Traffic Prioritization of H.264/SVC Video over 802.11e Ad Hoc Wireless Networks

Attilio Fiandrotti, Dario Gallucci, Enrico Masala and Enrico Magli[1]
Dipartimento di Automatica e Informatica / [1]Dipartimento di Elettronica
Politecnico di Torino, Italy
Email: [dario.gallucci, attilio.fiandrotti, enrico.masala, enrico.magli]@polito.it

***Abstract*—The H.264/SVC video codec extends the H.264/AVC standard with scalability features. In this paper we introduce a traffic prioritization algorithm suitable for the transmission of both H.264/SVC and H.264/AVC video over 802.11e ad hoc wireless networks. The proposed algorithm exploits the traffic prioritization capabilities offered by 802.11e to provide better protection to the most perceptually important parts of a video while achieving efficient network resource usage. We evaluate the algorithm by simulating video transmissions in an ad hoc network scenario. Results show that the H.264/SVC codec particularly benefits from the proposed algorithm, which enables a graceful video quality degradation in congested network conditions, as well as PSNR gains up to 2 dB with respect to the H.264/AVC codec using the same amount of network resources.**

I. Introduction

In recent years, the availability of multimedia devices with widely different processing capabilities has led to an increase in the demand for efficient scalable multimedia coding systems. In the case of video, the recently developed H.264/SVC amendment [1] aims at enhancing the H.264/AVC standard [2] with different types of scalability options. Such options are also useful to adapt the video to heterogeneous networks scenarios composed of, e.g., by high-speed DSL lines, WiMax, WiFi and 3G networks, which are becoming more and more common in digital communications.

Wireless multimedia communications, and video in particular, are however challenging. The intrinsic unreliability of wireless channels demands in fact robust error protection mechanisms. Some of these mechanisms are based, for instance, on unequal error protection schemes. Many schemes exploit the unequal perceptual importance of the data stream by protecting each part of it in proportion to its importance. Protection is provided either by means of application-level error control mechanisms or by exploiting the lower-level QoS capabilities offered by some network technologies.

The recently standardized 802.11e protocol [3] is especially suitable for the transmission of multimedia data because it offers different levels of traffic prioritization by means of different service classes. For instance, delay sensitive VoIP and video traffic can be advantaged over background traffic by means of the high priority service offered by 802.11e. However, while some multimedia flows generally present moderate bandwidth requirements (e.g., VoIP), video demands a considerable amount of bandwidth. Therefore an efficient video prioritization scheme is needed to optimally exploit the QoS capabilities offered by 802.11e.

Prior works focused on schemes which optimize the transmission of video over 802.11e networks. Transmission of H.264/AVC videos over 802.11e was investigated in [4], which shows that better video quality as well as more efficient network resources usage can be achieved if the video flow is distributed among multiple traffic classes instead over a single high priority class. The work in [5] addresses the transmission of 3D wavelets based scalable video in 802.11e networks. An architecture for SVC video transmission over various types of wireless networks, including 802.11e, was proposed in [6], but it is however complex as heavily relies on fuzzy logic.

In this work a new video prioritization algorithm for video communication over 802.11e networks is presented. While the algorithm was originally designed for SVC video transmission, it is also suitable for standard AVC videos. The algorithm exploits the information found in the SVC bitstream to order packets in decreasing order of importance, considering both the layer to which data belong and the dependencies in the decoding process. This scheme is then extended to the case of standard AVC videos by considering the information available in an AVC bitstream. Differently from [4] which addresses data-partitioned AVC streams only, the proposed algorithm is suitable also for unpartitioned data streams as well as for streams which encompass bipredictively encoded pictures. Another contribution of this work is a comparison between the video quality that the two encoding formats can provide by using the same prioritization algorithm and the same amount of network resources. The comparison quantifies the gains offered by the SVC scalable coding option of H.264 over the standard non-scalable AVC standard.

II. Background

A. The 802.11e Communication Standard

The recently standardized IEEE 802.11e amendment [3] supersedes the legacy 802.11 specifications for wireless channel access by introducing the Enhanced Distributed Channel Access (EDCA) mechanism. For each host station, four distinct transmission queues known as Access Categories (ACs) are introduced in place of the unique queue offered by the legacy 802.11 standard. Each queue is characterized by specific contention window and interframe spacing values, which are the parameters controlling the amount of time the

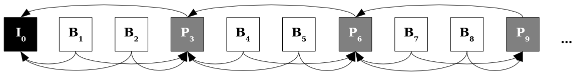

Fig. 1. GOP structure of the H.264/AVC encoding scheme used in this work.

transceiver has to wait before attempting access to the channel to transmit a packet. Actual settings of such parameters are different for each AC, resulting in high priority and low priority queues characterized by unfair chances of getting access to the channel. As a result, packets in high priority queues are elected for transmission before packets in low priority queues, resulting in an effective intra and inter hosts traffic prioritization mechanism.

B. The H.264 Scalable Video Coding Standard

The H.264 standard decomposes the functionalities of the encoder between a Video Coding Layer and a Network Abstraction Layer (NAL) [7]. The former layer encompasses all the encoder core functionalities such as macroblocks and slice encoding. The latter layer facilitates the transport of video over packet networks by encapsulating of each piece of encoded data into an independently decodable transport unit known as NAL Unit (NALU). Each NALU is prefixed by a header whose fields describe the characteristics of the payload. In particular, the *Type* field specifies the type of data the NALU contains (video slices, parameter sets etc.) and the *NRI* field indicates the number of pictures predicted from that NALU.

As in other video compression standards, pictures are encoded constraining the dependencies within the so called *Group of Pictures* (GOP). Figure 1 depicts a typical AVC GOP: each box represents a NALU, the letter inside the box indicates the type of the corresponding picture (Intra, Predictively or Bipredictively coded), the subscript number the picture display order and the arrows show the decoding dependencies among pictures.

The recently proposed H.264/SVC amendment extends H.264/AVC with temporal, spatial and SNR scalability options, allowing to encode a video as an independently decodable base layer and one or more enhancement layers. The base layer is encoded in AVC format to preserve backward compatibility with legacy H.264 decoders, while enhancement layers are decodable by the SVC decoder only. The header of a SVC NALU is extended with extra fields providing information about the type of the enhancement information and their level in the hierarchy of the layers.

Figure 2 depicts the GOP structure of a typical H.264/SVC encoding scheme which encompasses the AVC-compatible base layer, one spatial and one temporal SVC enhancement layers. Each GOP is 16 frames long and every 32 frames a picture is Intra-coded, therefore the video encoding scheme is $I_0B..BP_{16}B...BI_{32}$. The AVC base layer is used for prediction of SVC enhancement pictures as indicated by the arrows while enhancement pictures with a higher Temporal Index (TID) value depends on those with lower TID value.

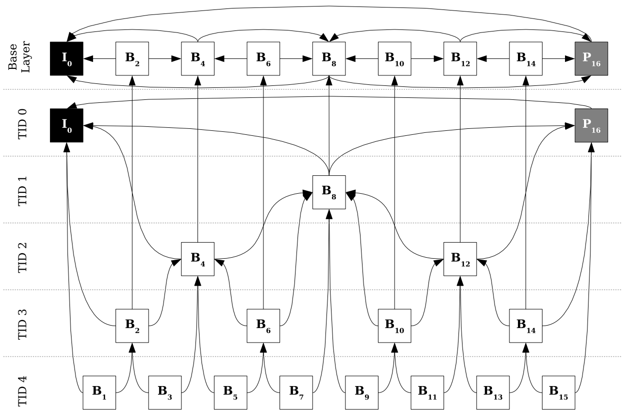

Fig. 2. GOP structure of the H.264/SVC encoding scheme used in this work.

Figures 1 and 2 show that, given a picture, the number of pictures depending on that one for decoding is highly variable. Since the loss of a picture usually propagates errors to all the pictures which depend on it, each picture should be protected according to its potential to propagate distortion to other pictures. Depending on the encoding scheme, different error propagation paths are possible. With the typical AVC encoding scheme shown in Figure 1 the loss of a P-type picture generates distortion which affects every picture which follows in the decoding order, until the next I-type picture is received. The loss of a B-type picture instead does not propagate distortion when considering the AVC codec. In the case of the SVC codec, which does use B-type pictures for prediction, the loss of any base layer picture generates distortion which affects the base layer and enhancement layers. The loss of an enhancement picture, instead, generates distortion which affects those enhancement layers with a higher TID index only.

III. The Prioritization Algorithm

In this section we illustrate a traffic prioritization algorithm for SVC encoded videos which operates on a GOP basis. An adaptation of the algorithm for standard non-scalable AVC videos is also presented. The algorithm, which is composed of two sequential stages, exploits the traffic differentiation capabilities of 802.11e to protect the most important parts of the video stream while achieving efficient network resource usage. The first stage operates on NALUs while the second stage packetizes NALUs and then assigns packets to access categories according to their importance.

A. Importance-Based NALU Ordering

NALUs are characterized by unequal relative importance which typically depends on their type. For instance, in AVC videos I-type pictures are more important than P-type and B-type ones. In SVC videos, the base layer is much more important than the enhancement layer. Although NALUs of the same type are characterized by comparable levels of importance, those required earlier in the decoding process are generally used to predict a higher number of pictures, thus they can be considered more important than the others.

The above considerations are at the basis of the operations of the first stage of the algorithm, which aims at sorting the NALUs of an AVC or SVC GOP in order of importance. Consequently, NALUs are sorted depending on their type.

For the AVC codec, three groups of NALUs are defined: I-type, P-type and B-type. As mentioned, the I-type group is more important than the P-type, which is in turn more important than the B-type. The type of a NALU can be determined by inspecting the *NRI* field in the NALU header, whose value depends on the picture type.

For the SVC codec, six groups are defined: the base layer and five enhancement NALUs groups. NALUs of the the most important base layer group are identified by exploiting the *type* header field. The *TID* field, which ranges from zero to four in the encoding scheme used in this work, provides the information needed to cluster SVC NALUs into one of the five proposed enhancement groups. Since enhancement NALUs are decoded in TID order as explained in Section II-B, the algorithm considers the groups containing NALUs with a lower TID value more important than other enhancement layer groups with a higher TID value.

Once every NALUs of the GOP has been assigned to a group, the algorithm extracts the NALUs one by one in picture decoding order starting from the most important group and queues NALUs in a temporary list which is therefore sorted in order of importance over the whole GOP.

For instance, in the case of the AVC encoding scheme depicted in Figure 1, NALUs would be sorted in the following order: $I_0P_3P_6...P_{27}P_{30}B_2B_3...B_{29}B_{31}$.

Similarly, in the case of the SVC encoding scheme in Figure 2, NALUs from the base layer group would be queued first in the order $I_0P_{16}B_8B_4B_{12}B_2B_6B_{10}B_{14}$. Enhancement NALUs would be then queued as $I_0P_{16}B_8B_4B_{12}B_2B_6B_{10}B_{14}B_1B_3B_5$-$B_7B_9B_{11}B_{13}B_{15}$, according to the TID value.

B. Packetization and Assignment to Traffic Classes

The second stage of the proposed algorithm operates on the list of NALUs produced by the first stage of the algorithm. First, each NALU is encapsulated into RTP packets according to RFC 3984 [8] (H.264/AVC) or [9] (H.264/SVC). Then, each RTP packet is assigned to an AC by the following algorithm. Since packets are considered in decreasing order of importance, the selection of the appropriate ACs is carried out starting from the one with the highest priority. Packets are assigned to the considered AC until it is full, i.e. the desired share of traffic for that AC has been achieved. Then, the process continues using the lower priority AC, until all packets have been assigned to an AC. By appropriately deciding the traffic share associated with each AC, the algorithm can be configured to deliver the video stream using any set of the four available 802.11e ACs: for example, half of the video traffic could be mapped to AC1 and half to AC2, reserving AC0 for VoIP and AC3 for background traffic delivery.

	AC3	AC2	AC1	AC0
Bitrate [kb/s]	33.6	560	720	1000

TABLE I
BACKGROUND TRAFFIC GENERATED BY EACH SOURCE NODE.

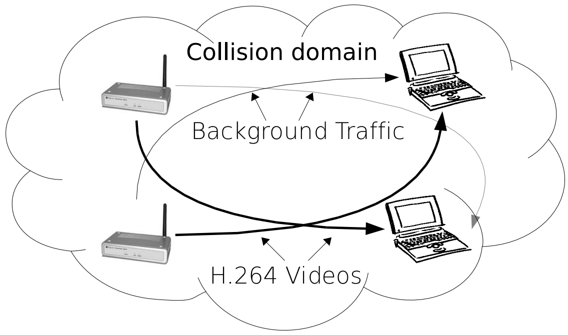

Fig. 3. The simulated network topology.

For the only purpose of simplifying the presentation of the results, the more important I-type (AVC) and base layer (SVC) groups of NALUs will be referred to as Gr_A in the remainder of the paper, while the less important P and B-type (AVC) and enhancement layer (SVC) groups will be referred to as Gr_B.

IV. SIMULATION SCENARIO

A. Network Setup

In a typical ad hoc network scenario multiple nodes contend for access to the channel while data flows of various nature (e.g. multimedia data and background traffic) compete for transmission opportunities within each node.

In order to investigate the performance of our algorithm in a typical scenario, an ad hoc wireless network consisting of several nodes located within the same collision domain was simulated using the NS2 network simulator [10]. Four nodes, two traffic sources and two traffic sinks, are simulated as shown in Figure 3. Within each source node, four UDP agents generate constant bitrate background traffic as it is summarized in Table I. Traffic in AC3, the highest priority class, simulates a bidirectional VoIP call. Traffic in AC2 simulates a constant bitrate video stream while traffic in AC1 and AC0 simulates generic background traffic that contributes to load the network. Finally, each of the two source nodes transmits an H.264 test video using the proposed traffic prioritization algorithm in order to assess its effectiveness. Thus, twelve data flows, two of which are H.264 video flows, offer an average traffic load of 5-6 Mb/s to the network.

The implementation of EDCA for NS developed by the Berlin Technishe Universität [11] is used. Such implementation was modified to introduce a timeout mechanism which drops packets after half a second of permanence in the queue, as recommended in the 802.11 standard.

Because our main focus is to investigate the effectiveness of the proposed traffic prioritization algorithm in exploiting the different QoS levels offered by 802.11e, in the simulations

Mobile	PSNR [dB]	BR [kb/s]	NALU#	Gr_A/Gr_B
AVC-1	31.60	807	298	1:2.34
AVC-2	32.19	916	961	1:2.79
SVC-1	31.57	795	600	1:2.29
Coastguard	PSNR [dB]	BR [kb/s]	NALU#	Gr_A/Gr_B
AVC-1	32.13	514	298	1:3.49
AVC-2	33.01	617	668	1:4.52
SVC-1	32.54	557	600	1:3.10

TABLE II
ENCODING CHARACTERISTICS OF THE VIDEOS USED IN THE SIMULATIONS.

we do not consider bit errors due to wireless channel noise. Therefore, packet losses are caused by the limited size of the transmission queues (set to 50 packets) as well as collisions in channel access.

B. Encoder Configuration and Simulation Testbed

The *Mobile* and the *Coastguard* sequences (CIF, 300 frames, 30 fps) are encoded using the H.264/AVC JM 10.2 and the H.264/SVC JSVM 9.4 encoders. Both encoders are configured to produce 32 frames long GOPs.

The JM encoder produces the two *AVC-1* videos listed in Table II using the scheme depicted in Figure 1, encoding one slice per picture and one NALU per slice. The JSVM encoder produces the two *SVC-1* videos using the scheme illustrated in Figure 2. The base layer is QCIF size (176×144 pixels), 15 fps, while the full resolution, full rate video is CIF size (352×288 pixels), 30 fps. The JM encoder has the ability to produce size constrained NALUs: this feature is capitalized to encode the two *AVC-2* videos, whose NALUs do not exceed the size of the network MTU (1400 bytes). This configuration is expected to improve the resilience to packet losses as no RTP fragmentation is involved.

Table II compares the characteristics of the six encoded videos (the rightmost column reports the size ratio between the Gr_A and Gr_B groups defined in Section III).

In our simulations, each encoded bitstream is fed to the traffic prioritization algorithm presented in Section III, which is tuned to accommodate one fourth of the GOP traffic in each AC to achieve an approximately equal usage of the four ACs. The algorithm produces a traffic trace which is loaded into NS and a packet loss trace is computed by simulation: the received video is reconstructed accordingly and decoded for visual inspection and PSNR computation. Simulations are run at three different physical channel rate, i.e. 11, 9 and 5.5 Mb/s. Thus, eighteen different scenarios are investigated. The amount of traffic offered to the network is equal in all simulations.

V. RESULTS

Table III reports, for each AC, the traffic loss rates (percentage of lost over sent data) at packet level for Mobile simulations. The results show the effectiveness of EDCA mechanism for traffic differentiation (AC3 is the highest priority class). As expected, packets mapped to high priority ACs receive a preferential treatment, while other packets are more likely to experience transmission delays, timeout expiration inside the MAC queue or even dropping before being queued due to the limited size of the queues.

5.5 Mb/s	AC3	AC2	AC1	AC0	Total
AVC-1	0.0	0.0	64.5	89.9	38.5
AVC-2	0.0	0.0	69.0	95.3	41.0
SVC-1	0.0	0.0	63.1	88.3	37.8
9 Mb/s	AC3	AC2	AC1	AC0	Total
AVC-1	0.0	0.0	0.0	71.9	17.9
AVC-2	0.0	0.0	0.0	79.3	19.8
SVC-1	0.0	0.0	0.0	65.8	16.4
11 Mb/s	AC3	AC2	AC1	AC0	Total
AVC-1	0.0	0.0	0.0	46.3	11.6
AVC-2	0.0	0.0	0.0	62.9	15.7
SVC-1	0.0	0.0	0.0	49.8	12.4

TABLE III
MOBILE SEQUENCE, PACKET LEVEL BYTE LOSS RATE [%] AT DIFFERENT SIMULATED CHANNEL RATES.

	5.5 Mb/s		9 Mb/s		11 Mb/s	
	Gr_A	Gr_B	Gr_A	Gr_B	Gr_A	Gr_B
AVC-1	0.0	66.1	0.0	28.6	0.0	20.3
AVC-2	0.0	54.0	0.0	26.4	0.0	20.8
SVC-1	0.0	89.9	0.0	27.6	0.0	22.4

TABLE IV
MOBILE SEQUENCE, NALU LEVEL BYTE LOSS RATE [%] AT DIFFERENT SIMULATED PHYSICAL CHANNEL RATES.

Mobile	5.5 Mb/s	9 Mb/s	11 Mb/s
AVC-1	19.25*	24.59*	27.09
AVC-2	19.45*	26.16	27.14
SVC-1	21.41	26.45	27.72
Coastguard	5.5 Mb/s	9 Mb/s	11 Mb/s
AVC-1	23.39*	27.50*	29.80
AVC-2	21.76*	27.82	30.35
SVC-1	23.69	28.00	29.33

TABLE V
AVERAGE PSNR OF THE RECEIVED VIDEOS AT DIFFERENT SIMULATED CHANNEL RATES.

The corresponding traffic loss rates at NALU level are reported in Table IV. Values are shown for the $Gr_A$and Gr_Bgroups of NALUs as defined in Section III. Coherently with Table III, loss rates of Gr_A NALUs, which are mapped to high priority ACs, are lower than those of Gr_B NALUs, mapped to low priority ACs.

Note that when transmitting an *AVC-2* video, every packet carries an independently decodable NALU. In the *AVC-1* and *SVC-1* cases a packet might instead carry just a fragment of a NALU. However, even if a single fragment is lost, the whole NALU is discarded. As a result, NALU loss rate in *AVC-1* and *SVC-1* simulations is higher than in the *AVC-2* case, despite in the latter case the packet loss rate is higher due to the higher video bitrate.

Table V shows that the AVC and SVC encoded videos presents similar PSNR trends and mean values when the physical channel rate is set to 9 and 11 Mb/s respectively. However, they significantly differ if the physical channel rate is reduced to 5.5 Mb/s.

In the *AVC-1* at 5.5 Mb/s scenario (Figure 4) the loss of Gr_B NALUs results in the loss of all the B-type and trailing

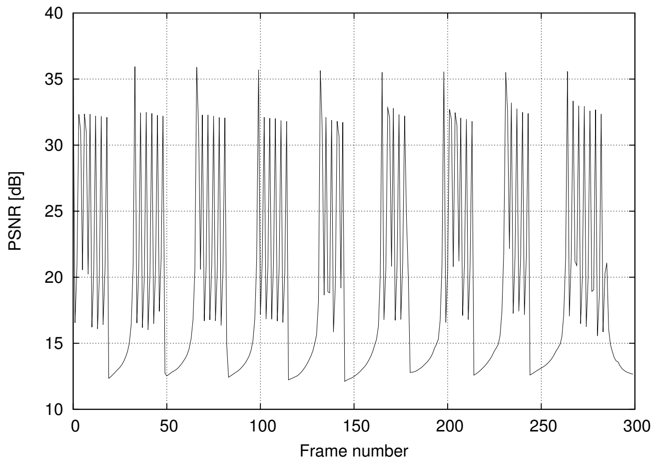

Fig. 4. Mobile sequence, AVC encoded, 5.5 Mb/s physical channel rate.

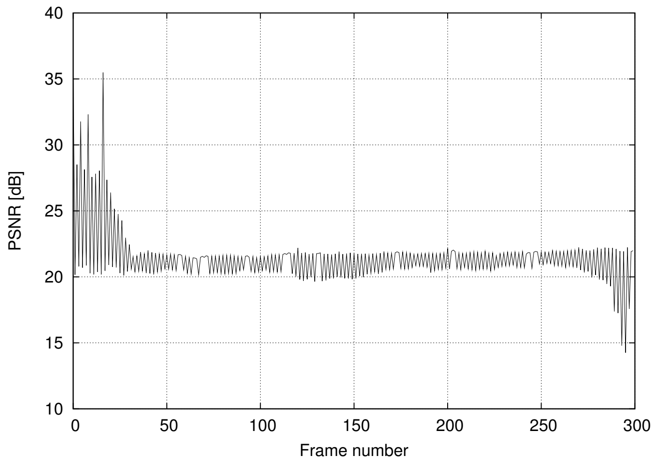

Fig. 5. Mobile sequence, SVC encoded, 5.5 Mb/s physical channel rate.

P-type pictures of each GOP. Lost pictures are concealed by the decoder using the previous available frame, thus producing frequent and long image freezes. The quality of the video is severely impaired, as the PSNR measures. It periodically drops in correspondence with each frame freeze and rises again in correspondence with the next correctly received I-type picture.

In the *SVC-1* at 5.5 Mb/s scenario (Figure 5) the enhancement layer is lost almost completely, but the base layer is received unaffected. The reconstructed sequence is less smooth than the original due to the loss of numerous odd frames which are concealed copying the previous even frame. Moreover, the sequence is somewhat blurry because some of the high spatial resolution frames are lost and concealed by up-sampling the corresponding lower resolution pictures of the base layer. However, since no long frame freezes are experienced differently from the case of the AVC-1 video, the playback of the reconstructed video is visually much more pleasant, despite pictures sometimes have less sharp details. The PSNR value drops almost immediately to the value corresponding to the base layer but remains stable during the whole simulation, hence providing an approximately constant quality throughout the playback.

The average PSNR values achieved with the proposed traffic prioritization technique are shown in Table V. The JM decoder was unable to decode the AVC videos marked with an asterisk: they were so severely affected by losses that we could decode them only using the more robust H.264 decoder provided by the FFmpeg project. In all other cases, both the decoders were able to decompress the received AVC videos, resulting in nearly identical PSNR values.

VI. Conclusions

In this paper we introduced a traffic prioritization algorithm for the transmission of both AVC and SVC encoded videos over 802.11e networks. The effectiveness of the algorithm is evaluated by simulation in an ad hoc network scenario. Our experiments demonstrate that the algorithm provides a comparable error protection level to AVC and SVC videos, yielding an acceptable quality of the reconstructed sequence even in poor network conditions. In addition, better visual quality can be obtained when the SVC codec is used in place of AVC for the same source coding bandwidth, in the same network conditions and using the same amount of network resources, especially in the case of high packet loss rate conditions. While the AVC codec offers in fact a very simple form of temporal scalability only, SVC provides advanced scalability options which enable more versatile traffic prioritization schemes such as the proposed one, which provides a better visual quality without consuming additional network resources.

References

[1] Joint Video Team of MPEG and ITU-T, "Joint draft 10 of SVC amendment (JVT-W201)," San Jose, California, July 2007.

[2] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," *ITU-T Recommendation H.264*, May 2003.

[3] "Wireless LAN medium access control and physical layer specifications amendment 8: Medium access control quality of service enhancements," *IEEE Std. 802.11e*, 2005.

[4] A. Ksentini, M. Naimi, and A. Guéroui, "Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture," *IEEE Communications Magazine*, January 2006.

[5] C. Foh, Y. Zhang, Z. Ni, and J. Cai, "Scalable video transmission over the IEEE 802.11e networks using cross-layer rate control," in *Proceedings of the IEEE International Conference on Communications*, 2007, pp. 1760–1765.

[6] 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, no. 3, pp. 317–330, Mar. 2007.

[7] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[8] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP Payload Format for H.264 Video," RFC 3984 (Proposed Standard), Feb. 2005. [Online]. Available: http://www.ietf.org/rfc/rfc3984.txt

[9] S. Wenger, Y. K. Wang, and T. Schierl, "RTP payload format for SVC video," draft-ietf-avt-rtp-svc-08 (work in progress), Feb. 2008.

[10] UCB / LBNL / VINT, "Network simulator - ns - version 2.31," http://www.isi.edu/nsnam/ns, 2007.

[11] S. Wiethölter, M. Emmelmann, C. Hoene, and A. Wolisz, "TKN EDCA model for ns-2," Telecommunication Networks Group, Technische Universität Berlin, Tech. Rep. TKN-06-003, 2006.