Dynamic Cross-layer Adaptation of Scalable Video in Wireless Networking

Esa Piri, Mikko Uitto, Janne Vehkaperä, and Tiia Sutinen
VTT Technical Research Centre of Finland
Kaitoväylä 1, FI-90571, Oulu, FINLAND
Email: firstname.lastname@vtt.fi

***Abstract*—The recent trend in Internet traffic indicates the proliferation of usage of multimedia services where a substantial part is related to some sort of video transmission. Moreover, an increasing number of Internet users employ wireless access technologies. High-quality video streaming over wireless access in unison with great mobility brings challenges to sustain the mobile user perceived video quality high. Capacities of wireless links vary due to, for instance, coverage area limitations, multipath propagation, and fading. However, novel video codecs utilize a layered encoding/decoding mechanism, which conveniently allows adapting the video quality, and thus the bitrate, by adjusting the number of layers transmitted. In this study, we exploit an extensive cross-layer signaling framework for a dynamic scalable video adaptation in varying network capacity. We focus on comparing a fast and fair MAC-layer packet scheduling with a relatively slow and long-term adaptivity taken place already at the application layer using a real H.264/SVC video. Our results attest the advantages of adaptation through the use of feedback signaling, which enables continuing the use of the current network access despite its capacity variation.**

I. Introduction

Internet traffic is rapidly increasing today and a growing proportion of the traffic is propagated over broadband wireless access (BWA) technologies such as Wireless LAN, Wireless MAN, and UMTS. Multimedia is expected to increasingly dominate the Internet traffic grow in future. According to Cisco [1], Internet video will account for 60% of all consumer Internet traffic by 2013. All above are a corollary of the trend where mobile phones are more and more transforming to multimedia devices with multiaccess capabilities and powerful processors. Thus, the usage of multimedia aside of mobility over various BWA technologies, providing different capabilities and characteristics, has proliferated. However, as the interest grows, the demand for bandwidth increases alike. Broadly, operators lag behind to rise to the bandwidth demand by offering decent resources to the prolific number of mobile users. In addition, fading and multipath phenomena typical of wireless networking fluctuate the available capacity.

In order to make the most of wireless networking, constant monitoring of current network conditions and available handover targets is needed. Nevertheless, not always a handover to another access point or network is feasible, although the current access is impairing. An alternative solution to continue using the current access is dynamic traffic adaptation and prioritization according to the varying network capacity. UDP, which is the dominant protocol for real-time multimedia transport, does not provide any congestion control mechanism to keep track of the success rate of a transmission flow. The flow monitoring and adaptation needs to be carried out elsewhere. Novel VoIP and video codecs come up with possibilities to overcome this issue. They are based on a layered coding mechanism where only base layers are needed for a successful decoding. Enhancement layers do increase the quality only.

In this work, we study the benefits of adapting video transmission using scalable video coding (SVC) [2] according to the wireless network capacity in real-time. Practically, there are two ways to dynamically adapt video transmission, at the medium access control (MAC) of the serving base station (BS) and in the video source. We evaluate the gains of these methods to the end user's experienced video quality in a congestion and a varying signal strength scenarios using a real H.264/SVC video stream. In both methods, the adaptation is realized through the prioritization of video packets upon insufficient channel capacity for the full-quality video.

An extensive cross-layer signaling architecture plays a notable role in the network traffic adaptation. To make intelligent and preventative decisions to keep the end-user experienced video quality high, traffic adaptors need timely information about the current state of the employed network and traffic flow to adapt. We use a signaling architecture introduced in [3], which uses IEEE 802.21 [4] as the signaling basis and expands it with Triggering Framework [5]. Triggering Framework is used in the signaling that is out of scope of IEEE 802.21.

The rest of the study is organized as follows. Section II presents and relates this work to other scalable video transmission studies. The signaling framework, adaptation methods and simulation environment are detailed in Section III. Section IV presents the results of our measurements. Finally, the paper is concluded in Section V.

II. Related Work

Capitalizing on the flexibility of SVC streams in improving video delivery performance with MAC-level prioritization and scheduling has been studied vastly during the recent years. Most of the existing solutions, such as the ones proposed in [6]–[9], are technology dependent, that is, they introduce video- or SVC-specific extensions to the standard MACs. Each solution defines a way for implementing differentiated treatment for video packets of differing priorities utilizing the QoS mechanisms supported by the access technology or the proposed extensions to those. The need for intra-traffic category differentiated services in the case of video traffic has been recognized also by standardization bodies. For example, the IEEE 802.11 working group has started the specification of MAC enhancements for robust audio video streaming under the task group 802.11aa [10] and these enhancements also include support for intra-access category service differentiation.

We focus on supporting video users with heterogeneous network connectivity. Thus, we propose using an access technology independent solution for MAC-level adaptation instead of using any of the currently available non-standard solutions for scalability. The same approach is taken also in [11], where a more general architecture for improving scalable video delivery is defined. Their model introduces Media Adaptation Layer (MAL) on top of the data link layer to prioritize and schedule SVC data packets according to the video packet type and priority information as well as the channel state. Although the model provides a technology-independent solution for the MAC-level SVC adaptation, its validation is rather limited.

The MAC-level adaptation solution proposed in this paper resembles the MAL approach. However, we present a more detailed design of the queue management solution and the scheduling algorithm as well as enhance the technology independency through the use of standard signaling. In our solution, the number of queues is limited and we consider also traffic other than video. Also, a more thorough performance evaluation is given taking into account the effect of different network conditions to the MAC-level adaptation performance.

As the scope of MAC-level adaptation is restricted to the next hop link, an adaptation dealing with congestion in the other parts of the end-to-end video path is also wanted. The most logical layer to perform the adaptation of multimedia is the application layer as this is the layer where the data rate of multimedia stream is originally set. Rate and congestion control are the most popular techniques to control the data rate and a lot of research has been conducted, especially, utilizing TCP-friendly rate control (TFRC) algorithms [12], [13]. TCP is commonly considered as an unsuitable protocol for real-time multimedia but it has some attractive features, such as fairly efficient and adaptive utilization of available network resources. It also provides fairness in case of several TCP streams sharing a network link. These are desirable features for rate control algorithms to be utilized in large-scale networks where network resources are shared with other traffic. Since TFRC-based algorithms provide many of the desired features, our application layer adaptation algorithm is based on TFRC.

III. Methodology

We have developed an access technology independent scheduling algorithm for scalable multimedia to ensure the transmission of the most important packets and to aim at losing only packets that has the least effect on overall service quality. Source-level adaptation is used as a comparison to the MAC-level scheduling. Overall, source-level adaptation provides a generic way to combat network traffic bottlenecks no matter where they reside.

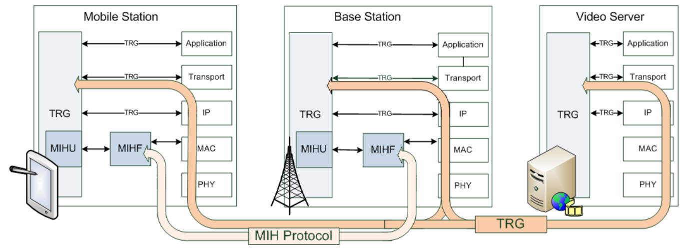

Fig. 1. Signaling framework

A. Signaling Framework

Fig. 1 illustrates the signaling framework exploited in the video adaptation. It is detailed in [3]. As shown in the figure, IEEE 802.21 [4], Media Independent Handover (MIH) Services, is used to facilitate Layer-2 (L2) operations in the BS and the mobile station (MS), while the end-to-end and cross-layer signaling between entities located on layers above L2 is realized through the use of Triggering Framework [5].

In the presented simulation model, BS and MS exchange events and information over L2 communication using an MIH protocol. In specific, MIH Functions (MIHFs) on MS and BS register with each other in order to use remote MIH services, that means, BS subscribes to MIH_Link_Parameters_Report event to receive current link parameters from the MS side.

Triggering Framework provides the required signaling for the TFRC-based source adaptation. In our system, each network entity runs event services of Triggering Engine (TRG). The event sources and consumers communicate only with their local TRG and the remote event delivery between TRG entities is enabled using cascaded TRG communication over IP. In the presented measurements, we pass only a video packet loss parameter from the TRG running in MS to the one in the server, but the framework leaves room for future extensions. Although the remote event delivery is originally specified to be performed over TCP, in our simulations UDP is employed.

B. MAC-level Adaptation

Video adaptation at MAC is carried out through the packet prioritization and scheduling, which is triggered upon a link capacity starts deteriorating. Both BS and MS constantly monitor the conditions of the employed link. BS monitors the link state through the rate of dropped packets, which are dropped due to exceeding of retransmission limit or transmission (TX) queue overflow. MS keeps track of the number of erroneously received packets, packet error rate (PER), and signal strength. In consequence of Automatic Repeat-reQuest (ARQ) procedure, PER does not imply the rate of lost packets.

Link condition is monitored by factoring in link parameter values from five previous link condition checks, performed every 0.1 s. Link condition and its trend is monitored using

$$\sum_{k=0}^{4} E_k \times M_k, M_k \in \{0.45, 0.3, 0.1, 0.1, 0.05\} \quad (1)$$

where E denotes failed transmissions (BS), queue size (BS), erroneously received packets (MS), or signal strength in percents (MS). M is a coefficient to accentuate the most recent values. As indicated, for example, in [14], the use of weighted coefficients for previous values increases the accuracy and reliability of predictive events.

MS periodically (every 0.5s) sends an MIH_Parameters_Report event to BS in order to refresh its view of the link condition. Parameters included in the events are nominal capacity, current bitrate, PER, and signal strength of wireless link. When MS notices that its weighted PER has exceeded the threshold of 4% or signal strength dropped below 70% of maximum, it generates an additional MIH_Parameters_Report event to also timely notify BS about this. Upon receving a parameters event from MS indicating the link condition change to worse, MIH User (MIHU) on BS triggers MAC to start prioritizing received IP packets by using MIH_Link_Actions command, unless prioritization is activated already. BS uses 1% as a threshold for the dropped frames and a weighted queue size of *max_queue_size - 2* to trigger prioritization on based on its own link monitoring. Once the link condition recovers, weighted rate of failed transmissions drops below 0.5%, BS returns to the state without packet prioritization and continue using one first-in first-out (FIFO) queue. Thresholds are chosen not to trigger the prioritization on too sensitively.

When prioritization is enabled, MAC classifies the received IP packets into three categories; high priority, medium priority, and low priority. IP packets including video packets with the highest priority (base layer frames) are put to the high priority category. Packets carrying first quality enhancement layer frames and all other than video packets are stamped as medium importance. Packets carrying second quality enhancement layer frames are inserted to the low priority queue. Upon full TX queue, IP packet received from the upper layer drops the oldest lower importance packet from the queue. If no packet to be dropped exist, the received packet is dropped.

The MAC scheduling algorithm factors in tolerated queuing delays for each video packet. This means, if the packet scheduled for sending next is a video packet and it still can wait for transmission, the front packet from the lower priority queue is sent first (if exist and older). Thus, the tolerated queuing delays do not refer to the actual time a packet queues at maximum. If the packet chosen to be sent next includes video data, video packets in the lower priority queues exceeded their tolerated queuing delays are dropped. We opted to use 4 milliseconds (ms), 6 ms, and 8 ms as tolerated delays for the base layer and enhancement layers, respectively.

The fairness with non-video traffic is realized in a way that non-video packet is merely dropped if TX queue is full when received and no packet exist in the low priority queue to be dropped. Moreover, the non-video packets are scheduled using the FIFO scheduling scheme on the turn of medium priority queue without the definitions of tolerated queuing delays.

C. Source-level Adaptation

In the application layer, the TFRC-based adaptation algorithm is used to control the data rate of the multimedia source. The source data rate *T* is calculated using TFRC equation [12]

$$T = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RT0}(3\sqrt{\frac{3p}{8}})p(1+32p^2)} \tag{2}$$

where t_{RT0} is $4R$ [13]. Packet loss rate *p* is calculated in MS every 0.5 s. This event with a filter excluding zero values is subscribed to by the video server. The decoder in MS continuously monitors the number of missing video packets and calculates the current loss rate by using Eq.(1).

We used a fixed packet size *s* (1450 bytes) and an estimate for the round-trip time *R*, factoring in the current video packet loss. Furthermore, the bitrate calculated by TFRC is utilized in the bitstream adaptation process every 0.6 s. The bitstream adaptation extracts the Network Abstraction Layer (NAL) units from the stream and finds spatial, temporal, and quality id values for each packet along with the bitrates for each layer. The bitstream adaptation process simply compares the target data rate calculated by TFRC to the layer bitrate and drops enhancement layers if the target data rate is lower.

D. Simulation Environment

We measure the video adaptation by exploiting a simulation environment developed in ICT-OPTIMIX project to OMNeT++ (see www.omnetpp.org) simulation framework. The environment consists of a video server, an IEEE 802.11g [15] BS, an MS connected to the BS, and an IPv6 wired network connecting the BS and the server. In our study, the wired network does not introduce any packet loss or congestion. The video server and the MS use RTP/UDP/IPv6 protocol stack for the H.264/SVC encoded video stream.

Table I lists the most important parameters used in our simulations. PHY simulates a log-normal shadowed uncorrelated Rayleigh fading channel (20 MHz) with additive white Gaussian noise (AWGN) and without path-loss. PHY does not use adaptive coding and modulation, but the modulation is kept fixed. Request to Send/Clear to Send (RTS/CTS) is not used.

The input video used in our simulations is a 500 frames (17 s) long H.264/SVC encoded sequence. It is encoded with JSVM 9.15 reference encoder and the output is decoded using the same software decoder. Frame copy as an error concealment technique is implemented into the decoder in order to cope with the packet losses in the base layers. The SVC video contains a base layer with two quality enhancement layers. The bitrate with the best quality and highest frame rate (30 Hz) is approximately 2 Mb/s. The resolution, Common Intermediate Format (CIF), is kept constant, which means no spatial scalability is included in this test sequence.

IV. Results

We conducted the video adaptation measurements in two scenarios: A) BS gets congested; and B) the link between BS and MS strongly varies. In both scenarios, we compare the performance of MAC- and source-level adaptations to the case without adaptation. For a fair comparison, physical medium is kept similar for each case by giving the same seed for the uncorrelated shadowing and Rayleigh channel. Each simulation run lasted 17 s.

TABLE I
SIMULATION PARAMETERS

MAC Parameters	
Retransmission limit	7
TX queue	3 queues of size 15 frames in total
PHY Parameters	
Modulation	QPSK (FEC: 1/2)
Coherence time	0.1 s
Log-normal shadowing time	3 s
Shadowing standard deviation (σ)	4 - 5 dB
E_s/N_0 (1 m distance from BS)	60 dB
Video Parameters	
Video encoding	H.264/SVC, base + 2 quality layers
GOP length and structure	8 and IPPPP ...

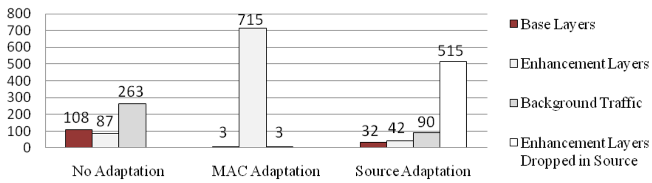

Fig. 2. Comparison of lost packets due to congestion

A. Congested network

In this scenario, the signal strength stays good during the simulation runs ($\sigma = 4$ dB) and it does not cause packet drops. The wireless link is congested by injecting UDP/IPv6 background traffic into the link with the maximum transfer unit (MTU) of 600 bytes. The link capacity with the variable MTU video stream and background traffic is 5.2 Mb/s with small loss (<1%). The background traffic starts at the point of 2.0 s with the throughput of 3 Mb/s and progressively increases every 2 s by 150 kb/s until the throughput of 3.6 Mb/s is reached. After that, the throughput decreases by 500 kb/s per 2 s until it ends at the point of 16 s.

Fig 2 shows the lost video and background traffic packets. With MAC adaptation, 99% of the dropped packets are video enhancement layers and only three base layer packets are dropped at MAC before the adaptation was triggered on. Without adaptation and with source adaptation, the dropped packets at MAC are quite evenly divided between the background and video traffic. However, with source adaptation large number of enhancement layers are being dropped already in the source, thus, resulting in 78% lower packet drop rate at MAC than with MAC adaptation. The large packet drop rate with MAC adaptation can also be explained by the scheduling algorithm's strict tolerated queuing delay restrictions. Therefore, the average throughput attained with MAC adaptation is 4.53 Mb/s, which is 86 kb/s lower than that without adaptation.

On average, the adaptation methods lower the packet transfer delays of video packets over the wireless link by 1 ms, resulting in 5 ms delays. With MAC adaptation, the delay variation between the base and enhancement layers is in the magnitude of 1 ms due to the various tolerated queuing delays.

Fig 3 shows the Peak Signal-to-Noise Ratio (PSNR) results

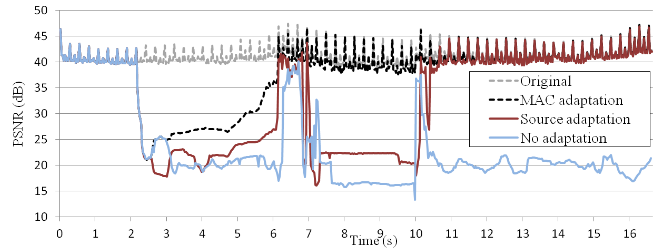

Fig. 3. Peak Signal-to-Noise Ratios

from the whole 500-frame sequence. For comparison, the original flawless video is also shown as a reference. As can be seen, the missing base layer slices negatively affect the quality by lowering the PSNR by several decibels. The used error-resilient SVC decoder conceals the missing base layer pictures using slice copy from the reference frame list. Additionally, the missing enhancement layer pictures are concealed by using the correctly decoded lower layer pictures as a reference. We noted that MAC adaptation gives the best results by achieving an average PSNR of 38 dB, being 14 dB higher than in the video without adaptation and only 3.5 dB lower than in the original. Source adaptation produces an average PSNR of 32 dB, which is a good result. The sharp cuts in the source adaptation PSNR decrease the mean value significantly. Around 7 s and 10 s, 5 and 9 base layer packets were lost, respectively. We noted that TFRC increases the bitrate too rapidly after a negligible frame loss rate is attained, although the current bitrate level would be ideal and we delayed the rise by taking a weighted value from the previous five bitrates, as proposed by [12].

We found that 0.5 s interval for receiving the video packet loss rate event from the decoder is sufficient to maximize the benefit of source adaptivity. With a longer interval, we observed problems when the adaptation algorithm started increasing the bitrate by supposing the available throughput larger than it actually was. Then, recovering from the bandwidth misestimation takes a longer time and more base layers are lost. With one client, the throughput of video packet loss feedback messaging with the 0.5 s interval is only 8.7 kb/s.

B. Varying link condition

Signal condition changes in wireless communication can be rapid, dramatic, and short-term, which is the case in this scenario. In contrast to Section IV-A, all frame drops are incurred by heavily deteriorating signal strength ($\sigma = 5$ dB). Between 3-6 s of the simulation runs, the log-normal shadowing gain jumps to 2.3 dB, by being elsewhere on range 0.7-1.2 dB. This causes an uncorrectable amount of bit errors to a large number of MAC frames received by MS. Although the mean bit error rate after correction coding during the 3 s period is only 0.5%, the mean frame drop rate at the receiver ranges between 25% - 32%, by being the largest one when source adaptation is in use (less packets to be transmitted).

Fig 4 illustrates the lost base layer packets at BS MAC. Without adaptation, the weak signal drops 77 base layer packets in total. Source adaptation decreases the number by

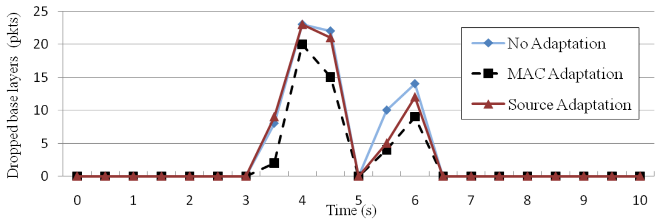

Fig. 4. Base layer packets dropped by BS MAC

7 packets and MAC adaptation mitigates the impact of the bad link even more by dropping 36% less base layer packets. Source adaptation follows the non-adaptation curve until the adaptation starts dropping enhancement layers, at the point of 3.9 s. From the point of 4.3 s on, all enhancement layers are being dropped in the source until 6.7 s. In source adaptation, most of the transmission effort is trained on the base layer packets as enhancements packets are absent. Thus, the gain of source-level adaptation is based on increasing the proportional success rate of the most important packets, that are base layers.

The additional gain of MAC adaptation compared to source adaptation is in great measure a result of the scheduling algorithm's feature which increases the retransmission limit by three when weak signal strength is faced. This small increase does not introduce any excessive queuing delays. Median L2 transfer delays were in the order of 1 ms during the weak signal, but high signal variation effected 6 ms and 8 ms average delays with the MAC and source adaptations, respectively.

Although one video base layer is sliced to multiple IP packets, the whole layer is not useless even though some of the slices are lost. The number of correctly received base layers at the decoder increased by 18 with MAC adaptation and by 4 with source adaptation. On average, one second in the employed video contains 88 base layer packets. Although the gain with respect to the received base layers is not substantial, an advanced video decoder can conceal frame drops the better the more it receives base layers. Thus, already a small increase in the number of base layers reduces the amount of disturbing errors in the video. Also, the enhancement layers can be utilized, even if part of the base layer frame they relate to is lost. With MAC adaptation, more than 500 enhancement layers were received during the weak signal strength period.

Overall, during very low signal strength the adaptation gains are observed minor. Although MAC adaptation increases the retransmission limit to ten upon weak signal, it is still not sufficient to reap significant benefit. Increasing the MAC retransmission limit is, practically, the most efficient way to bring more advantages here, which, on the other hand, may lead to excessive packet transfer delays.

V. Conclusion

In this study, we evaluated the benefits of adapting scalable video transmission at MAC and source level when wireless link gets congested or its signal strength impairs. Our results clearly show that both adaptation methods can better sustain the video quality at moderate level, although link conditions dramatically change. However, we found that MAC-level adaptation, as a fast way to conform the video stream to the current link conditions, clearly outperforms source adaptation in congestion. With a very weak signal strength, the observed differences were smaller. With MAC adaptation, only three base layer packets were lost when link got congested. Anyway, as the PSNR results indicated, source-level adaptation substantially increases the transport reliability of video base layers required in error-free decoding than the case without any adaptation. In addition, source adaptation handles cases where network problems reside somewhere else than in the link between MS and BS. There, MAC adaptation is useless.

Acknowledgment

This work is carried out in ICT-OPTIMIX, which is partially funded by European Commission.

References

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2008-2013," White Paper, 2009.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. CSVT*, vol. 17, pp. 1103–1120, September 2007.

[3] E. Piri, T. Sutinen, and J. Vehkaperä, "Cross-layer Architecture for Adaptive Real-time Multimedia in Heterogeneous Network Environment," in *Proc. European Wireless*, Aalborg, Denmark, May 2009, pp. 293–297.

[4] IEEE 802.21 WG, *IEEE Standard for Local and Metropolitan Area Networks. Part 21: Media Independent Handover Services*. IEEE Std 802.21-2008, January 2009.

[5] J. Mäkelä and K. Pentikousis, "Trigger Management Mechanisms," in *Proc. ISWPC*, San Juan, Puerto Rico, February 2007, pp. 378–383.

[6] T. Pliakas, G. Kormentzas, and S. Tsekeridou, "Joint Scalable Video Coding and Packet Prioritization for Video Streaming over IP/802.11e Heterogeneous Networks," in *Proc. Mobimedia*, Nafpaktos, Greece, August 2007, pp. 1–6.

[7] Y. Fallah, P. Nasiopoulos, and H. Alnuweiri, "Scheduled and Contention Access Transmission of Partitioned H.264 Video over WLANs," in *Proc. GLOBECOM*. IEEE, 2007, pp. 2134–2139.

[8] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer, "Advanced Wireless Multiuser Video Streaming Using the Scalable Video Coding Extensions of H.264/MPEG-4-AVC," in *Proc. ICME*, Toronto, Canada, July 2006, pp. 625–628.

[9] H.-H. Juan, H.-C. Huang, C. Huang, and T. Chiang, "Cross-layer Mobile WiMAX MAC Designs for the H.264/AVC Scalable Video Coding," *Springer Wireless Networks*, vol. 16, pp. 113–123, January 2010.

[10] IEEE 802.11 WG, *Draft standard for Local and Metropolitan Area Networks. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 8: MAC Enhancements for Robust Audio Video Streaming*, January 2010.

[11] J. Huusko, J. Vehkaperä, P. Amon, C. Lamy-Bergot, G. Panza, J. Peltola, and M. Martini, "Cross-layer Architecture for Scalable Video Transmission in Wireless Network," *Signal Processing: Image Communication*, vol. 22, pp. 317–330, March 2007.

[12] K. Tappayuthpijarn, G. Liebl, T. Stockhammer, and E. Steinbach, "Adaptive Video Streaming over a Mobile Network with TCP-Friendly Rate Control," in *Proc. IWCMC*, Leipzig, Germany, June 2009, pp. 1325–1329.

[13] I. Kofler, J. Seidl, C. Timmerer, H. Hellwagner, I. Djama, and T. Ahmed, "Using MPEG-21 for cross-layer multimedia content adaptation," *Signal, Image and Video Processing*, vol. 2, pp. 355–370, October 2008.

[14] T. Melia, A. de la Olivia, I. Soto, C. J. Bernardos, and A. Vidal, "Analysis of the Effect of Mobile Terminal Speed on WLAN/3G Vertical Handovers," in *Proc. GLOBECOM*, San Francisco, CA, USA, November 2006, pp. 1–6.

[15] IEEE 802.11 WG, *IEEE Standards for Local and Metropolitan Area Networks. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE Std 802.11-2007, January 2007.