divert- Sockets-mini-HOWTO.html#toc6.
- [19] R. Rejaie, M. Handley, and D. Estrin, "Quality adaptation for congestion controlled video playback over the Internet," ACM SIGCOMM Computer Communication Review, vol. 29, no. 4, pp. 189–200, 1999.
- [20] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equationbased congestion control for unicast applications," in *Proceedings of the ACM Annual Conference of the Special Interest Group on Data Communication (SIGCOMM '00)*, pp. 43–56, Stockholm, Sweden, August-September 2000.
- [21] M. Handley, S. Floyd, J. Pahdye, and J. Widmer, "TCP friendly rate control (TFRC): protocol specification," RFC3448, January 2003.
- [22] D. Wu, Y. T. Hou, W. Zhu, et al., "On end-to-end architecture for transporting MPEG-4 video over the Internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 923–941, 2000.
- [23] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Resource allocation for multimedia streaming over the Internet," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 339–355, 2001.
- [24] Z. Wang, S. Banerjee, and S. Jamin, "Media-friendliness of a slowly-responsive congestion control protocol," in *Proceedings* of the 14th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '04), pp. 82–87, Cork, Ireland, June 2004.
- [25] V. Samata, R. Olivera, A. Dixit, P. Aghera, P. Zerfos, and S. Lu, "Impact of video transcoding parameters on dynamic video transcoding," in *Proceedings of the 1st International Conference* on Communication System Software and Middleware (COM-SWARE '06), New Delhi, India, January 2006.
- [26] V. Arya and T. Turletti, "Accurate and explicit differentiation of wireless and congestion losses," in *Proceedings of the 23rd International Conference on Distributed Computing Systems* (*ICDCS* '03), Providence, RI, USA, May 2003.
- [27] S. Biaz and N. Vaidya, "Discriminating congestion losses from wireless losses using inter-arrival times at the receiver," in *Proceedings of the IEEE Symposium on Application-Specific Systems and Software Engineer (ASSET '99)*, pp. 10–17, Richardson, Tex, USA, March 1999.
- [28] Y. Tobe, Y. Tamura, A. Molano, S. Ghosh, and H. Tokuda, "Achieving moderate fairness for UDP flows by path-status classification," in *Proceeding of the 25th Annual IEEE Conference on Local Computer Networks (LCN '00)*, pp. 252–261, Tampa, Fla, USA, November 2000.
- [29] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," RFC 3550, July 2003.
- [30] F. Borgonovo and A. Capone, "Efficiency of error-control schemes for real-time wireless applications on the Gilbert channel," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 1, pp. 246–258, 2005.
- [31] A. Li, "RTP payload format for generic forward error correction," Internet Draft draft-ietf-avt-ulp-18, IETF, work in progress, June 2006.
- [32] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," RFC 2733, December 1999.

- [33] J. Padhye, V. Firoiu, D. F. Towsley, and J. F. Kurose, "Modeling TCP reno performance: a simple model and its empirical validation," *IEEE/ACM Transactions on Networking*, vol. 8, no. 2, pp. 133–145, 2000.
- [34] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis, and B. Basch, "Robust compression and transmission of MPEG-4 video," in *Proceedings of the 7th ACM International Conference on Multimedia (MULTIMEDIA '99)*, pp. 113–120, Orlando, Fla, USA, October 1999.
- [35] The VideoLAN server, http://www.videolan.org/vlc/streaming .html.
- [36] S. Cen, P. C. Cosman, and G. M. Voelker, "End-to-end differentiation of congestion and wireless losses," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 703–717, 2003.

# Research Article

# Power-Constrained Fuzzy Logic Control of Video Streaming over a Wireless Interconnect

### Rouzbeh Razavi, Martin Fleury, and Mohammed Ghanbari

Department of Computing and Electronic Systems, University of Essex, Colchester CO4 3SQ, UK

Correspondence should be addressed to Martin Fleury, fleum@essex.ac.uk

Received 29 September 2007; Accepted 6 May 2008

Recommended by David Bull

Wireless communication of video, with Bluetooth as an example, represents a compromise between channel conditions, display and decode deadlines, and energy constraints. This paper proposes fuzzy logic control (FLC) of automatic repeat request (ARQ) as a way of reconciling these factors, with a 40% saving in power in the worst channel conditions from economizing on transmissions when channel errors occur. Whatever the channel conditions are, FLC is shown to outperform the default Bluetooth scheme and an alternative Bluetooth-adaptive ARQ scheme in terms of reduced packet loss and delay, as well as improved video quality.

Copyright © 2008 Rouzbeh Razavi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. INTRODUCTION

Preservation of battery power is an essential feature of mobile devices, to reduce the frequency of recharges. Though Bluetooth (IEEE 802.15.1) [1] devices have hold, park, and sniff low activity modes, and the transceiver is designed to minimize power [2], it is still important that an application reduces the total data transmitted, as there is approximately a linear relationship [3, 4] between bit rate and energy consumption. A number of authors, for example [4–8], have investigated ways to manage power in a wireless network when streaming video. Although the enhanced data rate (EDR) of Bluetooth version 2.0 [9] now has a peak user payload of 2.2 Mbps (gross air rate 3.0 Mbps), which is the same average rate offered by some implementations of IP-TV, it must still compete with lower power alternatives, such as Wibree from Nokia, intended for button cell batteries, with a gross air rate of 1.0 Mbps. However, compared to IEEE 802.11 (Wi-Fi)'s [10] typical current usage of 100-350 mA, Bluetooth's consumption is 1-35 mA, implying that for mobile multimedia applications with higher bandwidth capacity requirements, Bluetooth is a preferred solution.

Many cellular phones are also equipped with a Bluetooth transceiver and larger resolution screens of CIF ( $352 \times 288$ ) and QCIF ( $176 \times 144$ ) pixel size. However, as in a group of

pictures (GOP), slices within one picture are predicted from previous ones, noise and interference on the wireless channel may corrupt slice-bearing packets, as they make the final hop before decoding and display on a mobile device. This suggests retransmission of corrupted packets should occur, which automatically increases the power budget, quite aside from the possibility for video of missed display deadlines. This is unfortunate, as in general automatic repeat request (ARQ) has proved more effective than forward error correction (FEC) [11] in ensuring statistically guaranteed qualityof-service (QoS) over wireless networks. FEC imposes an ongoing overhead, adding to the power budget, whereas typical channel errors come in bursts, with the channel state alternating between good and bad states. For example, in an indoor environment, fast fading occurs when persons walk across the line-of-sight between the communicating devices. Hybrid ARQ [12], in which reply packets advise the sender of errors, is complex to implement at the data link layer and, owing to the volatility of the wireless channel, may impose too great a latency if adaptive error control occurs at the application layer, at a remote encoder. In Bluetooth, fast ARQ comes for free by virtue of time-division duplex (TDD) polling, which is necessary for transmit/receive recovery, allowing a single-chip implementation, whereas data link layer FEC is only possible at the legacy basic rate (1.0 Mbps gross air rate).

Effective ARQ management is the key to both power management and ensuring acceptable video quality at the receiver device. However, it is a multifaceted control problem, as account must also be taken of wireless channel conditions, and of the display/decode deadlines of the picture type slices being conveyed. This paper proposes fuzzy logic control (FLC) of ARQ, as a way of combining all three factors: (1) channel state; (2) display/decode deadline; and (3) power budget. In our earlier work [13], we did not consider the need to meet a power budget. We have adopted a modular scheme whereby a two-input FLC stage with a single output is concatenated with a second FLC stage, with the output from the original FLC and an additional "remaining power" input. The two inputs to the first FLC stage are buffer fullness and the deadline margin of the packet at the head of the Bluetooth send queue, which gives a direct measure of delay. Assuming a fixed power budget for the duration of a video clip streaming session, the declining power budget as the stream progresses has the effect of modulating the ARQ retransmission count. A modular scheme reduces the construction complexity of the design and allows for future enhancements.

FLC, which has from its inception [14] been extensively used for industrial and commercial control applications [15], is a convenient tool for real-time control as unlike genetic algorithms or neural networks there is no long period of convergence or online training. Two factors imply that a mathematical model is unsuitable: the inputs are dependent on the outputs as there are feedback channels, implying that the problem is nonlinear; and the complexity of multiple constraints is an obstacle. Within video coding, FLC has already found an application [16, 17] in maintaining a constant video rate by varying the encoder quantization parameter according to the output buffer state, which is a complex control problem without an analytical solution. Therefore, FLC is a natural candidate for the solution of this problem. In general, a fuzzy scheme is easily tuned by adjustment of its membership functions. A fuzzy scheme is also well suited to the implementation on a mobile device, because not only are the decision calculations inherently simple (and can be made more so by adoption of triangular membership functions) but also, by forming a look-up table (LUT) from the fuzzy control surface, its operation can be reduced to simple LUT access. There is also a range of hardware designs [18] for FLC to aid real-time operation.

As is well known, real-time delivery of video is delaysensitive, as a frame cannot be displayed if its data arrive after their decode deadline. A further deadline exists for reference picture types if their presence contributes to decoding of future frames [19]. In practice, a playout buffer exists on a mobile device to account for startup delay and also absorbs delay jitter (variation of delay). Therefore, the maximum delay permissible corresponds to the startup delay deemed tolerable to the user. Packets may arrive too late for the frame to be displayed and, as error concealment at the decoder is implementation dependent, the net result is poor quality video. Not only do packets arrive after their display deadline, but while retransmission takes place, other packets may either wait too long in the send buffer or in the extreme case arriving packets may find the send buffer full. ARQ adds to delay and, therefore, the number of retransmissions should be minimized even before taking into account the impact on the power budget.

Adaptive ARQ is not a complete solution, as it fails to account for deadline expired packets remaining in the send buffer while retransmission takes place. The danger is that these packets will then be transmitted simply to be discarded at the receiver. The presence of expired packets in the send buffer, just like excessive ARQ delay, contributes to the queuing delay of other packets and possibly to buffer overflow. Therefore, an active discard policy for deadline expired packets is required as an addition to adaptive ARQ. In our system, the active discard policy is implemented as a deadline-aware buffer (DAB) and is also based on picture type. Picture type can be ascertained by inspection of application packet headers, whereas accounting for picture content rather than picture importance may require intervention at a source encoder. The DAB introduced by us has a threefold advantage: (1) queuing time of packets in the send buffer is reduced; (2) the possibility of send buffer overflow is effectively removed, except for the smallest of buffer sizes; and (3) power is conserved as deadline expired packets are no longer needlessly transmitted.

The remainder of this paper is organized as follows. Section 2 is a survey of related work, with a concentration on power-aware video streaming. Section 3 contributes background material on Bluetooth and explains the FLC in detail. The research methodology is also detailed. Section 4 contains our simulated results, while Section 5 summarizes and draws some general conclusions.

# 2. RELATED WORK

In [20], it was shown that transmission accounts for more than a third of the total energy consumption in communication on a mobile device. In [3], 78% of power consumption is attributed to transmission and playback at the receiver. In general, transmission consumes more power than reception, but this does not necessarily imply that in Bluetooth a master consumes more power than a slave receiver, because a receiver is unable precisely to anticipate when a transmission will occur. Thus a Bluetooth slave receiver on average consumes 46 mA, [21] as opposed to a master transmitter's 17 mA consumption.

In [4], assuming the aforementioned linear relationship between energy consumption and bit rate, within a GOP, Bpictures are first discarded, while if this does not succeed in reducing the bit rate then P and even I pictures are discarded. The authors propose spreading the discards to allow easier reconstruction at the decoder. However, this is an early work that gives no account of the impact on video quality of this rather simple policy. In [5, 6], the decoding capability of the receiver is signaled to the transmitter, which subsequently adjusts its transmission accordingly through fine-grained scalability. The transmitter encoder power budget is taken into account in [22], varying the power allocation between source and channel coding. However, the former approach apparently does not consider the effect of the channel,
whereas the latter is inappropriate for preencoded video. A transcoder at the wireless transmitter is assumed in [3] and the rate is controlled according to a linear model of power consumption, together with a piecewise linear model of playback power consumption. In [23], an energy constraint is introduced into a rate-distortion encoding model. In [24] also, content importance is factored in by annotating video segments through MPEG-7. Moderate improvements in user perception were reported. Despite the title in [8], the video content itself does not determine the transmit rate so much as the length of MPEG (*sic*) packets. The lengths are used to determine a packet burst profile for IEEE 802.11 networks. Depending on the video clip, approximately 60% energy savings are reported for this technique.

Our scheme considers a fixed playout buffer at the receiver and assumes single-layered video. Fixed-size playout buffers at the receiver are liable to underflow given that variable bit rate (VBR) encoded video is inherently "bursty." The burstiness occurs at multiple time scales, owing to changes in picture type within a GOP, within a scene with variable motion, and between scene cuts. Though in fixed networks large playout buffers (at up to several seconds of startup delay) may be applied in video-on-demand applications, web-based video clip distribution with click-level interactivity is less tolerant of startup delay. On a mobile device, memory contributes significantly to the power budget [25], resulting in relatively small buffers. For example, the experiments in [26] assumed a send buffer size of fifty packets, as also assumed in our experiments. In [26] also, selected packets are given priority transmission, rather than enforce rate changes at the encoder, which discriminates against preencoded video. However, layered encoding is assumed, while much content exists in nonlayered format.

For single-layer video, the packet type is a simple way of applying either a delay or a loss priority packet transmission. The packet type indicates content importance without the need for content awareness at the link layer. In [27], simple packet type discrimination is proposed as a means of implementing differentiated services QoS on the fixed Internet.

Varying the number of retransmissions as part of ARQ management is a feature of IEEE 802.11 wireless networks and in IEEE 802.11e it is also possible to set a maximum limit to the time spent in the transmitter buffer [28]. In [9], the packet loss rate over the wireless link is balanced with the loss rate from buffer overflow by incremental adjustments to the retry limit. Packet purging is also employed in [9], whereby packets dependent on lost packets are removed from queues. The problem with purging, as opposed to deadline-aware active discard (as in our paper), is that it appears only actionable when I-picture packets have been lost. The scheme in [9] was tested for a six-layered video stream, which increases the time taken in searching queues for packet purging, while the computational cost is less for the single queue nonscaleable video. Both IEEE 802.11's Point Coordination Function and IEEE 802.11e's Hybrid Coordination Function allow for centralized packet scheduling and, hence, techniques applicable to Bluetooth are to some extent transferable to these. IEEE 802.11e has a variable set of ARQ modes but a management policy is not part of the standard.

#### 3. METHODOLOGY

# 3.1. Bluetooth background

Bluetooth is a short-range (less than 10 m for class 2 devices), radio frequency interconnect. Bluetooth employs robust frequency-hopping spread spectrum (FHSS). It also has centralized medium access control through time division multiple access and TDD. These features indicate that Bluetooth is less prone to interference than from other Bluetooth networks. Bluetooth employs variable-sized packets up to a maximum of five frequency-hopping time slots of  $625 \,\mu$ s in duration. Every Bluetooth frame consists of a packet transmitted from a sender node over 1, 3, or 5 timeslots, while a receiver replies with a packet occupying at least one slot, so that each frame has an even number of slots. Therefore, in master to slave transmission, a single slot packet serves for a link layer stop-and-go ARQ message, whenever a corrupted packet payload is detected.

The timeout or retransmission limit value by default is set to an infinite number of retransmissions. On general grounds, this is unwise in conditions of fast fading caused by multipath echoes, as error bursts occur. Another source of error bursts is cochannel interference by other wireless sources, including other Bluetooth piconets, IEEE 802.11b,g networks, cordless phones, and even microwave ovens. Though this has been alleviated to some extent in version 1.2 of Bluetooth by adaptive frequency hopping [29], this is only effective if interference is not across all or most of the 2.402 to 2.480 GHz unlicensed band. However, both IEEE 802.11b and g may occupy a 22 MHz subchannel (with 30 dB energy attenuation over the central frequency at  $\pm 11 \text{ MHz}$ ) within the 2.4 GHz band. Issues of interference might arise in apartment blocks with multiple sources occupying the 2.4 GHz band or when higher power transmission occurs such as at WiFi hotspots.

For Bluetooth, an ARQ may occur in the following circumstances [30]: (a) failure to synchronize on the access header code; (b) header corruption detected by a triple redundancy code; (c) payload corruption detected by cyclic redundancy check; (d) failure to synchronize with the return packet header; and (e) header corruption of the return packet. Notice that a faulty ARQ packet can itself cause retransmission. The main cause of packet error [30], however, is (c) payload corruption, which is the simplified assumption in this paper.

#### 3.2. Analysis of ARQ impact

Given the probability of bit error,  $P_e$ , then  $P_s$ , the probability of a successful packet transmission is defined as

$$P_s = \left(1 - P_e\right)^L,\tag{1}$$

where L is the bit length of a packet. Variations of the following analysis (1) to (5) are well known, occurring,

for example, in [31]. Furthermore, the expected number of retransmissions, *N*, under the default ARQ scheme is

$$E[N] = 0 \times P_s + 1 \times P_s \times (1 - P_s) + 2 \times P_s \times (1 - P_s)^2 + \cdots,$$
$$E[N] = \frac{1 - P_s}{P_s}$$
(2)

which implies that the expected total number of transmissions, E[T], is simply

$$E[T] = E[N] + 1 = \frac{1}{P_s}.$$
(3)

More interestingly, for a maximum number of retransmissions M the expected number of retransmissions is

$$E[N] = P_s \times \sum_{n=1}^{M-1} n \times (1 - P_s)^n + M$$
$$\times \left( 1 - \left[ P_s \times \sum_{n=1}^{M-1} (1 - P_s)^n \right] \right), \tag{4}$$
$$E[N] = \frac{(1 - P_s)(1 - (1 - P_s)^M)}{P_s},$$

and again E[T] = E[N] + 1.

The mean packet departure rate, *S* packet/s, from the Bluetooth send buffer is given by

$$S = \frac{1}{(n+1) \times 625\,\mu s \times E[T]},\tag{5}$$

where *n* is the number of slots occupied by a Bluetooth packet. Assume that packets are fully filled (refer to Section 3.7) and, to find an upper bound on waiting time, that the buffer is fully occupied in a bad state. This means that a simple scaling may be applied to (5) based on the packet bit length. Figure 1 plots packet delay against the probability of a bit error for various retransmission policies. In Figure 1, the buffer size is set to 50 packets, assuming that just one picture type packet, I-picture, is in use. In practice, the buffer will not become fully occupied immediately and the effect of a DAB is to remove packets from the buffer but the plots in Figure 1 present the general situation for n = 5 (packet payload 1021 B). Clearly, delay climbs more rapidly under infinite ARQ within a critical region around  $P_e = 10^{-4}$ .

# 3.3. Fuzzy logic control

A fuzzy subset is expressed as a set of rules which take the form of linguistic expressions. These rules express experience of tuning the controller and are captured in a knowledge database. An inference engine is the intelligence of the controller, with the capability of emulating the human decision making process, based on fuzzy logic, by means of the knowledge database and embedded rules for making those decisions. Lastly, defuzzification converts inferred



FIGURE 1: Packet delay against  $P_e$  (logarthmic horizontal scale) for varying values of M (max number of retransmissions).

fuzzy control decisions from the inference engine to a crisp or precise value, which is converted to a control signal.

In a fuzzy subset, each member is an ordered pair, with the first element of the pair being a member of a set Sand the second element being the possibility, in the interval [0, 1], that the member is in the fuzzy subset. This should be compared with a Boolean subset in which every member of a set S is a member of the subset with probability taken from the set 0, 1, in which a probability of 1 represents certain membership and 0 represents nonmembership.

As a simple example, in a fuzzy subset of (say) "tall," the possibility that a person with a given height taken from the set S of heights may be called tall is modeled by a membership function, which is the mapping between a data value and possible membership of the subset. Notice that a member of one fuzzy subset can be a member of another fuzzy subset with the same or a different possibility. Membership functions may be combined according to a set of "if ... then" rules to make inferences such as if x is tall and y is old then z is happy, in which tall, old, and happy are membership functions of the matching fuzzy subsets and x, y, z are linguistic variables (names for known data values).

In practice, the membership functions are applied to the data values to find the possibility of membership of a fuzzy subset and the possibilities are subsequently combined through defuzzification to provide a precise output. We have applied a semimanual method of deriving the rules, combining human knowledge of network behavior with testing by simulator.

The fuzzy model behavior itself was examined through Matlab fuzzy toolbox v. 2.2.4. This results in a widely applicable but static set of rules. The FLC's behavior can be predicted from its output surface, formed by knowledge of its rule table and the method of defuzzification. For example, Matlab's toolbox allows a set of output data points to be calculated to a given resolution, allowing interpolation of the surface.



FIGURE 2: Overview of the FLC of ARQ system.

#### 3.4. Fuzzy logic control of ARQ

Figure 2 shows the complete two-stage FLC adaptive ARQ system. For the first stage, there are two inputs: buffer fullness and the normalized delay of the head of the queue packet. Bluetooth buffer fullness is a preferable measure (compared to delay or packet loss) of channel conditions and of buffer congestion, as was established in [32]. Buffer fullness is available to an application via the host controller interface (HCI) presented by a Bluetooth hardware module to the upper layer software protocol stack. As an FLC input, buffer fullness is normalized to the size of the send buffer.

The retransmission count of the packet at the head of the Bluetooth send queue will affect the delay of packets still to be transmitted. Retransmissions overcome the effect of noise and interference but also cause the send buffer queue to grow, with the possibility of packet loss from send buffer overflow, which is why it is necessary also to introduce a DAB. The second FLC input modulates the buffer fullness input by the already experienced delay of the head of queue packet.

The output of the first stage FLC forms the input of the second stage FLC. The other input to the second stage is normalized remaining power, assuming a predetermined power budget for streaming of a particular video clip, which diminishes with time and retransmissions. The output of the second stage is a transmission count, which is subsequently scaled according to picture type importance. Though it might be possible to modify the first stage output by nonfuzzy logic means, by keeping the whole within an FLC framework, the possibility of complex power models is allowed for.

The assigned membership functions, which were achieved heuristically, are shown in Figures 3(a) and 3(b), and once found remain fixed. The buffer fullness range in Figure 3(a) is [0-1] corresponding to a percentage fullness. In Figure 3(b), the horizontal axis represents the delay time of the packet at the head of the queue divided by the display deadline. In Figure 3(b), unit delay corresponds to expiration of playout deadline. It is important to note that any packet in the send buffer is discarded if its deadline has expired. However, this takes place after the fuzzy evaluation of the desired ARQ retransmission count. In practice, the inputs to the FLC were sampled versions of buffer fullness and packet delay deadline, to avoid excessive ARQ retransmission count oscillations over time. The sampling interval was every 20 packets. Table 1 shows the "if ... then" rules that allow input fuzzy subsets to be combined to form an output from stage one and an input to stage two. Notice more than one rule may apply because of the fuzzy nature of subset membership. The output of stage one is combined with a fuzzy input for "remaining power," and the "if ... then" resulting in the final nonscaled transmission count in Table 2.

The inputs were combined according to the well-known Mamdani model [33] to produce the output values for each stage. The standard center of gravity method was employed to resolve to a crisp output value, according to the output membership functions shown in Figures 3(c) and 3(e). The fuzzy control surfaces are represented in Figure 4, as derived from the Matlab Fuzzy Toolbox v. 2.2.4. As mentioned in Section 1, by means of an LUT derived from the surface, a simple implementation becomes possible.

Clearly a packet can only be transmitted an integer number of times but the final crisp output may result in a real-valued number. This difficulty was resolved by generating a random number from a uniform distribution. If the random number was less than the fractional part of the crisp output value then that value was rounded up to the nearest integer, otherwise it was rounded down. Notice that this means that, for (say) a less important B-picture packet very close to its display deadline, a packet at the head of the queue may never be transmitted because of the impact upon more important packets still remaining in the send buffer. The advantage of the randomization procedure over simple quantization is that, in the long term, the mean value of the output numbers of transmissions will converge more closely to a desired output level. The output value was subsequently scaled according to the priority of the packet's picture type. The complete algorithm including randomization and scaling is summarized in Figure 5.

A simple scaling of 5 : 3 : 2 was applied, respectively, for I-, P-, B-pictures, giving up to a maximum of five transmissions. The value of five retransmissions was selected to be inline with the experiments reported in [26]. Subsequently, the retransmission limit for the other picture types was scaled accordingly. In practice, the scaling was applied to the crisp value output after defuzzification. For example, if the crisp output value was 0.7, and a P-picture packet was involved then the value after scaling is  $0.7 \times 3.0 = 2.10$ . Then, the random-number-based resolution results in three transmissions if the random number is less than or equal to 0.10 and two transmissions otherwise.

# 3.5. Deadline-aware buffer

In the conservative send buffer discard policy of this paper, all packets of whatever picture type have a display deadline, which is the size of the playout buffer expressed as a time beyond which buffer underflow will occur. In a conservative policy, the deadline is set as the maximum time that the playout buffer can delay the need for a packet. In the



FIGURE 3: Fuzzy membership functions: (a) stage one, input buffer fullness; (b) stage one, input delay deadline; (c) output of stage one controller; (d) stage two input remaining power; (e) stage two output transmission count.

TABLE 1: FLC stage one if ... then rules used to identify output fuzzy subsets from inputs.

			Delay/deadline			
		Too low	Low	Normal	High	Too high
Buffer fullness	High	Normal	Normal	Low	Too low	Too low
	Normal	Too high	High	Normal	Low	Too low
	Low	Too high	Too high	High	Low	Too low

			Output1			
		Too low	Low	Normal	High	Too high
Remaining power	High	Too low	Low	High	Too high	Too high
	Normal	Too low	Low	Normal	High	High
	Low	Too low	Too low	Low	Low	Normal

TABLE 2: FLC stage two if ... then rules used to identify output fuzzy subsets from inputs.



FIGURE 4: (a) Stage one, FLC control surface resulting from FLC ARQ; (b) stage two, control surface giving the transmission count output (before subsequent scaling).

simulations of Section 4, the display deadline was set to 0.10 second.

In addition to the display deadline, all I-picture packets have a decode deadline, which is the display time remaining to the end of the GOP. This is because reference pictures (I- or P-) are still of value to the receiver as they serve in the decoding of subsequent pictures, even after their display deadline has elapsed. Thus, for a 12-picture GOP, this is the time to display 11 frames, that is, 0.44 second at 25 frame/s. For P-picture packets, the decode deadline will vary depending on the number of frames to the end of the GOP. For B-pictures the decode deadline is set to zero.

The decode deadline is added to the display deadline and a packet is discarded from the send buffer after its total deadline expires. By storing the GOP end time, an implementation performs one subtraction to find each decode deadline. Account has been taken of I- B- P-picture reordering at encode and send buffer output, which has an effect on buffer fullness. Reordering is introduced to ensure that reference pictures arrive and can be decoded before the dependent B-pictures. In the discard policy, packet handling and propagation delay are assumed (optimistically) to be constant. In all experiments, the buffer queue discipline is assumed to be first-in, first-out.

# 3.6. Channel model

Wireless channel errors are usually bursty and dependent in time, rather than independent and identically distributed. For this reason, we adopt a Gilbert-Elliott [34, 35] two state discrete-time, ergodic Markov chain to model the wireless channel error characteristics between a Bluetooth master and slave node. By adopting this model, it is possible to simulate burst errors of the kind that cause problems to an ARQ mechanism. The Gilbert-Elliott model was, in [36], applied to the same version of Bluetooth as herein.

The mean duration of a good state,  $T_g$ , was set at 2 seconds and in a bad state,  $T_b$ , was set to 0.25 second. In units of 625  $\mu$ s (the Bluetooth time slot duration),  $T_g = 3200$  and  $T_b = 400$ , which implies from

$$T_g = \frac{1}{1 - P_{gg}}, \qquad T_b = \frac{1}{1 - P_{bb}},$$
 (6)

that, given the current state is good (g),  $P_{gg}$ , the probability that the next state is also g, is 0.9996875 and, given the current state is bad (b),  $P_{bb}$ , the probability that the next state is also b, is 0.9975. The transition probabilities,  $P_{gg}$  and  $P_{bb}$ , as well as the bit-error rate (BER) are approximately similar to those in [37], but the mean state durations are adapted to Bluetooth. At 3.0 Mbps, the BER during a good state was set to  $a \times 10^{-5}$  and during a bad state was set to  $a \times 10^{-4}$ , where a is a scaling factor and is subsequently referred to as the channel parameter.

#### 3.7. Bluetooth adaptive ARQ schemes

Unfortunately, in respect to Bluetooth, we are not aware of other adaptive ARQ that would form a direct point of comparison to our FLC scheme, particularly if a power budget is factored in. As an alternative Bluetooth comparison, an adaptive ARQ scheme designed for audio streaming [34] was



FIGURE 5: FLC algorithm for processing a packet.

considered. For ease of reference, the details are summarized in this section.

In [38], the round-trip time (RTT) was measured at the link layer. The RTT was then smoothed over time, using a forgetting constant  $\gamma$  to form the smoothed RTT (SRTT). From these values, a retransmission timeout (RTO) was formed. The RTO forms a threshold on the number of ARQ retransmissions,

$$SRRT = (1 - \gamma) \times SRTT + \gamma \times RTT,$$

$$RTO = \begin{cases} \alpha \times RTO, & \text{if } RTT < SRTT, \\ \beta \times RTO, & \text{if } RTT > SRTT, \\ RTO, & \text{if } the previous packet was lost. \end{cases}$$
(7)

In simulations, the values of  $\gamma = 0.25$ ,  $\alpha = 1.1$ ,  $\beta = 0.9$  were adopted from [34] as bounds on RTO, namely RTO<sub>min</sub> was set to the total time to send a packet,  $T_{Packet}$ , which is the Bluetooth packet length divided by the arrival rate at the Bluetooth sender of the data forming that packet. The upper bound was set as follows:

$$RTO_{max} = T_{packet} \times Max(AvailBuff \times 75\%, 2),$$
 (8)

where AvailBuff is the remaining free space in the buffer.

Because this adaptive ARQ algorithm relies on a calculation of the available buffer space in the Bluetooth send buffer,



FIGURE 6: Distribution of slice sizes for the encoded video clip.

it is not possible to combine this algorithm with the use of a DAB. As the adaptive ARQ system relies on buffer fullness to adjust the number of retransmission, if a DAB is employed, expired packets will be actively removed from the buffer, keeping the buffer fullness at a low level. This will mislead the algorithm as it will interpret this low buffer fullness as a sign of the available capacity in the network and increase the number of retransmissions. Because our purpose was to make a fair comparison and because the absence of a DAB unfairly increases packet delays compared to default ARQ and FLC ARQ, in simulations with this adaptive ARQ algorithm, packets were not dropped at the receiver if their frame had missed its display deadline at the receiver. This compensates the calculated PSNR for this algorithm in the results in Section 4.

#### 3.8. Simulation setup

This research employed the University of Cincinatti Bluetooth (UCBT) extension (a download is available from http://www.ececs.uc.edu/~cdmc/ucbt/) to the well-known ns-2 network simulator (v. 2.28 used). The UCBT extension supports Bluetooth EDR but is also built on the air models of previous Bluetooth extensions such as BlueHoc from IBM and Blueware. The Gilbert-Elliott channel model was coded in C++ to be called by an ns-2 object tcl (otcl) script. All links were set at the maximum EDR 3.0 Mbps gross air rate. Each of the simulation runs was repeated twenty times and the results were averaged to produce summary statistics.

The simulations were carried out principally with input from an MPEG-2 encoded bitstream at a mean rate of 1.5 Mbitps for a 30-second video clip with moderate motion, showing a newsreader and changing backdrop, which we designate "News." (Other video inputs are summarized in Section 4.) PSNR was found by reconstructing with a reference MPEG-2 decoder. The display rate was 25 frames/s, resulting in 750 frames in each run. The source video was common intermediate format (CIF)-sized ( $366 \times 288$  pixels) with a GOP structure of N = 12, and M = 3 (where in standard codecs N designates the GOP length and M is the number of pictures between anchor pictures). The slice size distribution of the input video clip is shown in Figure 6.



FIGURE 7: Output from stage one of the FLC, with a = 2.



FIGURE 8: Output from stage two of the FLC, with a = 2.

In [39], fully filled Bluetooth packets were formed using maximal bandwidth five time slot packets, regardless of slice boundaries. These packets carry a 1021 B payload. While this results in some loss in error resilience, as each MPEG-2 slice contains a decoder synchronization marker, in [39], it is shown that the overall video performance is superior to choice of smaller packet sizes.

# 4. RESULTS

#### 4.1. Fuzzy logic model response

Figure 7 shows the output of stage 1 of the FLC as the "News" video clip of Section 3.7 was passed across a Bluetooth link with channel parameter *a* set to two. The high variability of the output is due to the repeated onset of bad states occasioned by the Gilbert-Elliott channel model (Section 3.5).

The normalized power budget for the clip declines with the number of bits passed across the link and the loss is exacerbated by repeated retransmissions during bad states. As the power budget changes linearly, this has the effect of modulating the original input, as illustrated in Figure 8, again with channel parameter set to two.



FIGURE 9: Buffer fullness input to stage one of the FLC, with a = 2.



FIGURE 10: Delay input to stage one of the FLC, with a = 2.

After the removal of deadline expired packets, through operation of the deadline aware buffer (DAB) described in Section 3.4, the buffer fullness input to stage one of the send buffer oscillates around a level well below the 50-packet maximum, Figure 9. Head-of-line packet delay, Figure 10, acts as a typical trimming input to the FLC stage one unit, as its pattern resembles that of buffer fullness over time. Notice that for the default ARQ scheme, Figure 11, delay is frequently over the 0.10 second display deadline and, therefore, B-picture packets face the possibility of being dropped without transmission if they have already spent longer than that time in the send buffer, while I- and Ppicture packets have the grace arising from their extra decode deadline time.

#### 4.2. Response of FCL, default ARQ, and adaptive ARQ

A comparison was made between the default scheme with infinite ARQ, the adaptive ARQ scheme of Section 3.6, and the FLC scheme. These schemes were all allocated an infinite power budget. The FLC scheme with power control was then introduced. To improve the comparison, the default static ARQ scheme was compared with a DAB in place, though,



FIGURE 11: Delay in default ARQ with DAB, with a = 2.



FIGURE 12: Packet loss during transmission of the "News" video clip, with the default scheme and the FLC power-aware scheme.

of course, a DAB is not a feature of the Bluetooth standard. The channel parameter, *a*, was varied in the tests to show the impact of differing channel conditions.

Figure 12 compares the ratio of packets lost to total packets arriving in the send buffer. The FLC ARQ is superior in worsening channel conditions both to default static ARQ and the adaptive scheme [34]. Even when compensating for a diminishing power budget, the FLC scheme shows a clear improvement. By monitoring the local (sender) buffer fullness and reducing the number of retries in the event of congestion, packet loss due to buffer overflow is reduced. In addition, as delay is also considered by the FLC, it is less likely that a packet's delay exceeds the display deadline (and therefore, removed by the DAB scheme). Therefore, the total packet loss rate is reduced when the proposed scheme is employed. Of course, when a power constraint is also considered, the packet loss rate will be compromised but as the Figure 12 shows the FLC still outperforms the other schemes.



FIGURE 13: Average packet delay during transmission of the "News" video clip, with the default scheme and the FLC power-aware scheme.

The average delay of successfully transmitted packets was also considerably reduced under the FLC schemes, Figure 13, while the default ARQ scheme results in a more rapid climb to its peak average value. Larger average delay will impact start-up time in one-way streaming and will add to overall delay in a two-way video exchange, such as for a videophone connection. Notice that removing the power budget results in more delay for the FLC scheme than with a power budget because the scheme is not handicapped by the need to reduce transmissions for power considerations. Either way the scheme is superior to default ARQ in delay (and also in reduced packet loss). As remarked in Section 3.6, the adaptive ARQ scheme is disadvantaged by the lack of a DAB and for that reason its results are not plotted in Figure 13.

Crucially, the FLC is able to save power over both the nonpower-aware default ARQ and the adaptive scheme, Figure 14. The impact is clearly greater as channel conditions worsen. Closer inspection of the distribution of packet losses between the picture types shows the advantage of FLC ARQ, Figure 16, as less B-picture packets and more reference picture packets are lost under default ARQ, Figure 15.

In fact, the loss pattern of the default ARQ replicates the distribution of packet types within the input video clip, Table 3, whereas FLC does not, as is clear by comparing the final two columns of Table 3. This is because the FLC is able to take account of packet type through the delay deadline of the head-of-line packet and because the number of transmissions output is scaled according to the picture type.

Considering the packet loss statistics of Figure 12 and the distribution of those packet losses between packet picture types from Figures 15 and 16, it is not surprising, Figure 17, that the mean PSNR of FLC ARQ is better than that of the other schemes and the relative advantage becomes more so as the channel conditions worsen. A significant part of that advantage is also due to the superiority of FLC ARQ and there is little difference between FLC ARQ with and without a power budget in better channel conditions. Notice that for

Picture type	Packets in input video bitstream (%)	Lost packets in infinite ARQ scheme (%)	Lost packets in fuzzy ARQ scheme (%)
Ι	17.97	17.29	7.50
Р	37.93	36.92	22.91
В	44.10	45.79	69.59

TABLE 3: Percentage distribution of input and lost packets by picture type.



Compared to infinite ARQ with DAB
 Compared to the adaptive scheme



FIGURE 14: Relative power saving of the FLC power-aware ARQ scheme compared to that of the default ARQ and the adaptive ARQ [34] schemes.



FIGURE 15: Contribution of I-, P-, and B-frame packet losses to total packet loss in the infinite ARQ with DAB scheme.

PSNR is improved by around 3 dB if an infinite power budget

power-aware control averaged PSNR, figures do not "show the whole story," as the achievable PSNR will deteriorate over time, as the available power becomes less. This confirms previous experience [40] that for the very worst channel conditions shown in Figure 17, that is, a = 5, then the mean

in the same conditions.

FIGURE 16: Contribution of I-, P-, and B-frame packet losses to total packet loss in the FLC with DAB and power-aware scheme.



FIGURE 17: Comparison of PSNR for the "News" video clip between the Bluetooth default and FLC ARQ schemes with DAB.

□ B-frame packets
 □ I-frame packets
 □ I-frame packets

	a	= 1	а	= 2	а	= 3	а	= 4	a =	= 5
	FLC	Inf ARQ								
News	32.22	31.12	31.45	30.05	31.18	28.78	30.19	27.56	29.29	25.33
Football	31.32	29.91	30.85	28.81	30.01	27.55	28.95	26.31	28.19	24.88
Friends	31.88	30.09	30.94	28.94	30.14	27.77	29.32	26.67	28.34	24.29
Italian job	30.97	30.41	30.39	29.01	30.02	27.62	29.19	26.55	28.45	24.09

TABLE 4: Comparison of video quality between power-aware FLC ARQ with DAB, and default ARQ with DAB for various video clips.

#### 4.3. Detailed comparison of video quality

As an illustrative example of the fluctuations of video quality over time, for the purposes of this test only, the power budget was artificially set to  $60 \times 10^6$  bits for both the Bluetooth default ARQ scheme and power-aware FLC ARQ. The choice of the power budget is arbitrary and the reader can refer back to Figure 14 for a quantitative comparison of the relative power savings, with no specific power budget imposed. Again, a DAB was in place for both tests. Figure 18 demonstrates a falling trend in PSNR quality over the course of the video clip stream, though quality is generally high for a wireless channel (channel parameter a = 2). In contrast, under the default ARQ scheme (see Figure 19), not only is the video quality lower but also there is an abrupt end to transmission as the bit budget runs out at about 652 frames. Figure 20 shows the equivalent result for adaptive ARQ of [38], when it will be seen that the objective video quality is more variable than FLC.

To verify the generality of the results, a comparison was made across a variety of video clips. Table 4 provides summary statistics (mean of 20 runs) for the different schemes, with four input video sequences: (1) "News," as in previous experiments in this section, (2) "Football" with rapid movement, (3) "Friends" from the well-known American situational comedy, with more "action" than in "News" and equally, (4) "Italian job," with an extract from the wellknown film including car chases. The additional clips had the same GOP structure as the "News" sequence and similarly were CIF-sized at 25 frames/s and were encoded at the same rate. The default ARQ with DAB scheme once again was deployed with an infinite power budget, whereas once again the FLC ARQ regulated its power allocation for each video clip over the course of the streaming session. Over changing channel conditions, Table 4, the results for video quality are broadly similar, except that the greater motion in the other clips results in a lower received video quality.

# 5. CONCLUSIONS

Power usage becomes an important factor when mobile stations are employed in ad hoc mode or when the receiver is a mobile device. The proliferation of such devices implies that any reduction in the recharge frequency is a welcome development. Certainly, alternatives to Bluetooth are considered in power terms, whether Wibree or as a commercial sensor network IEEE 802.15.4 (Zigbee). Transmission of higher quality video over a Bluetooth interconnect has been



FIGURE 18: PSNR over the 750 frames, showing a falling trend in PSNR over time, for the power-aware FLC ARQ with DAB.



FIGURE 19: PSNR over the 750 frames of the input video, under a fixed power budget with the default ARQ with DAB.

long sought. However, it is important to factor in power usage and not simply regard a wireless channel as a fixed channel with the addition of errors, to caricature one view. In this paper, fuzzy logic control of ARQ is able to respond to a fixed power budget, which diminishes over time. Other factors included are packet delay deadlines (both display and, for anchor picture packets, decode deadlines), and send buffer congestion. Because an ARQ management system is not also able to manage the send buffer size, a deadline aware buffer (DAB) removes expired packets. This avoids retransmission attempts on these packets and more importantly



FIGURE 20: PSNR over the 750 frames of the input video, under a fixed power budget with an adaptive ARQ scheme [34].

prevents send buffer overflow and excessive waiting times for other queued packets. For fairness, both the default infinite ARQ scheme and the FLC scheme were compared with the addition of a DAB. An adaptive ARQ scheme specifically for Bluetooth from the literature was also compared, though the nature of the algorithm did not permit the use of a DAB. However, FLC, which varies its transmission policy with packet picture type, still outperforms both the default static ARQ scheme and the conventional, adaptive ARQ, resulting in the end analysis in superior delivered video quality. This is despite the need to adjust the transmission policy as available power diminishes, whereas infinite battery power is assumed for the default scheme. The FLC framework, being modular, allows for a future power model that takes into account not only energy loss from transmission but also models energy taken up at the encoder and/or the decoder.

# ACKNOWLEDGMENT

This work was supported by the EPSRC, UK, under Grant no. EP/C538692/1.

#### REFERENCES

- [1] J. Haartsen, "The Bluetooth radio system," *IEEE Personal Communications*, vol. 7, no. 1, pp. 28–36, 2000.
- [2] S. Mattisson, "Low-power considerations in the design of Bluetooth," in *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED '00)*, pp. 151–154, Portacino Coast, Italy, July 2000.
- [3] M. Tamai, T. Sun, K. Yasumoto, N. Shibata, and M. Ito, "Energy-aware video streaming with QoS control for portable computing devices," in *Proceedings of the 14th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '04)*, pp. 68–73, Cork, Ireland, June 2004.
- [4] P. Agrawal, J.-C. Chen, S. Kishore, P. Ramanathan, and K. Sivalingam, "Battery power sensitive video processing in wireless networks," in *Proceedings of the 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communi-*

*cations (PIMRC '98)*, vol. 1, pp. 116–120, Boston, Mass, USA, September 1998.

- [5] K. Choi, K. Kim, and M. Pedram, "Energy-aware MPEG-4 FGS streaming," in *Proceedings of the 40th Design Automation Conference*, pp. 912–915, Anaheim, Calif, USA, June 2003.
- [6] A. Iranli, K. Choi, and M. Pedram, "Energy-aware wireless video streaming," in *Proceedings of the 1st Workshop on Embedded Systems For Real-Time Multimedia (ESTimedia '03)*, pp. 48–55, Newport Beach, Calif, USA, October 2003.
- [7] H.-S. Kim, D. T. Duong, J.-Y. Jeong, B.-K. Dan, and S.-J. Ko, "Power-aware rate control for mobile multimedia communications," in *Proceedings of the 5th International Conference on Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW '06)*, vol. 4104 of *Lecture Notes in Computer Science*, pp. 458–471, Ottawa, Canada, August 2006.
- [8] R. Cornea, A. Nicolau, and N. Dutt, "Content-aware power optimizations for multimedia streaming over wireless networks," Tech. Rep. CECS #06-13, University of California, Berkeley, Calif, USA, 2006.
- [9] Specification of the Bluetooth System—2.0 + EDR, 2004, http://www.bluetooth.com.
- [10] E. Ferro and F. Potortì, "Bluetooth and Wi-Fi wireless protocols: a survey and a comparison," *IEEE Wireless Communications*, vol. 12, no. 1, pp. 12–26, 2005.
- [11] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless Internet," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 123–133, 2005.
- [12] M. Chen and G. Wei, "Multi-stages hybrid ARQ with conditional frame skipping and reference frame selecting scheme for real-time video transport over wireless LAN," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 158–167, 2004.
- [13] R. Razavi, M. Fleury, and M. Ghanbari, "Fuzzy logic control of adaptive ARQ for video distribution over a Bluetooth wireless link," *Advances in Multimedia*, vol. 2007, Article ID 45798, 13 pages, 2007.
- [14] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [15] H. Takagi, Application of Neural Networks and Fuzzy Logic to Consumer Products, vol. 1 of IEEE Technology Updates Series: Fuzzy Logic Technology and Applications, IEEE Press, New York, NY, USA, 1994.
- [16] A. Leone, A. Bellini, and R. Guerrieri, "An H.261-compatible fuzzy-controlled coder for videophone sequences," in *Proceedings of the 3rd IEEE Conference on Fuzzy Systems*, vol. 1, pp. 244–248, Orlando, Fla, USA, June 1994.
- [17] P. M. Grant, Y.-S. Saw, and J. M. Hannah, "Fuzzy rule based MPEG video rate prediction and control," in *Proceedings of EURASIP ECASP Conference*, pp. 211–214, Prague, Czech Republic, June 1997.
- [18] I. Baturone, A. Barriga, S. Sánchez-Solano, C. J. Jiménez-Fernández, and D. R. López, *Microelectronic Design of Fuzzy Logic-based Systems*, CRC Press, Boca Raton, Fla, USA, 2000.
- [19] M. Kalman, P. Ramanathan, and B. Girod, "Rate-distortion optimized video streaming with multiple deadlines," in *Proceedings of the International Conference on Image Processing* (*ICIP '03*), vol. 3, pp. 661–664, Barcelona, Spain, September 2003.
- [20] J.-C. Chen, K. M. Sivalingam, P. Agrawal, and S. Kishore, "A comparison of MAC protocols for wireless local networks based on battery power consumption," in *Proceedings of the*

17th IEEE Annual Conference on Computer Communications (INFOCOM '98), vol. 1, pp. 150–157, San Francisco, Calif, USA, March-April 1998.

- [21] J.-C. Cano, J.-M. Cano, E. González, C. Calafate, and P. Manzoni, "Power characterization of a Bluetooth-based wireless node for ubiquitous computing," in *Proceedings of the 2nd International Conference on Wireless and Mobile Communications (ICWMC '06)*, p. 13, Bucharest, Romania, July 2006.
- [22] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 398–410, 2002.
- [23] Y. Liang and I. Ahmad, "Power and content aware video encoding for video communication over wireless networks," in *Proceedings of the IEEE Workshop on Signal Processing Systems* (*SIPS* '04), pp. 269–274, Austin, Tex, USA, October 2004.
- [24] M. Tamai, T. Sun, K. Yasumoto, N. Shibata, and M. Ito, "Energy-aware QoS adaptation for streaming video based on MPEG-7," in *Proceedings of the IEEE International Conference* on Multimedia and Expo (ICME '04), vol. 1, pp. 189–192, Taipei, Taiwan, June 2004.
- [25] M. Yokotsuka, "Memory motivates cell-phone growth," Wireless Systems Design, vol. 9, no. 3, pp. 27–30, 2004.
- [26] Q. Li and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through realtime retry limit adaptation," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 278–290, 2004.
- [27] W. Tan and A. Zakhor, "Packet classification schemes for streaming MPEG video over delay and loss differentiated networks," in *Proceedings of the 11th International Packet Video Workshop (PV '01)*, Kyongju, Korea, May 2001.
- [28] J. Wall and J. Y. Khan, "An adaptive ARQ enhancement to support multimedia traffic using 802.11 wireless LANs," in *Proceedings of the IEEE Global Telecommunications Conference* (GLOBECOM '04), vol. 5, pp. 3037–3041, Dallas, Tex, USA, November-December 2004.
- [29] N. Golmie, N. Chevrollier, and O. Rebala, "Bluetooth and WLAN coexistence: challenges and solutions," *IEEE Wireless Communications*, vol. 10, no. 6, pp. 22–29, 2003.
- [30] M. C. Valenti, M. Robert, and J. H. Reed, "On the throughput of Bluetooth data transmissions," in *Proceedings of the IEEE Wireless Communications and Networking Conference* (WCNC '02), vol. 1, pp. 119–123, Orlando, Fla, USA, March 2002.
- [31] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 2000.
- [32] R. Razavi, M. Fleury, and M. Ghanbari, "Detecting congestion within a Bluetooth piconet: video streaming response," in *Proceedings of the London Communications Symposium*, pp. 181–184, London, UK, September 2006.
- [33] J. S. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.
- [34] E. N. Gilbert, "Capacity of burst-noise channel," *Bell System Technical Journal*, vol. 39, no. 9, pp. 1253–1265, 1960.
- [35] E. O. Elliott, "Estimates of error rates for codes on burst noise channels," *Bell System Technical Journal*, vol. 42, no. 9, pp. 1977–1997, 1963.
- [36] L.-J. Chen, T. Sun, and Y.-C. Chen, "Improving Bluetooth EDR data throughput using FEC and interleaving," in *Mobile Ad-Hoc and Sensor Networks*, vol. 4325 of *Lecture Notes in Computer Science*, pp. 724–735, Springer, Hong Kong, 2006.

- [37] R. Fantacci and M. Scardi, "Performance evaluation of preemptive polling schemes and ARQ techniques for indoor wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 2, pp. 248–257, 1996.
- [38] L.-J. Chen, R. Kapoor, K. Lee, M. Y. Sanadidi, and M. Gerla, "Audio streaming over Bluetooth: an adaptive ARQ timeout approach," in *Proceedings of the 24th International Conference* on Distributed Computing Systems Workshops (ICDCSW '04), pp. 196–201, Hachioji, Japan, March 2004.
- [39] R. Razavi, M. Fleury, and M. Ghanbari, "An efficient packetization scheme for Bluetooth video transmission," *Electronic Letters*, vol. 42, no. 20, pp. 1143–1145, 2006.
- [40] R. Razavi, M. Fleury, and M. Ghanbari, "Fuzzy control of adaptive timeout for video streaming over a Bluetooth interconnect," in *Proceedings of 12th IEEE Symposium on Computers* and Communications (ISCC '07), pp. MW-27–MW-32, Santiago, Portugal, July 2007.

# Research Article

# An Adaptive Fair-Distributed Scheduling Algorithm to Guarantee QoS for Both VBR and CBR Video Traffics on IEEE 802.11e WLANs

# Saeid Montazeri,<sup>1</sup> Mahmood Fathy,<sup>2</sup> and Reza Berangi<sup>2</sup>

<sup>1</sup> Computer Group, Islamic Azad University, KhomeiniShahr Branch, Khomeinishar 84175/119, Iran
<sup>2</sup> Department of Computer Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran

Correspondence should be addressed to Saeid Montazeri, s.montazeri@iaukhsh.ac.ir

Received 2 October 2007; Revised 15 February 2008; Accepted 16 April 2008

Recommended by Jianfei Cai

Most of the centralized QoS mechanisms for WLAN MAC layer are only able to guarantee QoS parameters for CBR video traffic effectively. On the other hand, the existing distributed QoS mechanisms are only able to differentiate between various traffic streams without being able to guarantee QoS. This paper addresses these deficiencies by proposing a new distributed QoS scheme that guarantees QoS parameters such as delay and throughput for both CBR and VBR video traffics. The proposed scheme is also fair for all streams and it can adapt to the various conditions of the network. To achieve this, three fields are added to the RTS/CTS frames whose combination with the previously existing duration field of RTS/CTS frames guarantees the periodic fair adaptive access of a station to the channel. The performance of the proposed method has been evaluated with NS-2. The results showed that it outperforms IEEE 802.11e HCCA.

Copyright © 2008 Saeid Montazeri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. INTRODUCTION

The wireless LAN (WLAN) systems have received increasing popularity in recent years because they are cost effective, comfortable, and have high capacity. On the other hand, using video applications has been very popular in recent years. Therefore, effective using of video streams over the WLANs is an obligation these days. To achieve this goal, QoS parameters should be supported over WLANs.

Supporting QoS requirements in WLANs can be done in two ways: prioritized QoS and guaranteed QoS. A prioritized-QoS WLAN can only prioritize between different traffic streams while a guaranteed-QoS WlAN can guarantee QoS parameters such as delay, jitter, and throughput for traffic streams. Implementing QoS, either prioritized or guaranteed, is a challenge in WLAN because there are a large number of streams with different QoS requirements in a WLAN. Also some QoS requirements have variable characteristics during the time like a VBR video traffic. These characteristics lead to an adaptive QoS supporting approach in WLANs. In addition to the large number of streams and QoS requirements which vary during the time, wireless channel capacity is limited and must be shared among streams fairly. Thus, an adaptive fair algorithm which can guarantee QoS parameters is necessary in WLANs.

IEEE task group "e" worked on the support of QoS in a new standard, called IEEE 802.11e [1]. It introduces a new access method called hybrid coordination function (HCF), which combines functions from the DCF and PCF mechanisms in IEEE 802.11. HCF has two access mechanisms: enhanced distributed channel access (EDCA) and controlled channel access mechanism (HCCA). These two methods support QoS, which will be described further. In the HCCA, there is a scheduler for scheduling different traffic streams (TSs) on different stations. The HCCA can guarantee QoS parameters but it needs a centralized device that is called point coordinator (PC). On the other hand, the EDCA which does not use any PC could not guarantee QoS parameters. It can only operate for high-priority traffics sufficiently so it is not a fair method. In addition, both HCCA and EDCA have to tolerate high overhead to adapt to the network conditions.

Many works have been done to improve QoS in the IEEE 802.11e MAC layer. These works can be divided into two

categories: the works that improve QoS distributively and the works which improve QoS by using PC.

In [2], the authors proposed a new adaptive fairdistributed method. This method enhances the EDCA of IEEE 802.11e by increasing the contention window when the channel is busy. It also uses an adaptive fast backoff mechanism when the channel is idle. They computed an adaptive backoff threshold for each priority level by taking into account the channel load. In [3], the authors proposed a fully distributed MAC adaptation method. They achieve this by updating the MAC layer parameters like contention window based on the network condition. Adaptive EDCA is a new method based on the IEEE 802.11e EDCA that is proposed in [4]. The main idea in this method is to decrease CW [i] after a successful transmission and increase it after a collision slower than it is done in the EDCA. Also it takes into account both the network condition and application requirements. An improved EDCA is achieved in [5] by using the new backoff algorithm called agedependent backoff (ADB). ADB changes the persistence factor by using the age of packets in the transmission queue and their lifetime. In [6], the authors proposed a mechanism called A-DRAFT that supports both absolute and relative throughputs in adaptive distributed manner. This mechanism also provides fair throughput support with low variation. In [7], a new mechanism called differentiated service EDCA (DSEDCA) was proposed to provide both strict priority and proportional fair service for IEEE 802.11 WLANs. In this mechanism, resource is allocated to flows of higher priority, then the remaining bandwidth is shared proportionally among the other service class according to their assigned weights. In [8], authors proposed a surplus TXOP diverter (STXD) scheduling algorithm which allows each flow to exploit its granted TXOP time to reduce the delay when burst packets arrival.

In [9], authors proposed a new scheduling algorithm in link layer to support multimedia services with guaranteed QoS in WLAN. Their scheduling algorithm is based on the HCF. It reduces average packet loss ratio by setting constant bit-rate (CBR) RT to the highest priority followed by VBR RT, and after all NRT level. It also uses idle time, while satisfying required rate allocation, transmission delay bound, and system throughput. In [10], a fair QoS agent (FQA) is proposed to provide per-class QoS enhancement and per-station fair channel access simultaneously. Authors put the FQA algorithm above the MAC layer which enables algorithm to be implemented without any change in the MAC layer. Their algorithm satisfies the fairness in WLAN MAC layer. In [11], a novel QoS capable station (QSTA) uplink scheduler along with a QoS capable AP (QAP) HCF scheduler can provide the QoS requirement of delay bound for multimedia applications. In [12], the authors proposed a new scheduling algorithm for IEEE 802.11e which they called FHCF. It outperforms IEEE 802.11e HCF especially for VBR traffic. It uses queue length estimation to tune time allocation to stations. A new scheduling algorithm has been proposed in [13] which enables the IEEE 802.11e scheduler to work with different SIs for different TSs in the stations. In [14], the authors proposed a dynamic bandwidth allocation algorithm along with the measurement-based call admission control algorithm which can provide delay guarantee for real-time flow. It uses a classic feedback control system. There is an enhancement for IEEE 802.11e HCF in [15] that improves the admission control unit of HCF. By this method, each priority has a certain and limited amount of resources.

Although all distributed methods in [2–8] can improve EDCA in IEEE 802.11e, they can not guarantee QoS parameters in WLANs. However, methods in [9–15] can guarantee QoS in WLANs, they need point coordinator and cause high overhead to guarantee QoS parameters for VBR video traffic. This paper proposes an adaptive fair-distributed scheduling algorithm (AFDSA) [16] for both VBR and CBR video traffic streams in WLANs. AFDSA is a distributed method that operates better than centralized methods in all fields especially for VBR video traffics with low overhead.

The rest of the paper is organized as follows. The IEEE 802.11e is introduced in Section 2. The properties of AFDSA algorithm are described in Section 3. Section 4 describes the simulation results. The conclusion follows in Section 5.

# 2. IEEE 802.11e MAC

Hybrid coordination function (HCF) of IEEE 802.11e MAC has both contention-based access method and polling-based access method. EDCA introduces the concept of access categories (ACs), which can be considered as instances of the DCF access mechanism. It provides support for the prioritized delivery at each station.

# 2.1. Enhanced distributed channel access (EDCA)

Like DCF, EDCA uses CSMA/CA protocol to access the wireless media. It only operates during CP. In EDCA method, each AC within the stations contends for transmission opportunity (TXOP) independently. TXOP is defined as the interval of time when a particular station has the right to initiate the transmission onto the wireless channel. Each AC starts the backoff after detecting the channel to be idle for a time interval equal to the arbitration interframe space (AIFS). Each AC has its AIFS which depends on the assigned priority. Figure 1 demonstrates the eight different queues for eight ACs.

Each AC has its own queue,  $CW_{min}[AC]$ ,  $CW_{max}[AC]$ , and PF[AC]. Figure 2 shows the different ways to provide service differentiation.

For each AC, backoff is generated in the range of [1, CW[AC]+1]. The initial value for the CW is  $CW_{min}[AC]$ . CW is increased whenever the node involves in a collision by (1) up to  $CW_{max}[AC]$ :

$$newCW[AC] = ((oldCW[AC] + 1)*PF[AC]) - 1, \quad (1)$$

where PF is the persistence factor, which equals 2 by default. It determines the degree of increase for the CW when



FIGURE 2: Different AIFS for different priorities.

a collision happens. EDCA can only differentiate between different priorities.

# 2.2. HCF controlled channel access (HCCA)

In IEEE 802.11e standard, the polling-based scheme of 802.11 is extended in the form of HCCA, in which there is a hybrid coordinator (HC) usually colocated with a QoS AP (QAP). HC can access channel after waiting for a time which is shorter than each AIFS and DIFS. Thus, HC can get the channel in both CFP and CP. During CP, TXOP for each station can be received in two ways: by using EDCA rules or by receiving a poll from HC (polled\_TXOP). During CFP, TXOP is determined only by HC with poll frame. CFP is ended by a CF\_end frame which is transmitted by HC.

# 2.3. 802.11e HCF scheduling scheme

The HCF has a simple scheduler in IEEE 802.11e. If a QoSenhanced station (QSTA) needs a strict QoS support, it should send a QoS requirement packet to the QAP while the QAP can allocate the corresponding channel time for different QSTAs according to their requirements. Figure 3 shows the new beacon interval of 802.11e, CFP, and CP. The QAP can operate in both CFP and CP. During the CP, the QAP can start several contention-free bursts at any time to control the channel which are called controlled access periods (CAPs).

If a station requires a contention-free access to the channel by getting TXOP, it should send a QoS request frame to the QAP containing several parameters. These parameters are mean data rate of the application, the maximum service



FIGURE 3: CFP, CP, and CAPs in the 802.11e.

interval (MSI) and MAC service data unit (MSDU) size. Then the QAP calculates the TXOP in two steps. In the first step, it determines the minimum value of all MSIs required by different traffic streams. Then, it chooses the highest submultiples' value of the 802.11e beacon interval duration (duration between two beacons) as the selected SI which is less than the minimum of all requested MSIs. This selected SI is the time between two successive TXOPs for all streams. Since it is less than or equal to all MSIs, it is guaranteed that every station with different streams can reach desired MSI for their streams. In the second step, the QAP calculates the TXOP for each TSs in different QSTAs. Calculated TXOP should correspond to the duration required for transmitting all packets that is generated during one SI by the specific TS. Figure 4 shows the CPs, CFPs, selected SI, and EDCA time.

Equations (2) and (3) determine the TXOP, where  $\rho$  is the mean data rate of the application, and *L* is the MAC service data unit (MSDU) size:

$$N_i = \left\lceil \frac{\mathrm{SI} \times \rho_i}{L_i} \right\rceil. \tag{2}$$

Here,  $N_i$  is the number of packets that is generated during an SI for the *i*th priority. *R* is the physical transmission rate, *M* is the size of maximum MSDU (2304 bytes), and *O* determines the overhead in time units:

$$\mathrm{TXOP}_{i} = \max\left(\frac{N_{i} \times L_{i}}{R_{i}} + O, \frac{M}{R_{i}} + O\right). \tag{3}$$

It can be easily deduced that the TXOP<sub>*i*</sub> is the time required to send  $N_i$  packets for a specific application.

# 3. THE PROPOSED ADAPTIVE FAIR-DISTRIBUTED SCHEDULING ALGORITHM (AFDSA)

#### 3.1. Concept

All centralized channel access methods in WLANs, which are able to guarantee QoS parameters, have one PC that knows the QoS requirements of all TSs. These requirements, which are sent by each station to the PC before starting a transmission, enable the PC to schedule all TSs. The PC can manage QSTAs and guarantee QoS because of its awareness about the requirements of all traffic streams and its ability to get the channel in desirable time. As a matter of fact, PC polls the stations in a proper way by using its knowledge about the network condition. On the other hand, the distributed methods do not have such a PC or equivalent device to gather information and guarantee QoS by managing QSTAs.

The most important characteristic of our approach is to distribute the necessary information (which is different from the one in IEEE 802.11e) among QSTAs to make aware all stations about the network situation. The proposed AFDSA employs the RTS/CTS feature in IEEE 802.11 with some changes. By using this feature, we can reduce the overhead which is required for distributing QoS parameters. The RTS/CTS handshaking mechanism is used to solve the hidden terminal (hidden node) problem in IEEE 802.11 WLANs. A hidden node problem happens when two stations that communicate with a common station are not able to hear each other so their packets collide. Figure 5 demonstrates RTS/CTS packets in which the frame control field is related to the control functions, Duration field contains a value that shows the duration of a transmission, RA is the receiver address, TA is the transmitter address, and FCS is the frame check sequence of the packets.

A RTS/CTS protocol initiates with sending an RTS frame to the receiver (Figure 6). A transmission only starts when a CTS frame is replied to by the receiver. All the stations, which receive one of these frames, understand that a transmission will start and continue for duration equal to the duration field in the RTS/CTS frames. They set their network allocation vector (NAV) to the proper value to prevent themselves from disturbing transmission.

This local hand shaking between transmitter and the receiver provides an excellent opportunity to distribute necessary information to guarantee QoS parameters. To achieve this, the proposed AFDSA uses a modified RTS/CTS protocol with additional fields in the original IEEE 802.11 RTS/CTS frames. The new fields, that is, CurrentSI, FutureSI, and remainderSI, as shown in Figure 7, are added to both RTS/CTS.

What are the CurrentSI, FutureSI, and remainderSI? Before defining these fields, we should define service interval. Service interval is the time between two successive TXOPs that belong to the specific traffic stream. AFDSA uses this concept for the service interval (SI). When a WLAN works with a specific SI, a station can reach the channel for TXOP seconds and it is repeated each SI seconds. In AFDSA, CurrentSI is the service interval that the WLAN is working with at the time of transmission; FutureSI is the service interval that WLAN will work with after ending present SI; and remainderSI indicates the time that is remaining until the end of this service interval (after receiving the RTS/CTS packets). Also TA (transmitter address) is added to the CTS frames for the future development. The protocol needs two timers; a duration timer and a service interval (SI) timer.

The protocol sends the RTS/CTS frames only before the first few packets in each transmission to reduce the transmission overhead. The exact number of required RTS/CTS packets will be calculated in the next section. The QoS parameters can be only guaranteed for a traffic stream (TS) when the TS can have access to the channel for a special duration with a specific SI. Duration for *i*th TS in the *j*th QSTA can be obtained from

$$D_{i}^{j} = N_{i}^{j} * \left(\frac{M_{i}}{R} + 2\text{SIFS} + \text{ACK}_{\text{time}}\right) + \text{RTSNumber} * (\text{RTS}_{\text{time}} + \text{CTS}_{\text{time}} + 2\text{SIFS}) \quad (4) - (\text{RTS}_{\text{time}}),$$



Beacon TXOPi TXOP allocated to QSTAs

FIGURE 4: Structure of the 802.11e beacon interval.



FIGURE 5: RTS/CTS frame structure.



FIGURE 6: RTS, CTS, data, and ACK frames sequence.

Frame control	Duration	RA	TA	CurrentSI	FutureSI	<u>RemainderSI</u>	F C S
------------------	----------	----	----	-----------	----------	--------------------	-------------

FIGURE 7: New RTS/CTS frame structure.

where,  $D_i^j$  is the time required to transmit  $N_i^j$  packets with the length  $M_i$ , the physical rate R, plus the time required to transmit RTSNumber of RTS and CTS frames. Here,  $N_i^j$  is the number of packets in the queue of *i*th TS in the *j*th QSTA at the time of calculating  $D_i^j$ .

The question that is to be answered here is how to guarantee  $D_i^j$  repeats every SI seconds for the *i*th TS in the *j*th QSTA, without disturbing other QSTAs with different TSs requirements. Suppose that *i*th TS in the *j*th QSTA is the first one that starts the transmission in the WLAN with using EDCA method. It calculates  $D_i^j$  and puts it in the duration field of RTS. Then it sets the CurrentSI and FutureSI with maximum service interval for *i*th TS. The *j*th QSTA transmits

RTS frame and waits for receiving CTS. Destination receives the RTS and calculates the duration field for CTS by using

 $Duration_{CTS} = Duration_{RTS} - (SIFS + CTS_{time}).$  (5)

Then, it transmits the CTS frame to the *j*th QSTA. All the QSTAs that receive the RTS or CTS understand a new transmission will be started with the specific length (duration field in the RTS/CTS frames) and will be repeated with a specific period (CurrentSI field in the RTS/CTS frames). Therefore, they reserve this time for station *j* by setting and starting their duration timers with the duration field of RTS/CTS. They also set and start SI timers with the CurrentSI field of RTS/CTS as well as saving FutureSI field of RTS/CTS for the next SI timer restart. Finally, when source receives the CTS frame, it transmits data frame.

The duration field of RTS is updated by the value of the duration timer in the source station. After this, all stations have reserved the allocated turn for the *j*th station and they keep silent during this time. It is done by using an array. Each station has an array and saves the sequence of turns in it. If a station saves 0 in the *i*th place in array, it means that, in the SI, the *i*th turn is reserved for another station, yet if it saves 1 in the *i*th place in array, it means that in the SI the *i*th turn is reserved for itself. The exact duration is announced in the duration field of RTS/CTS by the *j*th station and other stations do not need to save the value of duration field. Therefore, duration field can be varied and updated each time. It is perfect for VBR video traffics and can adapt itself to the network condition.

After finishing  $D_i^{\prime}$ , all QSTAs start to compete for accessing the channel based on EDCA method. Suppose that *k*th QSTA gets the channel for its *l*th TS. It fills the duration field by using (5), and sets the CurrentSI with the value of CurrentSI of *j*th QSTA. However, it sets the FutureSI field with maximum service interval that *l*th TS required when

it is equal or less than previous FutureSI (related to *i*th TS in the *j*th QSTA). So, all the QSTAs that receive the new RTS/CTS understand that they must initialize their SI timer with FutureSI field of new RTS/CTS at the end of current SI. Therefore after a number of SIs, the network works with the sufficient SI. This SI is the minimum of maximum service intervals for all TSs.

After finishing the first SI, all stations check whether they have the first turn. They do this by using their arrays. The station which finds that it has the first turn, that is, *j*th station, starts to calculate the  $D_i^j$  and transmit the RTS. All the other stations keep quiet and wait until they receive an RTS or CTS frame. If a station receives an RTS/CTS, it starts its duration timer. When a duration timer goes zero, it is the time to go to the next turn and search the array. Also requesting stations must only compete in the free time at the tail of current SI (as shown in Figure 8).

Now we can describe why AFDSA is sufficient for transmitting video traffics. Other distributed method, EDCA, can only prioritize between various kinds of streams. As depicted in Table 4, CBR video traffic has the lowest priority and after that comes the VBR video traffic. In a WLAN with different kinds of streams and using EDCA, the video streams can not adequately access the channel in competition with other types of traffics. On the other hand, HCCA method needs specific characteristics of a stream-like data rate and packet size to allocate channel to it. However, data rate and packet size vary during the time for VBR video traffics. As a result, PC in HCCA method can not allocate the accurate time to the VBR traffics since it is obliged to select one of the following two choices. The first one is to allocate the channel to the VBR streams based on the mean data rate. It leads to some dropping packets, wasted channel, and increased jitter and delay because of the great changes in the amount of data. The other choice is to allocate the channel based on the peak data rate to prevent the packet loss. It causes to waste the channel much more than what it may happen in the first method.

AFDSA can adapt the allocated time of each station for transmitting packets by using number of packets that are available in the queue at the beginning of the  $D_i^j$ . This leads to improve the channel efficiently by letting the others to use the channel. This prevents packet dropping and channel wasting simultaneously.

The mentioned channel access process in AFDSA eliminates the need for a point coordinator, though each wireless station can act as an AP when it is connected to the wired network. This enhances the survivability of WLANs in case of an AP failure.

# 3.2. Special situations

This section reviews the performance of proposed protocol in special situation that might happen during a period in which a WLAN works.

#### 3.2.1. Missing the RTS/CTS

It is very important for all the stations to be synchronized so that their SI timers start and finish on time. If a station

RTS <sub>1</sub>	SIFS	CTS1	SIFS
Data <sub>1</sub>	SIFS	$Ack_1$	SIFS
RTS <sub>2</sub>	SIFS	$CTS_2$	SIFS
Data <sub>2</sub>	SIFS	Ack <sub>2</sub>	SIFS
RTS <sub>3</sub>	SIFS	$CTS_3$	SIFS
Data <sub>3</sub>	SIFS	Ack <sub>3</sub>	SIFS
	÷	:	:
RTS <sub>RTSNumber</sub>	SIFS	CTS <sub>RTSNumber</sub>	SIFS
Data <sub>RTSNumber</sub>	SIFS	Ack <sub>RTSNumber</sub>	SIFS
Data <sub>RTSNumber+1</sub>	SIFS	Ack <sub>RTSNumber+1</sub>	SIFS
Data <sub>RTSNumber+2</sub>	SIFS	Ack <sub>RTSNumber+2</sub>	SIFS
Data <sub>RTSNumber+3</sub>	SIFS	Ack <sub>RTSNumber+3</sub>	SIFS

TABLE 1: Data transmission sequence for a specific TS in AFDGP.



FIGURE 8: Free time and learning period.

misses the RTS or CTS, it must wait until it receives the next RTS or CTS. By receiving the next RTS or CTS, it can use the duration and remainderSI fields to synchronize itself with the others because these fields are always up to date. The process of sending and receiving RTS/CTS repeats RTSNumber times to assure that all the active stations in the communication range have received at least one RTS or CTS. After which only data frames will be transmitted. The RTSNumber depends on the BER of the channel and increases with increasing the BER. In our simulation, RTS number is set to 2. Table 1 shows data transmission sequences and shows the impact of RTSNumber on the data transmission.

#### 3.2.2. Entering a new station to the working WLAN

A new station needs remainderSI, FutureSI, and CurrentSI to synchronize with a working WLAN. So it must wait until it receives at least one RTS or CTS and it must wait at least one SI to learn about network condition. This SI which is referred to as the learning period is shown in Figure 8. Any new entering station is prevented to send data during its learning period. Not having permission to send data in learning period is a rule in AFDSA. It can only access the channel based on the EDCA rule in the free time after the learning period and after it receives at least one RTS or CTS packet too.

#### 3.2.3. Removing a duration between other durations

If a station stops using its allocated turn related to a specific TS (e.g., *i*th TS), it must send special RTS to the receiver RTSNumber times. Receiver replies to this special RTS by a special CTS frame. These special RTS/CTS frames mean that duration will not continue any more. So other stations that receive these frames understand that they must remove this special duration and its turn. The RTS duration field is calculated by

$$= \text{RTSNumber} * (\text{RTS}_{\text{time}} + \text{CTS}_{\text{time}} + 2\text{SIFS}) (\text{RTS}_{\text{time}}).$$
(6)

This process will be repeated RTSNumber times to assure that all the listening stations in the WLAN receive at least one RTS or CTS frame. In as much as the RTSNumber depends on the BER, it is possible to set its value based on the probability of missing RTS or CTS by one station. It is possible for this value to be less than a special limit. A WLAN with *m* station which has an active flow and RTSNumber gives this probability through (7) to (9)

$$P_{\text{RTS/CTS}}^{f} = 1 - (1 - \text{BER})^{\text{RTS}_{\text{Length}}} \approx \text{RTS}_{\text{length}} * \text{BER}, \quad (7)$$

$$P_{\text{Station}}^{f} = \left(P_{\text{RTS/CTS}}^{f}\right)^{\text{RTSNumber}},\tag{8}$$

$$P^{f} = 1 - (1 - P^{f}_{\text{Station}})^{n-1} \approx (n-1) * P^{f}_{\text{Station}},$$
 (9)

where  $P_{\text{RTS/CTS}}^{f}$  is the corruption probability of an RTS or CTS frame,  $P_{\text{Station}}^{f}$  is the probability that a station does not receive any of the sent RTSs or CTSs, and  $P_f$  is the probability of not receiving even one RTS/CTS by a station among m - 1 listening stations. By using Table 4 for RTS<sub>length</sub> and assuming that BER is equal to  $10^{-5}$  [1],  $P_{\text{RTS/CTS}}^{f}$  will be 0.00224. The power factor in (8) is RTSNumber rather than 2\*RTSNumber because in severe situations the listening station may only receive either RTS or CTS because of being in the signal range of either RTS transmitter or CTS transmitter. For a WLAN with 200 active flows and RTSNumber = 3, the  $P_f$  is equal to  $2.2*10^{-6}$ .

#### 3.3. AFDSA scalability

Since the AFDSA is a distributed algorithm, it has a good scalability. It can accept new stations until the channel saturates, or there is no bandwidth to assign. Since new stations only compete for TXOPs in free time, as depicted in Figure 8, no new station can reach channel if there is not any free time available. So if the number of stations is increased, network is accessible only for the number that can send their packets with adequate quality of service. It means by using AFDSA, a station can either send their packets with proper QoS or cannot have access to the channel for sending its packets. Perhaps it seems to be an unfair algorithm. However, the authors think assigning the network channel to a limited number of stations by good QoS parameters is better than sharing the channel among a large number of dissatisfied QoS stations.

TABLE 2: Scenario 1 nodes and traffic flows.

Node	Application	Arrival period (ms)	Packet size (bytes)	Sending rate (kbps)
1→6	Audio	4.7	160	64
$7 \rightarrow 12$	VBR video	≈26	≈660	$\approx 200$
13→18	MPEG4 video	2	800	3200

TABLE 3: Traffic specification.

Traffic type	Priority	$CW_{Min}$	$CW_{Max}$	Max delay (ms)
Voice	6	7	15	50
VBR video	5	15	31	100
CBR video	4	15	31	100

TABLE 4: The PHY and MAC layer parameters.

SIFS	16 µs	CCA Time	$4 \mu s$
DIFS	34 µs	MAC header	38 Bytes
ACK size	14 bytes	PLCP header length	4 bits
PHY rate	36 Mp/s	Preamble length	20 bits
Minimum bandwidth	6 Mp/s	RTS length	28 bytes
Slot time	9 µs	CTS length	28 bytes

# 3.4. AFDSA overhead

AFDSA sends RTSNumber RTS/CTS in addition to the packets that must be sent. AFDSA overhead can be calculated by

$$O_{\text{Total}} = [\text{RTSNumber} * \frac{1\text{Sec}}{\text{SI}} * \text{NumberofTotalFlows}$$

$$* (\text{CTS}_{\text{Time}} + \text{SIFS} + \text{RTS}_{\text{Time}} + \text{SIFS})],$$
(10)

where  $RTS_{Time}$  is the time required to transmit RTS packet,  $CTS_{Time}$  is the time required to transmit CTS packet, RTSNumber is defined in Section 4.1, SIFS is the time between two successive transmissions as depicted in Table 1. Since AFDSA sends RTS/CTS packets only at the beginning of each transmission, one second is divided by SI to find the number of SI repeats in one second. Also the total number of flows, NumberofTotalFlows, is calculated by

Number of Total Flows = 
$$\sum_{i=1}^{n} f_i$$
, (11)

where  $f_i$  is the number of flows in the *i*th station. Since in our simulation RTS<sub>Time</sub> = CTS<sub>Time</sub> = 12 µs, SIFS = 16 µs, RTSNumber = 2, and NumberofTotalFlows is 18, the O<sub>Total</sub> found from (11) is equal to 40320 µs. This is the time that AFDSA consumes for transmission of RTS/CTS packets in one second (4 percent). As it is clear form (10) and (11), neither the number of stations nor the size of packets affects the overhead. Only the NumberofTotalFlows, selected SI and RTSNumber, can affect the overhead. It must be mentioned that the NumberofTotalFlows in a WLAN can be increased until the channel saturates. After that, increase in the number

TABLE 5. Just for uncreativity of trane in uncreat methods.						
	HCCA	FHCF	AFDSA	EDCA		
Audio	14.2 (ms)	14.5 (ms)	14.1 (ms)	0.9 (ms)		
VBR video	460.57 (ms)	14.7 (ms)	19.4 (ms)	3.2 (ms)		
CBR video	20 (ms)	15.1 (ms)	13.7 (ms)	22.5 (ms)		

 TABLE 5: Jitter for different types of traffic in different methods.

of stations cannot influence the Numberof TotalFlows since the new stations can not access the channel. If these new stations are able to access the channel in a saturated manner, it is impossible to guarantee QoS parameters for any flow.

# 4. SIMULATION RESULTS

AFDSA is implemented using NS-2 simulator and compared with the three previously reported works for the distributed [2–8] and centralized [9–15] channel access mechanisms. Both distributed (EDCA) and centralized methods (HCCA) of 802.11e [1] are selected since they are widely used in the literature for comparison. The fair HCF (FHCF) proposed in [12] is also selected as the third scheme to compare with our method. Two kinds of simulation scenarios have been used. The first one contains 18 sources and one destination. The second contains 6 sources and one destination. In both scenarios, the destination is QAP that contains a PC to satisfy the requirements for HCCA and FHCF, yet it is an ordinary QSTA for our proposed method.

# 4.1. Scenario 1

In scenario 1, 6 QSTAs send a high-priority on/off audio traffic (64 kbps) each, another 6 QSTAs send a VBR video traffic (200 kbps of average sending rate) with medium priority each, and 6 QSTAs send a CBR MPEG4 video traffic (3.2 Mbps) with low priority each. Voice traffic is used to indicate that AFDSA in the presence of the high-priority traffic is still able to give desirable QoS parameters for both CBR and VBR traffics. Table 2 summarizes the different traffics used for this simulation. We model the audio flow by on/off source with parameters corresponding to a typical phone conversation [17]. UDP is used as transport protocol.

Figures 9 to 12 demonstrate the latency distribution of the simulated methods. It shows that AFDSA has a maximum latency for each traffic stream under its tolerable latency (Table 3). In contrast, VBR traffic latency in HCCA is uncontrollable and in EDCA exceeds the limit. Maximum VBR latency for the proposed method is 80 milliseconds but for FHCF is 50 milliseconds. This might seem to be an advantage for FHCF but it must be considered that FHCF is a centralized method that needs PC where AFDSA is a distributed algorithm that does not need any PC. Also the AFDSA latency is still lower than the tolerable latency for VBR video. These differences relate to the starting situations. FHCF starts with the SI = 50 ms and continues by this yet AFDSA starts with the SI = 100 ms and then changes it to 50 ms. So, the grater SI belongs to the starting TS, that is, a VBR video stream which its maximum latency is 100.



FIGURE 9: Latency distribution for FHCF.



FIGURE 10: Latency distribution for standard HCF.

Therefore, AFDSA sets the currentSI to 100 then it changes it to 50.

The same figure also shows that the latency distribution curve of the VBR flow has a stair shape. This shape relates to the packets interarrival time. Analysis of the VBR video trace file shows that the interarrival time of packets is 34 milliseconds (see Table 2) but some packets are received



FIGURE 11: Latency distribution for AFDSA.



FIGURE 12: Latency distribution for EDCA.

simultaneously so the mean arrival time is 26 milliseconds. With 34 milliseconds interarrival time, the arrival times repeat with a period near 400 ms. Therefore, packets can only get some specific latency between 0 and 50 milliseconds which causes a stair-shape latency distribution curve.

Figures 13 to 16 show the latency during the time. As depicted in these figures, the latency of AFDSA and FHCF methods are better than others. HCCA has problem in VBR video traffic and EDCA has problem with CBR video. Also the jitter of different flows in different methods is summarized in Table 5. Although EDCA has very low jitter, it suffers from very high-dropped packet number as illustrated in Table 6.

Table 3 demonstrates the characteristics of the selected traffics. The different VBR flows have been obtained with



TABLE 6: Dropped-packet number for different methods.

	HCCA	FHCF	AFDSA	EDCA
Audio	0	0	0	0
VBR Video	0	0	0	0
CBR Video	108	91	101	6794

VIC video-conferencing tool using the H.261 coding and QCIF format for typical "head and shoulder" video sequence. The PHY and MAC layer parameters used in the simulation are also summarized in Table 4.

There are two methods to increase the channel load: increasing the node number and increasing the packet size. The latter is selected for increasing the channel load in this simulation. It is a time-consuming method because CBR video packets need more time to be transmitted. The packet size of CBR MPEG4 video has been increased from 600 to 1000 bytes to achieve 96% channel load.

Figure 17 shows fairness for VBR and CBR video traffic streams when the load increases up to 96%. In order to compare the fairness of the different schemes for the same kind of traffic, Jain's fairness index has been employed [18]:

$$J = \frac{\left(\sum_{i=1}^{n} d_{i}\right)^{2}}{n \sum_{i=1}^{n} d_{i}^{2}},$$
(12)

where  $d_i$  is the mean delay of the flow *i* and *n* is the number of flows. Figure 17 indicates that FHCF and AFDSA are fairer than HCCA.

# 4.2. Scenario 2

In scenario 2 (see Table 5), there are 7 nodes, six of which are sources and another is destination. Each QSTA has three



FIGURE 14: HCCA latency.



different traffic flows (audio, VBR H.261 video, and CBR MPEG4 video flows) simultaneously through three different MAC layer priority classes. We increase the channel load by increasing the packet size of CBR MPEG4 traffic from 600 bytes (2.4 Mbps) to 1000 bytes (4 Mbps) using a 100 bytes increment and keeping the same interarrival period of 2 milliseconds.



Mean fairness of the VBR flows versus channel load





FIGURE 17: Mean fairness for VBR and CBR flows.



FIGURE 18: Mean latency for different flows versus channel load.

Figures 18 and 19 show the mean delay and fairness of several types of flows, obtained with the various schemes, for different loads of network, respectively.

#### Audio and VBR H.261 video flows

Figure 18 shows that the delay is almost constant for the FHCF and the AFDSA with increase in the load which indicates that delay does not strongly depend on the network load. In HCCA, VBR traffic has a high value of delay (300 milliseconds) that exceeds the limit for this kind of traffic.

In EDCA, mean latency is very low for audio and VBR video traffic streams because of the high priority that had been assigned for these streams. This increases the delay of CBR video traffic and it linearly increases with increase in traffic load. Figure 19 shows Jain index for all four methods. These methods are almost similar for audio traffic. It is



FIGURE 19: Mean fairness of different flows versus channel load.

apparent that AFDSA and FHCF are better than EDCA and HCCA for VBR video traffic.

# CBR MPEG4 video flows

In our simulation, CBR streams are responsible for increasing the traffic load. As we can see in Figure 18, latency is almost constant for HCCA, FHCF, and AFDSA but increases with the load increment for EDCA such that for loads more than 79% it exceeds the limit for CBR traffic (100 ms). Figure 19 shows that Jain's index for all methods is high with minor differences for CBR video traffic.

#### 5. CONCLUSION

A new distributed MAC scheduling algorithm (AFDSA) for upcoming 802.11e standard is proposed and evaluated.

The mechanism introduces three additional fields to the RTS/CTS frame to guarantee QoS. The EDCA method of 802.11e is used to access the channel for the first time. When time duration is reserved for a station, the rest of the stations only compete for accessing the channel in the unreserved periods. It is shown through extensive simulation that the AFDSA can guarantee QoS for both CBR and VBR video traffic. It does not need any point coordinator and each node can play an access point role if it is connected to the backbone.

# REFERENCES

- IEEE 802.11 WG, "IEEE std. 802.11e, Part 11: Wireless MAC and physical layer specifications: MAC Quality of Service Enhancements," Reference number ISO/IEC 15802-3, November 2005.
- [2] M. Malli, Q. Ni, T. Turletti, and C. Barakat, "Adaptive fair channel allocation for QoS enhancement in IEEE 802.11 wireless LANs," in *Proceedings of IEEE International Conference* on Communications (ICC '04), vol. 6, pp. 3470–3475, Paris, France, June 2004.
- [3] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, "Distributed MAC adaptation for WLAN QoS differentiation," in *Proceedings* of *IEEE Global Telecommunications Conference (GLOBE-COM '03)*, vol. 6, pp. 3442–3446, San Francisco, Calif, USA, December 2003.
- [4] L. Romdhani, Q. Ni, and T. Turletti, "Adaptive EDCF: enhanced service differentiation for IEEE 802.11 wireless ad hoc networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '03)*, vol. 2, pp. 1373–1378, New Orleans, La, USA, March 2003.
- [5] G. W. Wong and R. W. Donaldson, "Improving the QoS performance of EDCF in IEEE 802.11e wireless LANs," in *Proceedings of IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing (PACRIM '03)*, vol. 1, pp. 392–396, Victoria, Canada, August 2003.
- [6] W. Pattara-Atikom, S. Banerjee, and P. Krishnamurthy, "A-DRAFT: an adaptive QoS mechanism to support absolute and relative throughput in 802.11 wireless LANs," in *Proceedings of the 7th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '04)*, pp. 117–125, Venezia, Italy, October 2004.
- J. F. Lee, W. Liao, and M. C. Chen, "A MAC-layer differentiated service model in IEEE 802.11e WLANs," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol.
   6, pp. 3290–3294, St. Louis, Mo, USA, November-December 2005.
- [8] Q. Deng and A. Cai, "A TXOP-based scheduling algorithm for video transmission in IEEE 802.11e networks," in *Proceedings* of the 6th International Conference on ITS Telecommunications (ITST '06), pp. 573–576, Chengdu, China, June 2006.
- [9] C. Liu and C. Zhou, "Providing quality of service in IEEE 802.11 WLAN," in *Proceedings of the 20th IEEE International Conference on Advanced Information Networking and Applications (AINA '06)*, vol. 1, pp. 817–822, Vienna, Austria, April 2006.
- [10] E.-C. Park, D.-Y. Kim, C.-H. Choi, and J. So, "Improving quality of service and assuring fairness in WLAN access networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 4, pp. 337–350, 2007.
- [11] J. Jackson Juliet Roy, V. Vaidehi, and S. Srikanth, "A QoS weight based multimedia uplink scheduler for IEEE 802.11e

WLAN," in Proceedings of the International Conference on Signal Processing Communications and Networking (ICSCN '07), pp. 446–451, Chennai, India, February 2007.

- [12] P. Ansel, Q. Ni, and T. Turletti, "FHCF: a simple and efficient scheduling scheme for IEEE 802.11e wireless LAN," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 391–403, 2006.
- [13] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE802.11E networks," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 36–43, 2003.
- [14] G. Boggia, P. Camarda, L. A. Grieco, and S. Mascolo, "Feedback-based bandwidth allocation with call admission control for providing delay guarantees in IEEE 802.11e networks," *Computer Communications*, vol. 28, no. 3, pp. 325– 337, 2005.
- [15] B. A. Venkatakrishnan and S. Selvakennedy, "An enhanced HCF for IEEE 802.11e wireless networks," in *Proceedings of the 7th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '04)*, pp. 135–142, Venezia, Italy, October 2004.
- [16] S. Montazeri, R. Berangi, and M. Fathy, "A new distributed scheduling algorithm to guarantee QoS parameters for 802.11e WLAN," in *Proceedings of the International Conference* on Information Networking (ICOIN '06), vol. 3961 of Lecture Notes in Computer Science, pp. 132–145, Sendai, Japan, January 2006.
- [17] P. M. Soni and A. Chockalingam, "Performance analysis of UDP with energy efficient link layer on Markov fading channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 769–780, 2002.
- [18] R. Jain, The Art of Computer Systems Performance Analysis, John Wiley & Sons, New York, NY, USA, 1991.