

RESEARCH

Open Access

Towards ubiquitous video services through scalable video coding and cross-layer optimization

Tiia Sutinen^{*}, Janne Vehkaperä, Esa Piri and Mikko Uitto

Abstract

Video content as one of the key features of future Internet services should be made ubiquitously available to users. Moreover, this should be done in a timely fashion and with adequate support for Quality of Service (QoS). Although providing the required coverage for ubiquitous video services, wireless networks, however, pose many challenges especially for QoS-sensitive video streaming due to their inadequate or varying capacity. In this article, we propose a cross-layer video adaptation solution, which may be used for optimizing network resource consumption and user experienced quality of video streaming in wireless networks; thus improving the availability of video services to mobile users. Our solution utilizes the flexibility of the Scalable Video Coding (SVC) technology and combines fast and fair Medium Access Control (MAC) layer packet scheduling with long-term application layer adaptation. The proposed solution both improves the usage of network resources by dropping video data based on its priority when the network is congested but also reduces efficiently the number of useless packet transfers in a congested network. We evaluate our solution with a simulation study under varying network congestion conditions. We find that already application layer adaptation gains over 60% less base layer losses, momentous for SVC video decodability and quality, than in the case without any adaptation. When our MAC layer scheduling is enabled, nearly a zero loss situation with respect to packet losses carrying base layers can be attained, resulting in peak-signal-to-noise ratio values very close to the original.

Keywords: SVC, adaptation, TCP friendly rate control, MAC layer scheduling, IEEE 802.11e

1 Introduction

The future Internet should ensure seamless and ubiquitous access to media through heterogeneous networks and terminals by implementing dynamic scalability across the whole delivery chain. Media, and especially video traffic, is expected to dominate the Internet traffic growth while the main access method is shifting from wired to wireless, as indicated by Cisco [1]. This trend creates problems for Broadband Wireless Access (BWA) operators as the ever-increasing traffic loads [1] can no longer be handled efficiently with today's technology. Moreover, the wireless medium has its own challenges for supporting Quality of Service (QoS) sensitive services such as video streaming due to its fluctuating capacity. Thus, there is a demand for new solutions for

the service providers and BWA operators to ensure the required level of QoS to their customers while providing them access to their favourite Internet services anytime and anywhere.

In the case of video streaming, the inability of wireless networks to guarantee the required bandwidth and QoS for the services has boosted the development of novel video coding and adaptation solutions to improve the robustness and QoS for video transmission. For instance, the novel Scalable Video Coding (SVC) technology [2] provides both bitrate and device capability adaptation, which are especially useful in heterogeneous network environments. In addition, several algorithms and protocols for controlling the video stream bitrate to match the available network capacity have been proposed in the literature. The typical solution of adapting video streams in the application layer has been studied, for instance, in [3,4]. In this case, video bitstream adaptation takes place

* Correspondence: tiia.sutinen@vtt.fi

VTT Technical Research Centre of Finland, Vuorimiehentie 3, Espoo, 02044 VTT, Finland

in the server or an intermediate network node and the decision-making relies on client feedback information of the streaming performance (e.g., delay and loss metrics). However, due to this very feedback signaling requirement, application layer adaptation is not responsive enough to quick wireless link capacity fluctuations.

To overcome the deficiencies of application layer adaptation solutions in wireless networks, several proposals for adapting video streams in the data link layer have appeared during the recent years (e.g., [5-9]). Medium Access Control (MAC) layer video adaptation employs selective packet discarding and prioritized transmission in order to ensure that the most important video packets get transmitted over the wireless link with the highest probability. The MAC-level solutions, however, regulate video streaming only in the scope of the wireless link, thus potentially wasting transmission resources in the wired core network. Therefore, it can be acknowledged that local adaptation within a single system layer is not the most efficient way to achieve dynamic scalability, but cross-layer solutions should be used instead.

In this article, we propose an architecture and implementation approach for cross-layer adaptive video streaming required for ubiquitous video stream delivery. Our solution relies on the SVC technology for implementing wireless bandwidth-adaptive video streaming services without adding any extra redundancy to the streaming. We present an end-to-end architecture for scalable video transmission enhanced with different cross-layer signaling and adaptation capabilities. Our solution is based on the OPTIMIX system architecture [10], which supports novel controlling modules for cross-layer optimization as well as a signaling framework for transmitting timely cross-layer context information within the video streaming system.

Of the diverse cross-layer optimization approaches supported by the OPTIMIX architecture for video streaming, we focus on considering the solutions for application layer (i.e., source) as well as MAC layer bitrate adaptation of SVC-encoded video streams. The application layer adaptation is implemented using a TCP Friendly Rate Control (TFRC) based adaptation algorithm [3], and the corresponding feedback information delivery is realized using the cross-layer signaling framework supported in the OPTIMIX architecture [10]. TFRC based adaptation performs well in terms of TCP friendliness and smoothness and it is well suited for multimedia applications. Since we are primarily considering IEEE 802.11 WLAN networks in this paper, our proposed MAC layer adaptation employs an IEEE 802.11e Enhanced Distributed Channel Access (EDCA) [11] based solution for MAC-level video packet differentiation. EDCA is a standardized mechanism for implementing distributed QoS management for WLANs. In our

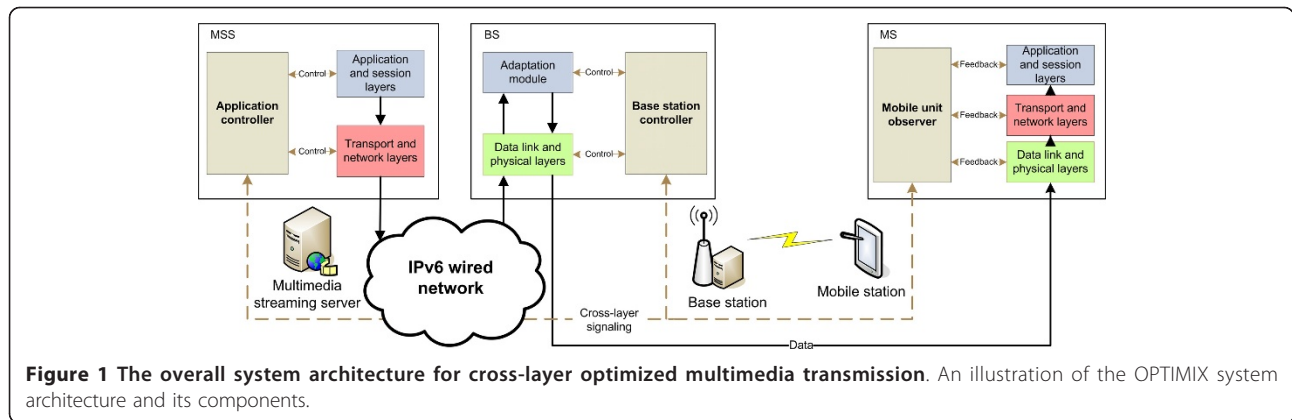
system, we extend the standard EDCA queuing and scheduling solution by adding extra video queues and a video scheduler [12] to achieve differentiated treatment for SVC video packets. We also show the advantages of the proposed system with a simulation study conducted in the OMNeT++ environment [13]. The results show how the application- and MAC-level adaptation complement one another in optimizing the Quality of Experience (QoE) for the video user and saving network as well as terminal resources by reducing the number of useless transmissions under congestion.

Cross-layer optimization approaches with distributed video adaptation have been studied earlier in the literature to some extent. For example, the authors of [14] propose a communications architecture which utilises both application and MAC layer adaptation. However, this work proposes to change the transmission rate at the radio/MAC level instead of buffering and uses non-scalable video coding instead of SVC, which provides more advanced adaptation capabilities. The article [15], on the other hand, proposes a cross-layer approach for congestion control of real-time video. The authors indicate that the fairness in the usage of network resources together with application layer adaptation increase the resource usage efficiency. But also this work does not include SVC. Thus, the novelty in our work lies in the usage of SVC in implementing cross-layer optimized video streaming.

The rest of the article is organized as follows: First, we introduce the OPTIMIX system architecture, which forms the basis for the SVC optimization with its cross-layer controlling and signaling capabilities. Second, we discuss the optimization approaches designed for SVC transmission, namely the application and MAC layer adaptations. Third, we introduce the simulation model and scenario for evaluating the proposed solution along with selected results. Finally, we give the conclusions with some insights of the future work.

2 System architecture

The overall OPTIMIX system architecture is depicted in Figure 1. The architecture consists of routers and nodes which use both wireless and wired connections for communications and where the last hop is assumed to be wireless for the application scenarios of our interest. In addition to the traditional network blocks, namely the Multimedia Streaming Server (MSS), the Base Station (BS), and the Mobile Station (MS), the architecture includes both data link/physical and application layer controller units and a cross-layer signaling system providing the controllers timely information regarding the changing network and channel conditions. Although the OPTIMIX architecture supports a multitude of optimization mechanisms for multimedia transmission, in this article, we will



focus on discussing only the most relevant features for optimizing SVC streams over wireless networks. For a more complete overview of the whole architecture, the reader is advised to refer to [10].

In this section, we will shortly introduce different aspects of the OPTIMIX architecture as it acts as a framework for the cross-layer optimization approaches proposed later in this article. Cross-layer optimization means that several different system layers are optimized together, and in order to see how this affects to the overall system performance, it is important to first get a picture of the whole system involved in our study.

2.1 Application and session layers

The application layer of the OPTIMIX architecture enables the usage of different video coding schemes such as H.264/AVC and SVC. For session initialization and maintenance, Real Time Streaming Protocol (RTSP) and Real-time Transport Control Protocol (RTCP) are used. In this article, we will focus on the transmission of pre-encoded SVC videos.

SVC, the scalable extension of the H.264/AVC standard (Annex G), is a novel video coding standard with build-in features optimal for adaptive video transmission. SVC provides both bitrate and device capability adaptation which are desirable features especially in error-prone wireless heterogeneous networks [2]. The term scalability in the video coding context means that physically meaningful video information can be recovered by decoding only a portion of the compressed bitstream. The following options are possible, separately or combined:

- *Temporal (framerate) scalability*: complete pictures can be dropped from the bitstream and the stream can still be decoded.
- *Spatial (resolution) scalability*: video is encoded at multiple spatial resolutions.

- *SNR (quality) scalability*: video is encoded at different levels of fidelity.

An SVC bitstream consists of a Base Layer (BL) and one or several Enhancement Layers (EL). The BL is compliant with non-scalable H.264/AVC, which means that it is decodable also by legacy devices that do not have support for SVC. SVC bitstream consists of a sequence of Network Abstraction Layer (NAL) units identified by a NAL unit header. Different types of NAL Units (NALU) are defined by the standard and NALUs can carry video data, information about parameter or decoding settings and additional information useful for the decoder (SSEI NALUs). The content of the NALU payload is identified by the NALU type and the NALUs, which include SVC video data, are identified with three additional bytes defining the layer in which this specific NALU belongs to. This layer information included in the header can be utilized during the adaptation process to easily identify the layer of each NALU. Adaptation of the SVC stream is computationally lightweight since it can be performed by truncating the original stream without the need for transcoding. Scalable video needs to be encoded only once at the highest resolution or with the best quality, but multiple scalable sub-streams can be decoded depending on the target characteristics [16].

Even though SVC and the OPTIMIX architecture support application layer adaptation of SVC bitstreams in different locations (i.e., the server or a media aware network element), in this article, we assume that the application layer adaptation takes place in the server (i.e., the source). The entity controlling the adaptation is the application controller, presented later in this section with more details.

In the MS, the application layer implements error concealment techniques for robustness against packet losses [17]. The target of this work is not to measure

the effectiveness of the error concealment but to ensure a stable decoder with three supported methods: frame copy is implemented into the decoder as an error concealment technique in order to cope with the packet losses in the BL. Additionally, to avoid severe error propagation, pixel-value interpolation is used for the missing I-slices in order to provide smaller error drifting. Finally, error concealment for the EL slices is done by concealing the missing slice by using the lower quality slice. For further details on how different error concealment algorithms perform in an error-prone transmission environment, please refer to [17].

2.2 Transport and network layers

The transport and network layers are based on a traditional multimedia streaming protocol stack, including the RTP, UDP, and IP protocols, and no modifications has been introduced to these layers. RTP provides end-to-end transport functionalities for multimedia transmission and supports different data formats such as the generic H.264/AVC and SVC payload formats. The network layer is based on the IPv6 protocol. In our work, we use the RTP/UDP/IPv6 protocols for SVC video transmission.

2.3 Data link and physical layers

The data link and physical layers supported by the OPTIMIX architecture do not strictly follow any specific standard and multiple transmission schemes can be introduced. The reasoning behind this is that different cross-layer techniques can be evaluated more easily with different access schemes, channel codes, and modulation schemes when the restrictions of certain standards are not valid. In this article, we focus on investigating IEEE 802.11 *g* and *e* [11] based WLAN MAC technologies built on top of a proprietary physical layer. The entity controlling the data link and physical layer functions for optimizing multimedia transmission in the proposed system architecture is the BS controller, introduced in the next subsection. In addition, more details of the data link layer structure are provided later in this article.

2.4 Cross-layer control and signaling

An efficient signaling solution is crucial for the success of cross-layer adaptation and control. The cross-layer and end-to-end signaling solution used in the proposed architecture is based on the triggering framework introduced in [18] as well as the IEEE Media Independent Handover (MIH) services standard [19]. The triggering framework is used for transferring cross-layer signals both locally, that is, between entities located on the different layers of the local protocol stack, and remotely, between entities in the different network nodes. MIH, on the other hand, provides standardized link level

signaling support between the MS and the wireless access network. The proposed cross-layer signaling system is illustrated in Figure 2. A more detailed overview of the integration of the triggering framework with MIH can be found from [20].

The control of the cross-layer adaptation is divided into two separate units in the OPTIMIX architecture, namely the application controller and the BS controller. The reason behind the separation is to split the overall cross-layer adaptation into two sub-units which can react to either long-term or rapid variations in the transmission environment [10]. The controlling units can cooperate when optimizing the transmission parameters by using the proposed cross-layer signaling solution.

The application controller is used for adjusting the properties of video (e.g., bitrate and frame rate) to the prevailing transmission conditions. The aim is to utilize the available transmission capacity as efficiently as possible while trying to maximize the viewing experience of the end-user. The application controller adjusts the adaptation process on a regular time interval (e.g., once in a second). In this article, we focus on investigating the case of rate adaptation of pre-encoded SVC streams where the application controller chooses the appropriate bitrate for the streaming based on client feedback information. The decision is then enforced by a specific SVC adaptation module that adds or drops SVC layers from the stream accordingly.

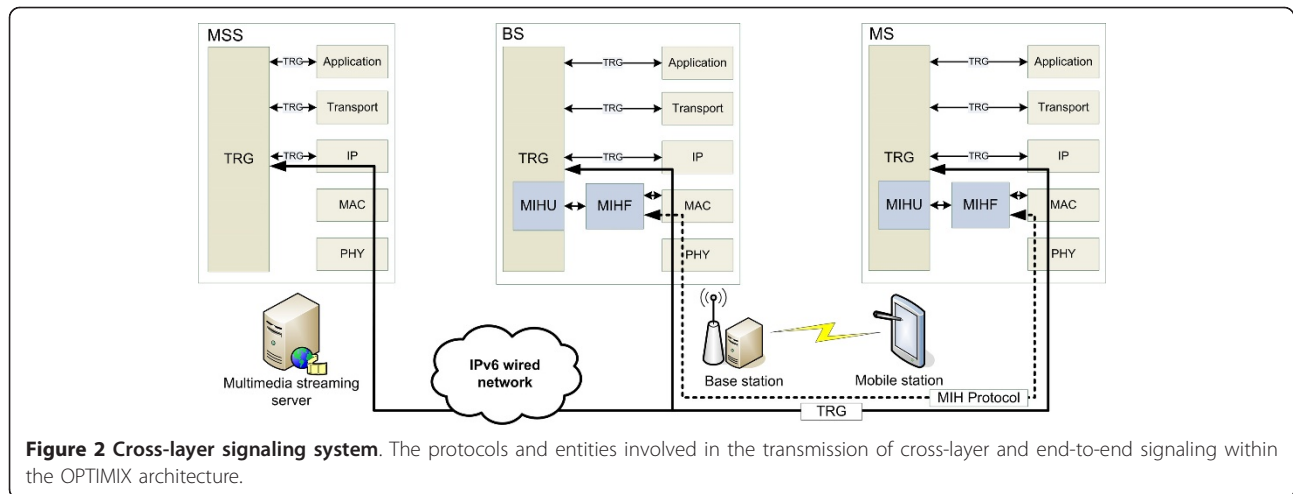
The BS controller may be used for controlling the faster adaptation of video streams and efficient usage of radio resources by utilizing scheduling as well as adaptively selecting the coding and modulation parameters. In this article, we focus on improving scalable video delivery across a wireless link through cross-layer enhanced MAC layer QoS techniques and use the BS controller to trigger the data link layer SVC adaptation on and off dynamically.

2.5 Summary

This section provided an overview of the OPTIMIX system architecture. This architecture with its controllers and cross-layer signaling system provides a framework for the cross-layer optimization solution and corresponding algorithms proposed in the next section. Moreover, the different system components and protocols discussed in this section are included in the simulation model used for evaluating the developed SVC optimization solutions and presented later in this article.

3 Cross-layer optimization of scalable video streaming

In order to adapt SVC streams efficiently to wireless network conditions, our system implements application



and MAC level adaptation in adjusting the SVC stream bitrate according to prevailing transmission conditions. The individual components and functions for the cross-layer adaptation are discussed in this section. The solution builds on the OPTIMIX architecture introduced with details of the entities and protocols deployed at each layer of the transmission architecture in the previous section.

3.1 Application layer control and adaptation

The purpose of the application layer control and adaptation is to handle the long-term adaptation of the SVC-encoded video bitstream in the end-to-end scope. The adaptation process at the application layer can be divided into two sub-processes: the adaptation decision-taking process and the bitstream adaptation process. The first defines the adaptation parameters according to an adaptation algorithm and the second performs the actual adaptation. In the OPTIMIX architecture, the application controller takes the adaptation decision while the bitstream adaptation is handled by a separate SVC adaptation module.

3.1.1 Adaptation decision-taking process

The adaptation decision-taking process performs the optimization of a given adaptation problem. The goal of this process is to find a set of parameters for the bitstream adaptation which do not violate the constraints of the system and content and which achieve a set of objectives given to the adaptation process. In our system, the adaptation decision-taking process aims to adapt an SVC bitstream to match the available bandwidth in the end-to-end transmission path as well as to react to its changes.

The application controller estimates the available bandwidth using TFRC based bandwidth estimation [4], in which the available bandwidth is calculated using Equation 1:

$$T = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RTO} \left(3\sqrt{\frac{3p}{8}} \right) p(1 + 32p^2)} \quad (1)$$

where T is the proposed data rate for the multimedia stream, s is the packet size, R is round-trip time, p is the packet loss rate, and t_{RTO} is TCP retransmission time. t_{RTO} can be estimated by setting it to $4R$ and the packet size s can be considered fixed (e.g., 1,450 bytes) since fixed size segments are expected to be used [4]. Round-trip time R is also considered as an estimate calculated in relation to the packet loss rate value.

The packet loss rate p is calculated in the client by monitoring the number of the received packets. The client calculates the packet loss rate periodically (e.g., every 0.5s) neglecting zero values and sends it to the application controller located at the MSS via the OPTIMIX signaling architecture. In addition, the client applies a weight to the calculated current and earlier packet loss rate values using Equation 2:

$$\sum_{k=0}^4 E_k \times M_k, \quad M_k \in \{0.45, 0.3, 0.1, 0.1, 0.05\} \quad (2)$$

where E denotes the erroneously received packets in the client. M is a coefficient to accentuate the most recent values. As indicated, for example, in [21], the use of weighted coefficients for previous values increases the accuracy and reliability of predictive events. Furthermore, in order to remove large fluctuations from the estimated data rate provided by TFRC, an average of four last data rate estimates is calculated and used as the target bitrate. The available bandwidth estimates calculated using the TFRC algorithm can vary quite a lot since TFRC reacts easily to even small levels of packet loss. The target bitrate calculated by TFRC is utilized in

the bitstream adaptation process on a regular time interval (e.g., once in a second).

We have selected to use a TFRC based adaptation algorithm because it is well suited for multimedia applications and it provides smooth but responsive controlling without high throughput variations that are possible with other controlling schemes [4]. TFRC is also standardized by IETF and can be considered to co-exist with other existing protocols in the Internet.

3.1.2 Bitstream adaptation process

The bitstream adaptation process for scalable videos utilizes the scalability features of SVC streams in order to adapt the video stream according to the application controller's decision. Firstly, the SVC priority information contained in the NALU header is a key tool for SVC bitstream adaptation. Any part of an SVC bitstream can be simply extracted by removing NALUs based on these parameters. The priority information is contained in specific fields of the SVC NALU header, namely the *dependency_id*, *temporal_id*, *quality_id* (QID), and *priority_id*, indicating the spatial, temporal, quality, and priority layers the SVC NALU belongs to, respectively. Secondly, the Scalable Supplemental Enhancement Information (SSEI) NALUs can be used for defining the bitrates and other characteristics of the SVC stream and its sub-streams. The SSEI information can be sent in-band as well as out-of-band.

Our SVC bitstream adaptation solution relies on the above mentioned parameters when adapting the bitstream according to the adaptation decision. The bitstream adaptation extracts NALUs from the video stream and finds the spatial, temporal, and quality id values for each packet along with the bitrates for each layer. The bit and frame rates for each layer can be obtained from the SSEI message. The bitstream adaptation process then simply compares the target bit rate calculated using TFRC to the layer bitrate and drops the EL(s), if needed.

3.2 MAC layer adaptation

The MAC layer adaptation of SVC is handled with an EDCA based MAC QoS architecture that provides SVC packet prioritization and adaptation under poor link conditions (e.g., congestion) in the BS. The MAC layer adaptation allows scaling the video bitrate to the prevailing transmission conditions of the wireless link in a timely manner. Moreover, to save processing in the BS, the BS controller can be used to trigger the MAC-level SVC adaptation on and off dynamically based on the wireless link status information collected using the OPTIMIX cross-layer signaling architecture.

3.2.1 MAC QoS architecture

The proposed MAC layer QoS architecture for SVC optimization builds upon a standard QoS mechanism that supports multiple transmission queues and

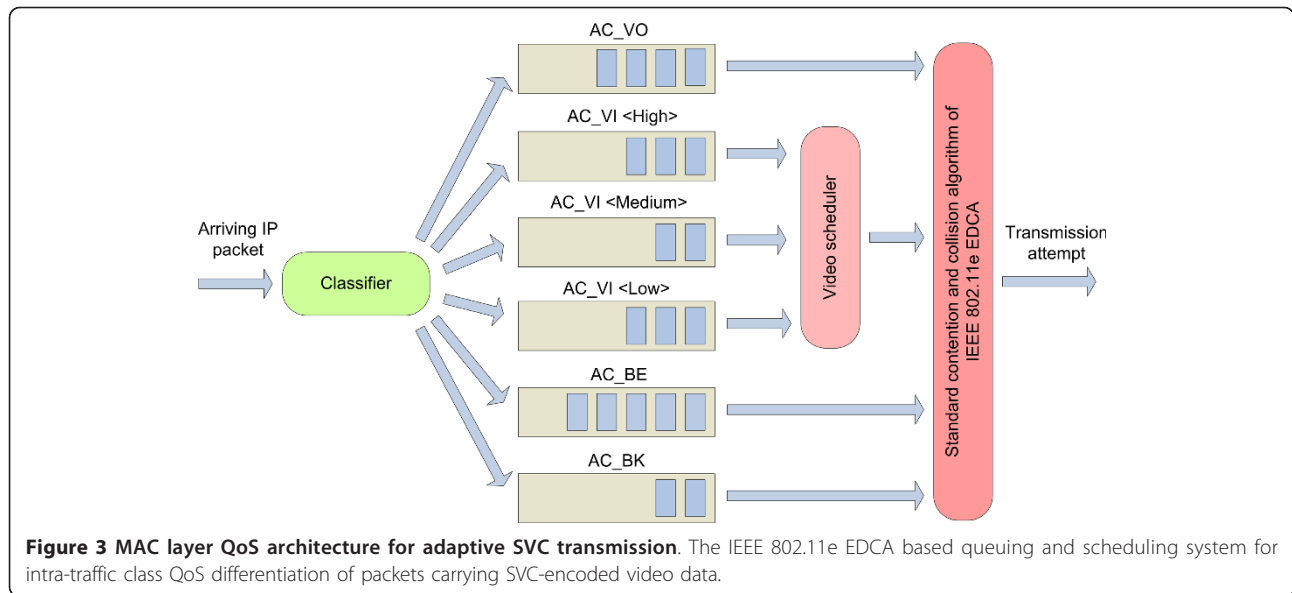
prioritized medium access for different types of traffic [12]. In this article, we consider IEEE 802.11 WLAN as the wireless access technology and thus we have selected the IEEE 802.11e EDCA [11], as the basis for our MAC layer QoS solution. EDCA is a distributed QoS management mechanism, which provides differentiation of voice (VO), video (VI), best effort (BE), and background (BK) traffic across a WLAN link. As discussed in [12], our MAC layer QoS architecture extends the standard approach by adding extra queues to the EDCA video access category (AC_VI) and introducing a second scheduler for selecting video packets for transmission. That is, our solution does not change the EDCA operation as such but extends it to allow for prioritizing more important video packets over lower important ones while leaving EDCA's inter-traffic class QoS differentiation and access control untouched. In this section, we focus on discussing the proposed extensions, and the reader is asked to refer to [11] for more details on EDCA operation and the handling of different types of traffic.

Figure 3 illustrates the MAC layer QoS architecture built on top of EDCA. The architecture supports three queues in AC_VI, each having a different priority: high, medium, and low, and a video scheduler, which selects video packets from the queues accordingly for transmission.

The prioritized video packet handling is accomplished by placing the incoming video packets into the video queues based on their type (e.g., SVC BL or EL) with a packet classifier, and triggering the video scheduler every time the video category is granted access to the medium by EDCA. The number of additional video queues used depends on the level of video packet differentiation required in the system. When invoked, the video scheduler selects the next video packet for transmission considering factors like the type of each Head of Line (HOL) video packet, the time the packet has spent in the queue, and whether there are higher priority video packets currently stored in the queues.

The packet classifier needs to access additional information to detect the types of the incoming higher layer packets. For this, the DiffServ Assured Forwarding (AF) per-hop behaviours [22] can be utilized as discussed in [12]. An example of a viable cross-layer packet-marking scheme for a three-layer quality scaled SVC stream is given in Table 1. Here the priority (i.e., layer) of a video frame is indicated by the value in the QID field of the SVC NALU header. During the streaming, the QID values are mapped to DiffServ priorities and onwards to EDCA AC_VI priority queues.

Besides supporting prioritized transmission of packets within AC_VI, our MAC QoS solution implements active queue management. Video traffic typically has strict QoS requirements in terms of end-to-end delay as frames arriving past their playout deadline are useless to



the receiver. Thus, the queue management discards video packets from the queues, if a specific maximum queuing time (e.g., 500 ms) is passed.

Finally, during congestion, the system implements priority-based discarding of video packets. Finite buffers are assumed, meaning that under congestion incoming packets that do not fit into their corresponding queue will be dropped. With priority-based discarding of video packets, the MAC always drops the packets belonging to the least important SVC layer first to fit the more important layers into the queues. This way, the MAC ensures that the most important BL packets have the highest probability to get transmitted, thus helping to maintain service continuity for the client. In case two or more ELs are placed into the same queue, the lower importance packets need to be marked as drop eligible, as discussed in [12], to allow priority-based discarding.

3.2.2 Video scheduler

The video scheduler selects video packets for transmission from the different MAC-level video queues whenever AC_VI is granted access to the medium. The scheduler implements an algorithm that factors in tolerated queuing delays for each HOL video packet. That is, if the HOL packet in a higher priority queue can wait for transmission, the HOL packet from a lower priority queue will be sent first (if exists and older). Figure 4

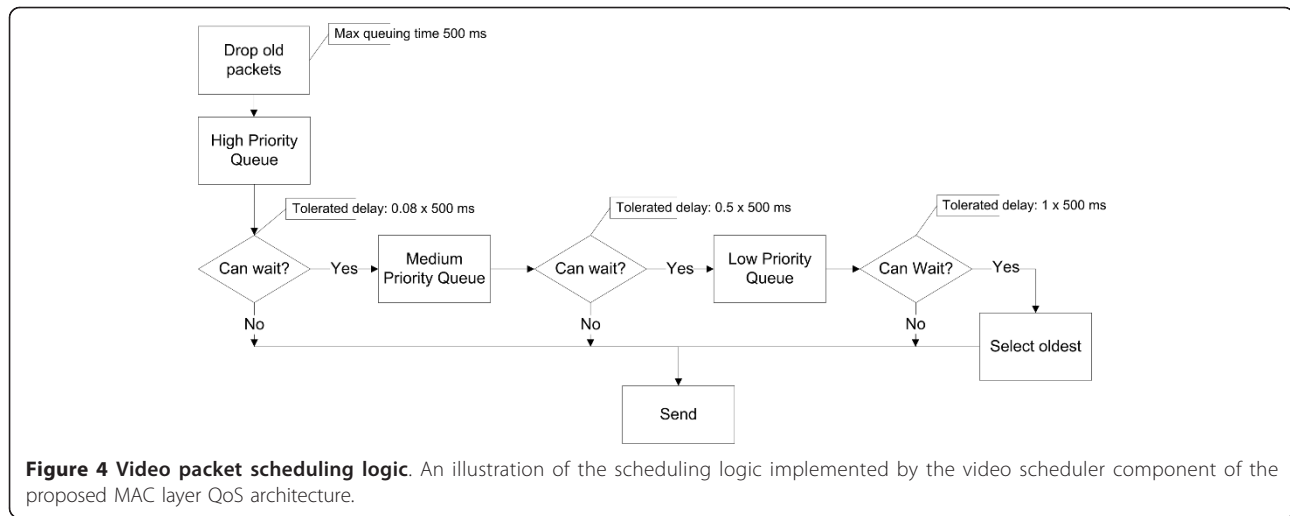
illustrates the video scheduler operation in a three queue case. The tolerated queuing delays thus do not refer to the actual time a packet spends in the queues at maximum but instead they are weighted values of the maximum queuing delay allowed for the video packets (e.g., 500 ms). Whenever the maximum queuing delay is exceeded for a HOL packet, it is discarded by the video scheduler. The weights are assigned so that the scheduler provides differentiated QoS for the video queues. We have chosen to use 0.08 as the weight for the high priority queue, and 0.5 and 1.0 as the weights for the medium and low priority queues, respectively, since they were tested to provide the desired performance.

3.2.3 Dynamic triggering of MAC layer SVC adaptation

In order to save processing in the BS, the BS controller is able to dynamically trigger the MAC-level video adaptation on whenever the link capacity starts deteriorating and off when the link capacity recovers. For this, the BS controller collaborates with the MS to constantly monitor the condition of the wireless link between them. The required signaling is supported with the low level signaling function of the OPTIMIX cross-layer signaling architecture (i.e., MIH). We utilize here the same approach as presented in [23], modifying some of the parameters and thresholds for the EDCA based QoS architecture. In this approach, the parameters monitored by the BS controller, namely the rate of video packets dropped due to buffer overflow or exceeding of retransmission limit and the video queue size, are AC_VI specific. This is to ensure that the MAC-level prioritized transmission and adaptation of SVC are triggered on based on the link conditions experienced by the video traffic. The parameters monitored and reported by the MS, that is, the Packet Error Rate (PER) and signal strength, and the thresholds

Table 1 Cross-layer packet QoS mapping for SVC differentiation

Priority	SVC QID	DiffServ class	EDCA AC	Video queue
High	0	AF41	AC_VI	High priority
↓	1	AF42	AC_VI	Medium priority
Low	2	AF43	AC_VI	Low priority



applied to them are kept the same as in [23]. All monitored parameter values are calculated from five previous parameter check rounds in order to see their trends and to ignore short-term peaks (see Equation 2). In addition, to ensure that the triggering is efficient enough when the AC_VI buffer is filling up too quickly, we apply an additional 75% threshold to the AC_VI maximum queue size in the BS side. Whenever the AC_VI buffer size exceeds this threshold, the MAC layer adaptation is triggered on immediately by the BS controller.

4 System evaluation and results

We evaluate the performance of the proposed cross-layer optimization solution for scalable video delivery in terms of an optimal network resource usage and QoE by exploiting a simulation environment developed in the FP7 ICT-OPTIMIX project. The simulator was built using the OMNeT++ [13] simulation framework. In this section, we present our experimental setup as well as selected simulation results to attest the advantages of our proposed solution for wireless video streaming services.

4.1 System model

The simulation model of the OPTIMIX system consists of a video server, an IEEE 802.11g BS and a MS, receiving a video stream from the server through the BS, and an IPv6 wired network connecting the BS and the server as depicted in Figure 1. Table 2 lists the most important parameters used in our simulations. In our study, the wired network does not introduce any packet loss or congestion. The video server and the MS use an RTP/UDP/IPv6 protocol stack for the SVC encoded video stream. The wireless physical layer simulates a log-normal shadowed uncorrelated Rayleigh fading channel with Additive White Gaussian Noise (AWGN) and without path-loss. The physical layer does not use adaptive coding and

modulation, but the modulation is kept fixed throughout the simulation run.

The wireless MAC implements the QoS architecture introduced in this article. The channel access for the four EDCA ACs is controlled using the default EDCAF parameters listed in Table 3. The default parameters were selected as we have only one client connected to the BS and our purpose is not to examine EDCA performance as such but the video packet differentiation and scheduling capability of our proposed MAC QoS solution. Each AC has a maximum queue length of 50 packets. No fragmentation is supported in the MAC but the incoming packets have already been fragmented to a 1,024B limit in the application layer. Request to send/clear to send is also not used. The video adaptation in the MAC is triggered on dynamically utilizing the signaling architecture. The triggering is event based and it can occur either if the video buffer size stays constantly (see Equation 2) beyond 50% of the maximum or exceeds 75% of the maximum or if the MS observes a PER exceeding 4% or the received signal strength being less than 70% of the maximum. When the MAC layer video adaptation is enabled, the maximum queuing delay allowed for video packets is 500ms and the tolerated queuing delays for the high, medium, and low priority video categories are 40, 250, and 500 ms, respectively.

Due to the novelty and advantages of the on-the-fly adaptation of SVC, the input video is a 500 frames (17s) long SVC-encoded sequence. The well-known *Foreman* and *Hall* were chosen for the test sequences. *Foreman* contains partitions of relatively static background, average motion, and a moving camera. On the contrary, *Hall* has a static camera with only a couple of slowly moving objects. We use the JSVM 9.15 reference encoder and the same version of the reference decoder extended with three error concealment algorithms [17].

Table 2 Simulation parameters

MAC parameters	
Retransmission limit	7
EDCA ACs	Four ACs with TX queues of size 50 frames each
Video TX queues in AC_VI	Three with differing priorities
PHY parameters	
Modulation	BPSK (FEC:1/3)
Bandwidth	20 MHz
Coherence time	0.1 s
Log-normal shadowing time	3 s
Shadowing standard deviation (σ)	4 dB
E_s/N_0 (1m distance from BS)	60 dB
Video parameters	
Video encoding	SVC video, BL + 2 quality enhancements
GOP length and structure	8 and IPPP...
Resolution	CIF (352 × 288)
Frame rate	30 Hz
Slices	Three in each layers' frame
BL bitrate (<i>Foreman; Hall</i>)	530; 300 kb/s
1st EL bitrate (<i>Foreman; Hall</i>)	1060; 1080 kb/s
2nd EL bitrate (<i>Foreman; Hall</i>)	2330; 2500 kb/s

The encoded video contains a BL and two quality ELs. The bitrate with the best quality is approximately 2.3 Mb/s for the *Foreman* sequence and slightly higher, 2.5 Mb/s, for *Hall*. Furthermore, the Peak-Signal-to-Noise Ratio (PSNR) difference is roughly 2dB between the layers. The resolution, common intermediate format (CIF), is kept constant for the both sequences, as is also the frame rate (30 Hz).

Before sending, the server tags the RTP packets containing SVC data with SVC layer information, which is accessible also at the MAC layer. In the MS, we use an RTP receiver buffer of 1 s to cope with jitter and the reception of unordered packets.

The TFRC based bandwidth estimation algorithm used in our simulations has been implemented into the application controller, which subscribes to the packet loss information trigger generated at the client. The client sends the packet loss trigger in every 0.5 s and the reported loss follows the weighted average according to Equation 2. In the TFRC calculation (Equation 1), we use a fixed packet size of 1,450B, which follows the average NALU size of the encoded SVC sequence. The used average Round-Trip Time (RTT) is an estimation between 50

and 70 ms depending on the received packet loss ratio. During the simulation, the average delays stay low when the traffic is low and the RTT is based on these values. The main factor of the TFRC calculation is the packet loss ratio whereas the small changes in RTT accuracy affect only little the bandwidth estimation. Finally, we calculate a weighted average of the TFRC based bandwidth estimate by weighting the previous four estimates with Equation 3:

$$\sum_{k=0}^3 T_k \times A_k, \quad A_k \in \{1, 0.5, 0.33, 0.25\} \quad (3)$$

where T_0 is the current bandwidth estimate given by Equation 1, T_1 the previous one, and so on. A_k is the weighting value for each of these estimates.

4.2 Simulation scenario

The signal strength stays good during the whole simulation run, and congestion is thus the main reason for packet drops in the BS MAC. However, a few packets may be lost due to failed transmissions caused by collisions or wireless transmission errors that could not be recovered with the allowed maximum of seven MAC layer retransmissions.

The SVC video streaming starts after the initial RTSP session signaling in the beginning of the simulation. The video traffic is allocated to AC_VI in the BS MAC. When MAC adaptation is enabled, the video packets are split into the three AC_VI queues by the packet classifier depending on which SVC layer they belong to.

Table 3 EDCA function parameters

AC	CWmin	CWmax	AIFSN	TXOP limit
VO	1	3	2	0
VI	3	7	2	0
BE	7	255	3	0
BK	7	255	7	0

We congest the wireless link by injecting UDP/IPV6 traffic into the link with the maximum transfer unit of 600B. The generated traffic starts at the point of 2 s with the throughput of 1.5 Mb/s and progressively increases every 2 s by 200 kb/s until the throughput of 2.3 Mb/s is reached. After that, the throughput decreases by 500 kb/s per 2s until it ends at the point of 14 s. In the BS MAC, the generated traffic is split to AC_BE and AC_VO categories in a 50-50% ratio to introduce real congestion into the wireless link and to cause packet losses to the video traffic. This way, we aim to show the benefits of the proposed cross-layer adaptation approach.

4.3 Main results

The evaluation was done for four different cases: no adaptation, MAC layer adaptation, application layer adaptation, and combined MAC and application layer adaptation. The test cases were run equally for the both video sequences for comparison. In this section, we present the main results that show how our cross-layer SVC optimization approach can be used to achieve better performance in terms of optimal network resource utilization and improved QoE for the user.

4.3.1 Results for network resource utilization

To evaluate network resource utilization efficiency of the different adaptation schemes, we investigate the number of packets dropped in the BS MAC during the simulation. The results presented here are averaged over ten simulation runs to ensure convergence.

The packets that affect the video quality the most are the BL ones. Figure 5 shows the number BL packets dropped by the BS MAC in the different cases for the *Foreman* and *Hall* sequences. Without adaptation, 516.8BL packets were dropped due to congestion, in average, in the case of *Foreman* and 408.9 in the case of *Hall*. The employed adaptation schemes lower this number significantly. The application layer adaptation reduces BL packet drops by over 60% for the both sequences: 64% for *Foreman* and 62% for *Hall*. Even more gain can be

achieved by enabling the MAC layer adaptation: when using only MAC layer adaptation, no BL packets were lost, and in the case of the MAC plus application layer adaptation, only 0.6 BL packets were lost, in average, for the *Foreman* sequence. For *Hall*, the corresponding figures are 0.5 for MAC adaptation and 0.9 for MAC plus TFRC adaptation. The higher loss rate in the combined case is due to the dynamic triggering of the MAC layer adaptation. The use of TFRC causes the MAC to temporarily stop adapting and when congestion builds up again, some BL losses may occur as the high priority queue gets congested. This is a place for improvement in our scheme in the future.

In any case, MAC layer adaptation produces performance gains. This is because when it is disabled, all video packets are treated with the same priority by the MAC layer and the BL packets get dropped in the same degree as the EL ones as can be seen in Figures 5, 6, and 7. When MAC layer adaptation is used, all or most of the BL and most of the 1st Enhancement Layer (EL1) packets are saved by prioritizing them over the 2nd Enhancement Layer (EL2) packets. The average number of EL1 packet drops in the case of MAC adaptation was as low as 0.4 and for MAC plus application layer adaptation 1.1 for *Foreman*; and 1.3 and 1 for *Hall*, respectively.

As discussed earlier in this article, the application layer adaptation provides a slow and long-term adaptation scheme. This can be clearly seen in Figures 5, 6, and 7. The application layer adaptation starts working with a delay and before that lots of video packets are dropped at MAC. Thus, we can state that the TFRC based adaptation improves the video transmission performance under wireless link congestion but is not perfect. Nevertheless, improving the performance of TFRC was not the focus of this article and was left for future work.

Figure 7 illustrates the dropped EL2 packets at the BS MAC. Here, we observe significantly larger drop rates with the adaptation schemes employing the MAC layer adaptation. In fact, the sole MAC layer adaptation results in about two times higher EL2 drop rate at the BS MAC

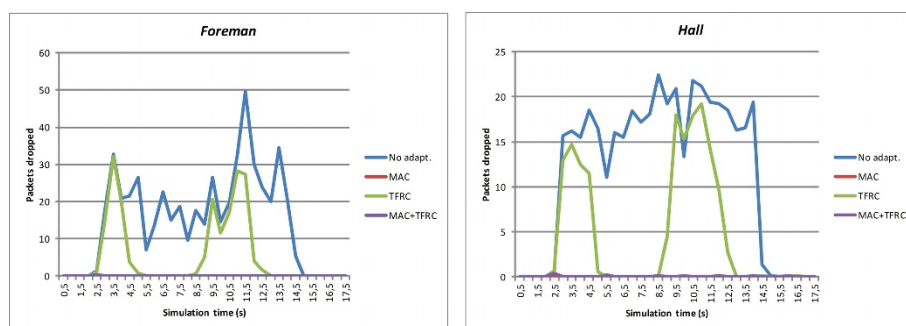


Figure 5 Simulation results for SVC BL transmission performance. Number of BL packets dropped in BS MAC for the two test sequences.

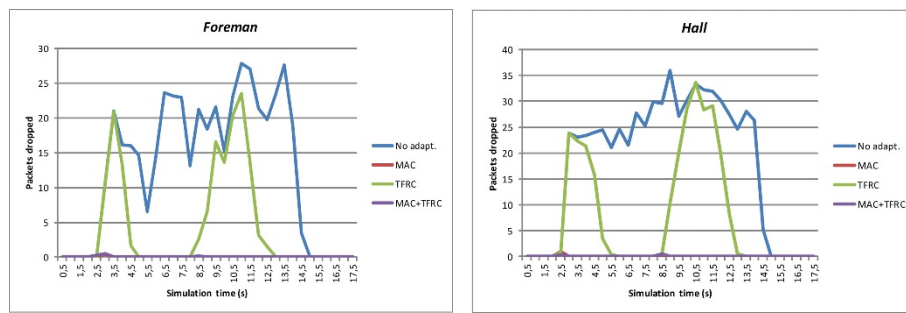


Figure 6 Simulation results for SVC 1st EL transmission performance. Number of 1st EL packets dropped in BS MAC for the two test sequences.

(1181.1 packets in average for *Foreman* and 1431.9 packets for *Hall*) than when no adaptation is enabled (585 packets *Foreman*, 744.2 packets *Hall*). This is because the BL and EL1 packets are prioritized over EL2 packets. The usage of application layer adaptation on top of MAC adaptation helps reducing the amount of EL2 packet drops in the BS, and thus improves the video transmission efficiency also in the wired network: The number of EL2 drops with sole application layer adaptation and MAC plus application layer adaptation for the *Foreman* sequence are in average 188.6 and 480.7, respectively. The corresponding figures for the *Hall* sequence are 286.8 and 628.8 EL2 packets.

Figure 8 summarizes the amounts of lost packets in the BS MAC. In the figure, we illustrate also the amount of lost interfering traffic that was generated into the network to induce congestion and of which 50% was allocated to AC_VO and 50% to AC_BE.

Although the EDCA based MAC always prioritizes video traffic over the BE one, it can be seen in Figure 8 that the video adaptation schemes considered in this article are fair to other types of traffic. That is, the video traffic gives away a share of its network resources to the other traffic under congestion when adaptation is enabled. Here, the schemes using TFRC perform the best as they drop part of the video traffic already in the

source: For *Foreman*, 1442.4 and 1180.8 packets, in average, are dropped by the source in the application layer adaptation and in the combined case, respectively. For *Hall*, an average of 1465.9 packets are dropped in the source in the application layer adaptation case, and 1140.5 packets in the combined application plus MAC layer adaptation case.

All in all, based on the results, we can conclude that the combined MAC and application layer adaptation results in the smallest number of video packet drops in the BS MAC, thus being the most efficient approach in terms of network resource usage. Furthermore, the results show that the proposed scheme works similarly for different SVC sequences. The minor differences in the obtained results can be explained by the sequences' differing SVC layer characteristics (bitrate) and the nature of the content, which also affects the shape of the video bitstream.

4.3.2 Results for video quality

Considering the nature of the packet drops at BS MAC, the adaptation schemes using the MAC layer adaptation should clearly outperform the application layer adaptation in terms of the resulting video quality or QoE. We use PSNR as the main gauge for the video quality. The PSNR values averaged over the ten simulation runs are shown in Figures 9 and 10 for each of the

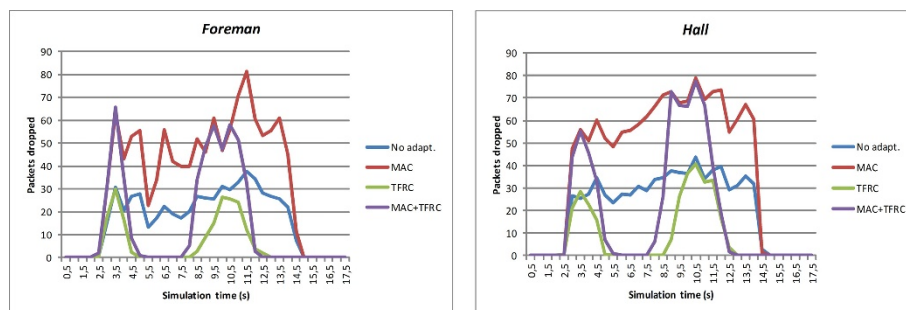


Figure 7 Simulation results for SVC 2nd EL transmission performance. Number of 2nd EL packets dropped in BS MAC for the two test sequences.

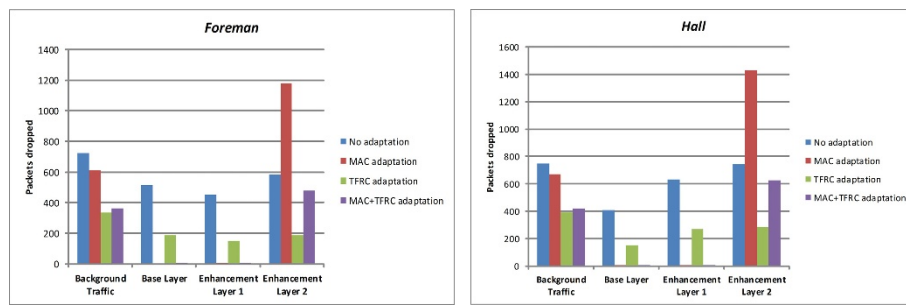


Figure 8 Simulation results: summary of the transmission performance results. Total number of packets dropped in BS MAC in the different cases for the two test sequences.

test cases and for the two video sequences, *Foreman* and *Hall*. The total PSNR averages for the test cases and sequences are also illustrated in Table 4. It should be noted here that the time axis of Figures 9 and 10 refer to the video and do not include the network delays. As a reference, we also show the PSNRs of the original videos, which have average PSNRs of 42.1 (*Foreman*) and 42.6dB (*Hall*). Moreover, Figure 11 clarifies the video quality for an individual *Foreman* frame at the point of 12.8 s.

From the results we can see that when no adaptation is used, the PSNR values decrease dramatically. The average PSNR of the whole *Foreman* sequence is 20 and 22.6dB for *Hall*, which give a very poor QoE. Naturally, the main reason for this originates from the great amount of BL packets being lost, although the frame copy algorithm in BL error concealment works slightly better for static sequences, such as *Hall*. Indeed, whole frames are

missing from the sequence since the EL packets cannot be used, if the entire corresponding BL frame is missing.

The application layer adaptation with the TFRC based algorithm provides at least satisfying results in terms of PSNR. The weakness of the TFRC based algorithm is its slow reaction to the decreasing channel bandwidth, which leads to packet losses before the actual application layer adaptation is triggered on. The average PSNR for the application layer adapted sequence is 29.7dB for *Foreman* and 31.6 dB for *Hall*, which results in a satisfactory level. The number of losses at the BL before the adaptation is triggered on leads to the situation where the PSNR falls temporarily to an unacceptable level. However, when the TFRC based algorithm begins to produce bandwidth estimate for the bitstream adaptation process, the PSNR rises again quickly to a good level. As a conclusion for the application layer adaptation, it works well under slowly changing channel conditions.

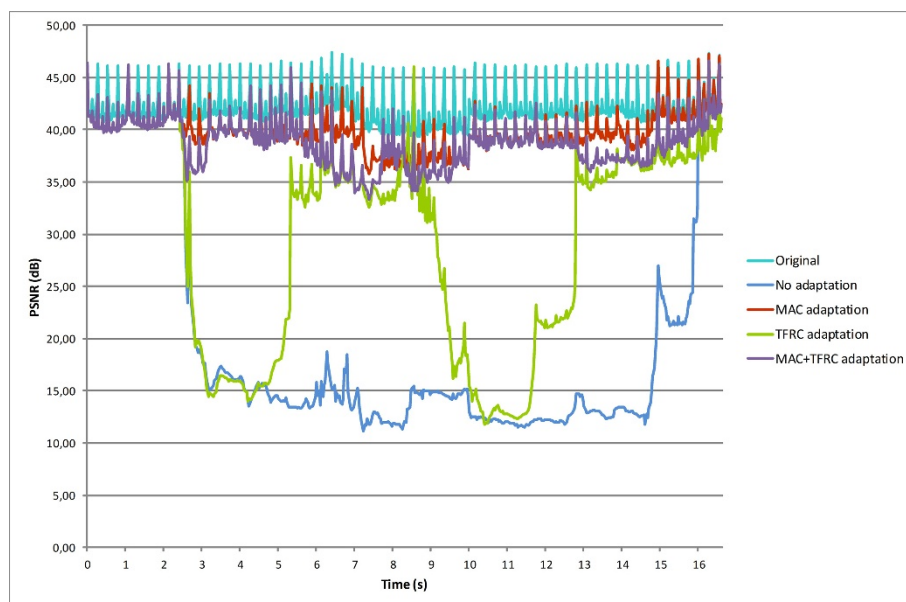


Figure 9 Simulation results for visual video quality. Received video quality measured in PSNR in the different cases for *Foreman*.

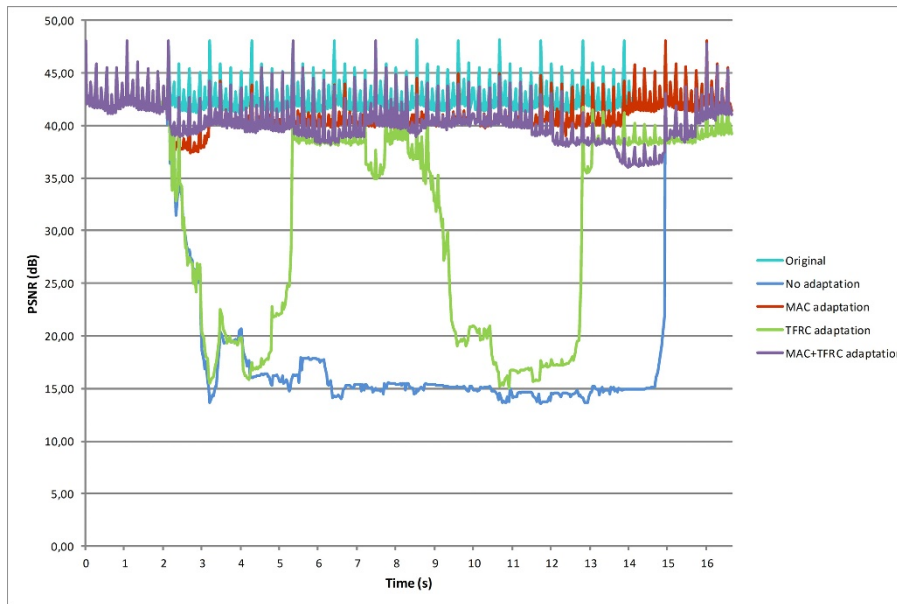


Figure 10 Simulation results for visual video quality. Received video quality measured in PSNR in the different cases for *Hall*.

Clearly, the best video quality can be attained when the MAC layer adaptation is enabled. Since very little or no BL packets are lost, the PSNR values stay near the original, that is, 39.9dB in average for *Foreman* and 41.2dB for *Hall*. The PSNR drops with the MAC layer adaptation follow well the drop rate of ELs. With MAC plus application layer adaptation, the effect of the application layer adaptation can be seen, for example, between 5-8 s in Figure 9. During this time period, a significant part of all EL packets are not sent at all, which negatively affects the PSNR. However, because only few BL packets are lost, the PSNR recovers fast and the error resilience algorithms of the decoder are less used. The average PSNR achieved with the combined MAC and application level adaptation is 38.8dB for *Foreman*, and 40.3dB for *Hall*.

Based on the obtained results, we can conclude that the combined MAC and application level adaptation approach proposed in this article can be used to achieve the best performance in terms of both optimal network resource usage and video QoE. As shown in the previous subsection, the proposed cross-layer optimization scheme reduces efficiently the number of useless packet transfers in the end-to-end path as most of the packets transmitted

also reach the receiver, but was also shown to produce a QoE near to that achieved with just MAC layer adaptation. Moreover, additional QoE performance gains could be expected when using a more efficient application adaptation solution.

5 Conclusions

To meet the requirement of ubiquitous media access for the future Internet services, we proposed in this article to use the OPTIMIX architecture for optimized media delivery and cross-layer signaling. We considered the different types of adaptations supported by the architecture for the optimization of scalable video transmission over wireless networks. We also presented the results obtained from a simulation study conducted using a complete OMNeT++ model of the OPTIMIX system architecture. In the study, we evaluated both the network resource efficiency and QoE performance of four adaptation cases: no adaptation, MAC layer adaptation, application layer adaptation, and combined MAC and application layer adaptation. Based on the obtained results, we propose to use the combined MAC and application layer adaptation approach for implementing wireless bandwidth adaptation support for the future ubiquitous video services. The proposed scheme performs the best when jointly considering the two criteria.

As future work, we plan to test and develop further our cross-layer SVC optimization strategy. The system is to be evaluated in different network scenarios and with a larger number of video clients to verify its performance and scalability. Also we aim to improve the awareness of

Table 4 Average PSNR comparison

	No adapt.	TFRC	MAC +TFRC	MAC	Original	
avg PSNR (dB)	20.0	29.7	38.8	39.9	42.1	<i>Foreman</i>
avg PSNR (dB)	22.6	31.6	40.3	41.2	42.6	<i>Hall</i>

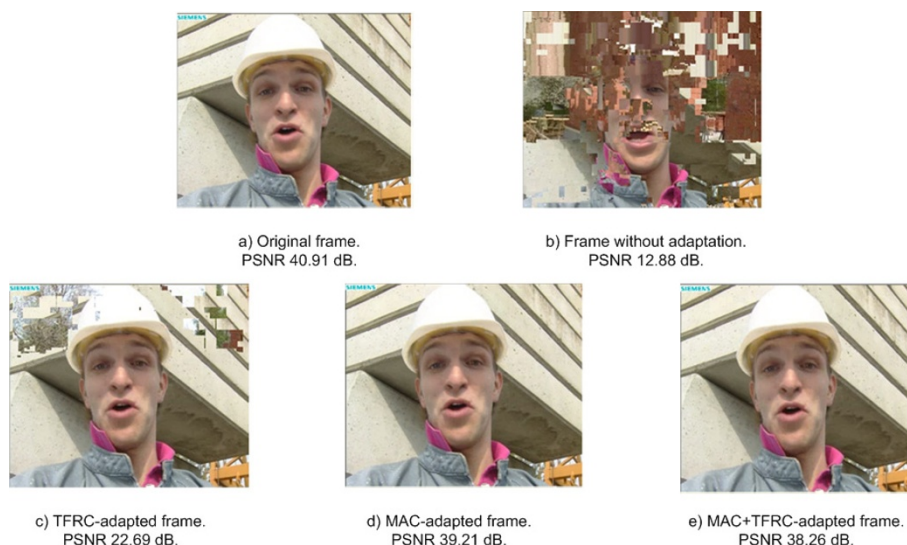


Figure 11 Visual video quality comparison for an individual frame. Visual frame quality comparison and PSNR values in the different cases.

the application level adaptation of the network state to increase its performance. For this, we envisage using link status information from the wireless MAC. Also the impact of host mobility is to be taken into account when developing the more advanced solution. For the MAC QoS architecture, we will investigate its integration also to other wireless systems. Parts of the proposed SVC streaming system are also to be implemented and included into a real testbed environment.

Acknowledgements

This work was carried out in ICT-OPTIMIX and ICT-CONCERTO projects, which are partially funded by the European Commission. The authors would like to thank their colleagues who have participated in the two projects and especially those who have contributed to the development and implementation of the simulation framework. For further information about the projects, please visit <http://www.ict-optimix.eu> and <http://www.ict-concerto.eu>.

Competing interests

The authors declare that they have no competing interests.

Received: 16 June 2011 Accepted: 23 January 2012

Published: 23 January 2012

References

- Cisco Visual Networking Index: Forecast and Methodology (2010-2015), H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Technol.* **17**, 1103-1120 (2007)
- K Tappayuthpijarn, G Liebl, T Stockhammer, E Steinbach, Adaptive video streaming over a mobile network with TCP-friendly rate control, in *Proceedings of IWCMC'09*, Leipzig, Germany, 1325-1329 (June 2009)
- I Kofler, J Seidl, C Timmerer, H Hellwagner, I Djama, T Ahmed, Using MPEG-21 for cross-layer multimedia content adaptation. *Springer J. Signal Image Video Process.* **2**, 355-370 (2008)
- T Pliakas, G Korkmentzas, S Tsekeridou, Joint scalable video coding and packet prioritisation for video streaming over IP/802.11e heterogeneous networks, in *Proceedings of the 3rd International ICST Mobile Multimedia Communications Conference*, Nafpaktos, Greece, 31:1-31:131:6 (2007)
- G Liebl, T Schierl, T Wiegand, T Stockhammer, Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG-4-AVC, in *Proceedings of ICME'06*, Toronto, Canada, 625-628 (July 2006)
- HH Juan, HC Huang, C Huang, T Chiang, Cross-layer mobile WiMAX MAC designs for the H.264/AVC scalable video coding. *Springer Wirel. Netw.* **16**, 113-123 (2010)
- G Bianchi, A Detti, P Loreti, C Pisa, F Proto, W Kellerer, S Thakolsri, J Widmer, Application-aware H.264 scalable video coding delivery over wireless LAN: experimental assessment, in *Proceedings of IWCLD'09*, Palma de Mallorca, Spain, 1-6 (11-12 June 2009)
- Y Fallah, P Nasiopoulos, H Alnuweiri, Scheduled and contention access transmission of partitioned H.264 video over WLANs, in *Proceedings of IEEE Global Communications Conference*, Washington DC, USA, 2134-2139 (Nov 2007)
- C Lamy-Bergot, R Fracchia, M Mazzotti, S Moretti, E Piri, T Sutinen, J Zhuo, J Vehkaperä, G Feher, G Jeney, G Panza, P Amon, Optimisation of multimedia over wireless IP links via X-layer design: an end-to-end transmission chain simulator. *Multimedia Tools and Applications*. Springer Netherlands. **55**(2), 261-288
- IEEE, IEEE Standard for Local and Metropolitan Area Networks-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (2007)
- T Sutinen, J Huusko, MAC layer QoS architecture for optimized and fair transmission of scalable video, in *Proceedings of IEEE Consumer Communications and Networking Conference*, Las Vegas, Nevada, USA, 421-425 (9-12 Jan 2011)
- OMNeT++ Network Simulation Framework, <http://www.omnetpp.org>
- I Haratcherev, J Taal, K Langendoen, R Lagendijk, H Sips, Optimized video streaming over 802.11 by cross-layer signaling. *IEEE Commun Mag.* **44**, 115-121 (2006)
- CW Huang, M Loiacono, J Rosca, JN Hwang, Airtime fair distributed cross-layer congestion control for real-time video over WLAN. *IEEE Trans Circuits Syst Video Technol.* **19**(2), 1158-1168 (2009)
- JR Ohm, Advances in scalable video coding. *Proc IEEE.* **93**, 42-65 (2005)
- M Uitto, J Vehkaperä, Spatial enhancement layer utilisation for SVC in base layer error concealment, in *Proceeding of 5th International Mobile Multimedia Communications Conference*, London, UK, 1-8 (27-9 Sept 2009)
- J Mäkelä, K Pentikousis, Trigger management mechanisms, in *Proceeding of ISWPC'07*, San Juan, Puerto Rico, 378-383 (Feb 2007)
- IEEE, IEEE Standard for Local and Metropolitan Area Networks-Part 21: Media Independent Handover (2009)
- E Piri, T Sutinen, J Vehkaperä, Cross-layer architecture for adaptive realtime multimedia in heterogeneous network environment, in *Proceeding of European Wireless Conference*, Aalborg, Denmark, 293-297 (17-20 Mat 2009)

21. T Melia, A de la Oliva, I Soto, C Bernardos, A Vidal, analysis of the effect of mobile terminal speed on WLAN/3G vertical handovers, in *Proceeding of IEEE Global Communications Conference*, San Francisco, USA, 1–6 (Nov 2006)
22. J Heinanen, F Baker, W Weiss, J Wroclawski, Assured forwarding PHB group. IETF Request for Comments (RFC 2597). <http://www.ietf.org/rfc/rfc2597.txt>
23. E Piri, M Uitto, J Vehkaperä, T Sutinen, Dynamic cross-layer adaptation of scalable video in wireless networking, in *Proceedings of IEEE Global Communications Conference*, Miami, Florida, USA, 1–5 (6–10 Dec 2010)

doi:10.1186/1687-1499-2012-25

Cite this article as: Sutinen et al.: Towards ubiquitous video services through scalable video coding and cross-layer optimization. *EURASIP Journal on Wireless Communications and Networking* 2012 **2012**:25.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
