Content-adaptive traffic prioritization of spatio-temporal scalable video for robust communications over QoS-provisioned 802.11e networks

A. Fiandrotti, D. Gallucci, E. Masala*, J.C. De Martin

Control and Computer Engineering Department, Politecnico di Torino, corso Duca degli Abruzzi, 24 — 10129 Torino, Italy. Phone: +39-011-5647036, Fax: + 39-011-5647198

Abstract

In this work a low complexity traffic prioritization strategy to transmit H.264 Scalable Video Coding (SVC) video over 802.11e wireless networks is presented (the approach applies to any DiffServ-like network). The distinguishing feature of the proposed strategy is the ability to adapt the amount of error protection to the changing characteristics of the video content. First, we estimate the perceptual impact of data losses in the different types of enhancement layers, i.e., temporal or spatial, for a large set of H.264/SVC videos. The experiments show that perceptual impairments are highly correlated with the level of motion activity in the video sequence. Then, we propose a content-adaptive traffic prioritization strategy based on the identification of the most important parts of the enhancement layers of the video sequence by means of a low complexity macroblock analysis process. The prioritization algorithm is tested by simulating a realistic 802.11e ad hoc network scenario. Simulation results show that the proposed traffic prioritization strategy consistently outperforms, particularly for dynamic video sequences, fixed a priori approaches, as well as a traditional FEC-based UEP strategy.

Key words: H.264 scalable video, Robust 802.11e video communications, Content-adaptive prioritization

1. Introduction

The increasing availability of inexpensive devices with wireless communication capabilities is fostering the demand for reliable wireless multimedia communications. However, wireless devices are highly heterogeneous in terms of communication bandwidth and processing capabilities. Therefore, efficient mechanisms are needed to effectively handle such heterogeneity of networks and devices, such as scalable multimedia coding systems. For the case of video communications, advances in the development of video compression technologies, e.g., the recently standardized H.264/SVC [1], provide significant efficiency gains compared to previous standards [2], therefore the H.264/SVC is expected to be widely adopted in the near future. The H.264/SVC standard offers temporal, spatial and fidelity scalability mechanisms which allow to trade-off temporal and spatial video resolution as well as image quality for different bandwidth requirements and types of client devices (mobile devices, personal computers, etc.).

Wireless multimedia communications, and video in particular, are challenging because of the intrinsic unreliability of wireless channels. Robust error protection mechanisms are, therefore, necessary. Protection is usually provided either by means of application-level error control mechanisms or by exploiting the QoS capabilities offered by DiffServ-like networks. In the context of multimedia communications, unequal error protection (UEP) schemes are often employed to optimize resource usage by protecting each part of the multimedia stream in proportion to its perceptual importance [3]. For these schemes to be effective, a reliable data importance measure has to be defined. Such importance is often defined a priori, e.g., on the basis of the encoding algorithm. For non-scalable video coding, for instance, I- and P-frames are considered more important than B-frames, due to the higher number of frames which depend on them, there-

fore they are prioritized by transmitting them first as in [4, 5], or by assigning them a high network protection level supplied by a DiffServ-like network, such as 802.11e, as in [6].

Scalable video coding techniques are particularly well suited for UEP schemes, as they separate the content into layers with different levels of importance: a base layer which contains the most important information and one or more enhancement layers with less important refinement information. In a UEP scheme designed for scalable video coding, high priority is assigned to the base layer data to allow, even in the worst case, the reconstruction of a base quality version of the content, while low priority is assigned to enhancement layers, thus achieving an efficient trade-off between network resources and overall video quality [7]. Moreover, within the same layer, decoding dependency can be used to determine data priority [5], so that error propagation in case of data loss is mitigated. The work in [8] suggests an analytical formulation of an FEC-based UEP assignment problem for video transmission relying on SNR scalability over a wireless packet-erasure channel. However, for complexity reasons, the proposed solution is based on a heuristic algorithm. Others jointly consider different types of enhancement layers, as in [9], proposing an algorithm to optimize the performance under a given total bit rate constraint. Layers are first sorted according to some parameters such as temporal and SNR levels, quality improvement and source bitrate. Then, a decreasing amount of FEC is assigned to each layer as previously sorted, also considering the residual loss probability in each layer after FEC decoding. However, both [8] and [9] require to know the packet loss probability at the network level which might be difficult to determine and it might be strongly time variant, particularly in wireless networks. For scalable video transmission over burst-error channels, an adaptive UEP and packet size assignment scheme is proposed in [10]. Extensive comparisons with distinct scalable video transmission schemes using MPEG-4 fine granular scalability (FGS) are conducted, showing that the proposed transmission scheme can react to varying channel conditions with less and smoother quality degradation than the other considered scalable video transmission schemes.

Whenever multiple scalability options (e.g., spatial and temporal) are employed, as H.264/SVC allows, it is, however, not straightforward to decide which enhancement information should receive more protection. In some cases, the user might prefer to receive the temporal enhancement layer, thus trading off details for smoother motion, while in other cases, the spatial enhancement layer, thus experiencing better image sharpness at the cost of lower temporal resolution. As shown in [11], the user preference depends on the considered video content. For instance, for high-motion content such as sport clips, a higher temporal resolution, i.e., a higher frame rate, is visually more pleasant than an increased spatial resolution [11].

In this work we design a content-adaptive traffic prioritization algorithm based on motion-analysis and evaluate its performance with respect to fixed a priori UEP schemes relying on network-provided QoS classes and a traditional FEC-based UEP scheme [9]. Although the proposed approach is general and applies to any DiffServ-like network, we focus on 802.11e, which is increasingly adopted to provide QoS capabilities in an 802.11-based network architecture. First, a reference a priori UEP algorithm for H.264/SVC communications based on frame dependency information is designed and adapted to the 802.11e scenario. Then, the performance of such an algorithm is investigated in details by assessing the perceptual impact of data loss affecting different types of enhancement layers over a wide range of H.264/SVC encoded videos. The experiments suggest that the amount of distortion strongly depends on both the motion activity of the video sequence and the type of enhancement layer affected by losses. On the basis of these results, a low complexity algorithm is designed to dynamically adapt the protection strategy to the characteristics of the video content. The perceptual performance of the proposed algorithm is then evaluated using the NS network simulator [12] in an 802.11e network scenario. A comparison between the proposed content-adaptive technique, fixed a priori schemes and the FEC-based UEP scheme proposed in [9] is presented in details.

The rest of the paper is organized as follows. In Section 2 we briefly review the 802.11e network architecture and summarize the main features of the H.264/SVC video coding standard. Then, in Section 3, we present a UEP method for H.264/SVC communications over 802.11e networks. In Section 4 we perform a preliminary statistical analysis of the perceptual importance of the different enhancement layers. Such analysis is the basis of the proposed content-adaptive video prioritization algorithm, described in details in Section 5. Section 6 introduces the simulation scenario, followed by the simulation results in Section 7. Conclusions are drawn in Section 8.

2. H.264/SVC video over 802.11e networks

2.1. The IEEE 802.11 and 802.11e standards

The IEEE 802.11 standard for wireless communications [13] is widely adopted by many different types of devices. In its basic implementation, the standard supports unstructured communications by means of the distributed coordination function (DCF), which is based on a carrier sense multiple access with collision avoidance (CSMA/CA) channel access mechanism. Numerous amendments have been developed to improve some characteristics, e.g., to increase physical data rates or to add some new features. In particular, the 802.11e amendment [14] introduces the enhanced distributed channel access (EDCA) mechanism to provide a network-level prioritization mechanism in channel access. For each host station, four distinct transmission queues known as access categories (ACs) are introduced in place of the unique queue offered by the 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 transceiver has to wait before attempting to access the channel. Actual settings of such parameters are different for each AC, resulting in high priority and low priority queues characterized by a different probability 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.

2.2. The H.264/SVC coding standard

The H.264/SVC [1] amendment extends the earlier H.264/AVC standard [15] with spatial, temporal and fidelity scalability options, allowing to encode a video as an independently decodable AVC-compatible base layer and one or more SVC enhancement layers. The H.264/SVC inherited from H.264/AVC the partition of the encoder functionalities between a video coding layer and a network abstraction layer [2]. While the former layer encompasses all the encoder core functionalities (e.g., macroblock coding), the latter is responsible for the encapsulation of encoded data into independently decodable transport units known as NALUs. Each NALU is prefixed by a header whose fields describe the characteristics of the data it contains, such as the *Type* field which specifies whether the NALU contains AVC base layer or SVC enhancement layer data. Enhancement NALU headers provide additional information, such as the scalability level which is, therefore, easily accessible without decoding the whole NALU.
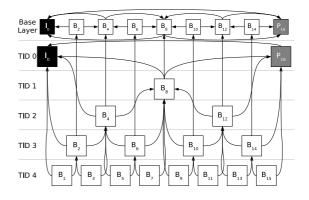

Figure 1: GOP structure of the H.264/SVC encoding scheme used in this work.

As in other video coding standards, frames are considered in groups called group of pictures (GOP) for encoding and decoding purposes. Such a structure is important since all coding dependencies are kept within the group. Figure 1 shows the H.264/SVC encoding scheme used in this work: the depicted GOP includes the AVC-compatible base layer, one spatial and one temporal enhancement layer. Each box represents a NALU, letters inside the box the corresponding picture type (intra, predictively or bipredictively coded) and subscript numbers define the picture display order. Each GOP is 16 frames long and every 32 frames a picture is intra-coded, therefore the video encoding scheme is $I_0B..BP_{16}B...BI_{32}$. As it can be surmised from Figure 1, the base layer provides the decoder with the data needed to reconstruct a half spatial resolution, half frame rate video. NALUs whose temporal index (TID), as defined in the standard [1] and shown in Figure 1, ranges from zero to three provide the decoder with the information needed to reconstruct a full spatial resolution, half frame rate video (i.e., composed by even frames only). Finally, NALUs with TID equal to four, i.e., the *temporal enhancement layer*, provide the remaining odd numbered pictures, which are needed to reconstruct the full frame rate video. In Figure 1 the arrows indicate the NALU decoding order, which takes place in increasing TID order.

To transmit the encoded video stream over the network, data must be encapsulated into packets. The IETF RTP protocol provides suitable mechanisms to transmit multimedia traffic over IP networks [16]. Transport of H.264/SVC data (i.e., NALUs) in RTP packets is currently undergoing standardization [17].

3

3. A priori H.264/SVC prioritization over 802.11e

In this section we introduce an *a priori* traffic prioritization strategy, that is an algorithm which determines the importance of each NALU on the basis of the encoding scheme of the video and protects it accordingly. The decoding dependencies depicted in Figure 1 suggest in fact that the TID value could be used as a coarse index of the importance of enhancement NALUs (lower values meaning higher importance). Indeed, the higher is the TID value of a given NALU, the lower is the number of NALUs affected by distortion in case that NALU is lost. This importance index for each NALU allowed us to design a simple a priori traffic prioritization algorithm [18] which exploits the traffic differentiation capabilities of a QoS-provisioned network architecture such as 802.11e.

The algorithm operates on a GOP basis and assigns a global priority value to each NALU. Apart from base layer NALUs, which have no TID value and are handled separately, NALUs are considered one group at a time, each group characterized by a specific TID value. The TID value can be easily determined by inspecting the enhancement NALU header. With the encoding scheme of Figure 1, six groups are considered, i.e., the base layer NALUs (without TID) and five enhancement NALU groups, corresponding to TID values from zero to four. Then, within each group, NALUs required earlier in the decoding process are given priority over the others. In the case of the SVC encoding scheme in Figure 1, NALUs from the base layer group are considered in the following decreasing priority order: $I_0 P_{16}$-$B_8 B_4 B_{12} B_2 B_6 B_{10} B_{14}$, then the less important enhancement NALUs as: $I_0 P_{16} B_8 B_4 B_{12} B_2 B_6 B_{10} B_{14} B_1 B_3 B_5 B_7$-$B_9 B_{11} B_{13} B_{15}$. This approach can also be applied to the case of more complex scalability schemes by forming more NALU groups determined by means of the additional information available in the extended SVC NALU header.

NALUs are then encapsulated into RTP packets [19] which are assigned to the 802.11e ACs depending on the priority corresponding to that of the NALU in the packet payload. In more details, the packets are assigned to the ACs in decreasing order of packet priority with a water-filling algorithm. When the desired share of traffic for a given AC has been achieved, the assignment process continues using the lower priority AC, until all packets have been assigned to an AC. By appropriately setting 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, a third of the video traffic could be mapped to AC2, a third

Sequence	Average bitrate [kb/s]	Encoding PSNR [dB]
Akiyo	66.50	37.38
Bus	489.00	30.46
Coastguard	337.82	30.28
Container	108.45	33.62
Flower	514.79	29.24
Football	571.57	31.07
Foreman	212.54	33.13
Irene	169.42	34.94
Mobile	501.62	29.52
Mother	80.58	36.68
News	165.16	35.13
Paris	305.61	31.99
Silent	150.87	33.30
Soccer	331.08	32.98
Students	141.74	34.28
Tempete	377.56	30.46

Table 1: Test videos characteristics: bitrate and encoding PSNR.

to AC1 and a third to AC0, reserving the high priority AC3 class for VoIP traffic.

Simulations with this traffic prioritization scheme in various network scenarios showed that the temporal enhancement layer is the first to be lost in case of data losses since it is transmitted in the low priority AC0 [18].

4. Importance analysis of enhancement layers

In this section we investigate the effects of two different traffic prioritization strategies on a set of H.264/SVC videos encoded with the spatio-temporal scalability scheme shown in Figure 1. For spatio-temporal scalable streams, as anticipated in Section 1, it is in fact not easy to determine the a priori prioritization scheme which optimizes the visual communication performance because such a decision depends on the video sequence characteristics. Table 1 reports the bitrate and encoding PSNR of the tested videos, which encompass different types of content (e.g., sport, news, etc.). The videos are encoded at CIF resolution (352×288 pixel), 30 fps, and their length ranges from 260 to 300 frames.

Two sets of experiments have been performed. First, the loss of the whole temporal enhancement layer has been simulated on our test video sequences. Therefore, according to the encoding scheme depicted in Figure 1, all odd-index frames are lost. The second experiment simulated the loss of the spatial enhancement layer. Preliminary, note that the loss of the NALUs with temporal

index zero, one and two systematically results in either undecodable or extremely distorted video sequences. Therefore, we investigate only the loss of NALUs with TID equal to three. Even if all NALUs whose TID value ranges from zero to three contribute to enhance the spatial resolution, for simplicity the term *spatial enhancement layer* is used in the following to indicate the set of NALUs whose TID is equal to three.

For completeness, all the three error concealment algorithms available in the JSVM encoder [20] were considered. A detailed description of their behavior can be found in [21] and it is briefly summarized in the following. The first algorithm is named *BLSkip* and it first upsamples the residual from the base layer (provided that it is available), then performs motion compensation using an upsampled motion field. With the second algorithm, named *Frame Copy* (FC), pixels of the concealed frame are copied from the corresponding position of the first frame in the reference picture list #0. Finally, the *Temporal Direct* (TD) algorithm conceals missing data by motion compensation from the available frames as if blocks were coded using the temporal direct mode. The motion vector value of a newly generated direct mode macroblock is estimated as the weighted mean between the motion vector values in the reference picture lists #0 and #1.

Table 2 presents the PSNR reduction with respect to the encoded video due to the loss of either the spatial or the temporal enhancement layer. In the case of the loss of the spatial enhancement layer, the BLSkip algorithm shows the smallest distortion for most of the video sequences. In the case of the Soccer sequence, for example, the distortion drops from about 8 dB (FC and TD) to less than 1 dB (BLSkip). In the case of sequences such as Akiyo or Container the distortion gain of the BLSkip algorithm over TD and FC is smaller, although BLSkip still achieves the best video quality on average. Hence the most effective strategy to conceal the spatial enhancement layer loss is to exploit the corresponding information in the base layer, especially in the case of video content such as Football, Soccer or Bus, since the FC and TD concealment algorithms are not able to form a prediction as good as the upsampled base layer due to the high amount of motion in the video scene. If the temporal enhancement layer is lost, a corresponding base layer is not available, therefore the BLSkip concealment algorithm cannot be used. Table 2 presents that the lack of such base layer information, as it happens for the case of the temporal enhancement layer loss, yields a lower reconstructed video quality for a significant subset of the test sequences (81%) compared to the case of the spatial enhancement layer loss.

The loss of the spatial or the temporal enhancement layers is evaluated using PSNR values, presented in Table 2, as well as by means of informal vision tests. Concerning PSNR values, for each sequence in Table 2, comparing the PSNR values in bold, which correspond to the best concealment strategies when the spatial or temporal enhancement layer are lost, allows to determine the most important enhancement layer, i.e., the one which would cause the highest PSNR decrease (Δ PSNR) in case of loss (marked with an asterisk in the table).

Informal vision tests confirm those results and give some insight on the perceptual quality experienced by the users. In more details, in the case of sequences characterized by fast movements of objects, such as Football or Soccer, the loss of the temporal enhancement layer results in a very annoying motion jerkiness effect, which is reflected by a nearly 5 dB PSNR drop with respect to the error-free video. The a priori, TID-based, traffic prioritization strategy proposed in Section 3 thus does not provide optimal results in such cases, as it would be better to give higher transmission priority to the temporal enhancement layer rather than to the spatial layer. For those sequences, in fact, the loss of the spatial enhancement layer causes a much less perceptible quality degradation, since the presence of the temporal enhancement layer minimizes motion jerkiness effects while the artifacts due to error concealment of the lost spatial enhancement layer are hardly noticeable. We attribute such behavior to the fact that in high-motion scenes the user attention is generally captured by fast-moving objects, thus artifacts in image details are less perceived.

For sequences such as News or Paris, in which objects move very slowly, no significant motion jerkiness is noticeable if the temporal enhancement layer is missing. For such sequences, instead, a higher quality degradation is observed when the spatial layer is lost, which is confirmed by an average PSNR reduction close to 2.5 dB.

5. Content-adaptive H.264/SVC prioritization

Since we observed that the visual performance is strongly influenced by the presence of fast-moving objects in the video, we use this information to decide which layer, temporal or spatial, should receive better protection on a GOP-by-GOP basis. In principle, this information could be precomputed and stored in, e.g., hint-tracks. However, this approach assumes the possibility of pre-computing information, that limits the applicability of the proposed techniques, for instance excluding scenarios such as live video streaming.

Sequence	Δ PSNR [dB] due to loss of				
	spatial layer			temporal layer	
	BLSkip	FC	TD	FC	TD
Akiyo	**0.17**	1.52	0.87	0.74	**0.26***
Bus	**1.07**	8.08	6.88	6.17	**5.40***
Coastguard	**1.47**	4.08	3.07	2.82	**1.86***
Container	0.31	1.06	**0.30***	0.47	**0.12**
Flower	**0.80**	7.17	6.20	5.44	**4.45***
Football	**0.66**	7.22	6.16	5.61	**4.80***
Foreman	**0.60**	5.85	4.47	3.78	**2.66***
Irene	**0.70**	4.50	3.46	2.98	**2.09***
Mobile	**0.76**	6.82	5.50	4.27	**2.53***
Mother	**0.24**	1.83	1.08	0.95	**0.45***
News	**2.52***	3.70	2.54	2.14	**1.42**
Paris	**1.97***	3.05	2.10	1.93	**1.13**
Silent	**0.24**	2.93	2.00	1.84	**1.14***
Soccer	**0.77**	8.95	7.64	6.40	**5.33***
Students	**0.26**	1.48	0.81	0.74	**0.34***
Tempete	**0.47**	4.00	2.69	2.56	**1.48***

Table 2: PSNR reduction due to the loss of either the spatial or the temporal enhancement layer. For each combination of sequence and type of lost layer, the best concealment algorithm is highlighted in bold. Moreover, for each sequence, an asterisk indicates the highest Δ PSNR value.
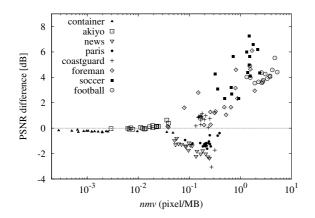

Figure 2: Scatter plot of PSNR difference, for each GOP, between the performance of the two a priori layer prioritization strategies (temporal minus spatial) versus the normalized motion vector length per macroblock (*nmv*).

Therefore, we propose a simple and fast macroblock motion measurement algorithm to decide which layer, temporal or spatial, should receive better protection on a GOP-by-GOP basis. As a first measure of the amount of motion in the scene, a normalized motion vector length per macroblock (*nmv*) is computed as the sum of the length (in pixels) of all motion vectors present in all frames contained in a given GOP, divided by the num-

ber of frames and number of macroblocks per frame. Figure 2 allows to investigate the relationship between *nmv* and the traffic prioritization strategy which minimizes the distortion. For each GOP of every video sequence, Figure 2 shows the *nmv* value and the difference in PSNR between the case of loss of the temporal enhancement layer and the case of loss of the spatial layer. Points with positive PSNR difference correspond to GOPs whose temporal enhancement layer is perceptually more important than the spatial enhancement layer. In this case, it is better to assign a higher protection level to the temporal enhancement layer than to the spatial enhancement layer. The opposite consideration holds for points with negative PSNR difference. Note that in Figure 2 part of the tested sequences have been omitted for picture clarity. However, sequences were carefully selected not to alter the trend which is present in the comprehensive picture.

Computing *nmv* might be, however, computationally expensive, thus we introduce the lower complexity motion measure M. To calculate *nmv* requires, in fact, to access motion vector information for each macroblock of the video stream. The low complexity algorithm we propose only relies on the macroblock type information to determine a motion complexity index M, ranging from zero to one. For every picture, the number of macroblocks coded in *Direct* mode, $N_{DirectMB}$, is determined. Such a number is then subtracted from the total
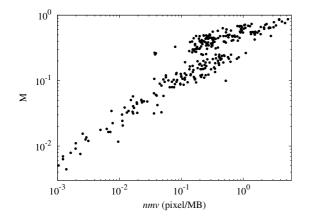
Figure 3: Correlation between the *nmv* and the *M* motion indexes. Each point represents a GOP of one of the tested video sequences.
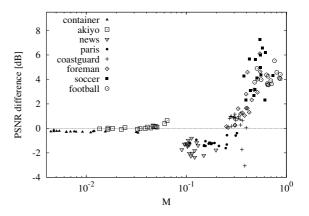

Figure 4: Scatter plot of PSNR difference, for each GOP, between the performance of the two a priori layer prioritization strategies (temporal minus spatial) versus the motion complexity index *M* of the GOP, as defined by Eq. (1).

number of macroblocks of the picture, N_{MB}. The difference is divided by N_{MB} and then averaged over the N pictures that constitute the GOP, as shown in Eq. (1):

$$M = \frac{1}{N} \sum_{i=1}^{N} \frac{N_{MB} - N_{DirectMB}}{N_{MB}}. \tag{1}$$

The underlying idea is that the direct mode is selected by the encoder when the motion vector prediction can be used as is to motion-compensate and encode the macroblock, i.e., the movement of the considered macroblock can be predicted with good accuracy. Therefore, the amount of macroblocks not coded in direct mode is an index of the motion "complexity" of the scene. Moreover, a high number of such non-direct macroblocks is more likely to cause jerkiness effects in the video sequence in case of data loss.

In Figure 3 a comparison between the *nmv* value and the proposed motion complexity index *M* is shown. When the *M* value is close to zero, i.e., nearly all macroblocks are encoded in direct mode, the magnitude of motion vectors is generally very low. When the *M* value is higher than a given threshold, e.g., about 0.3, more and more macroblocks are encoded without using the direct mode, i.e., with more complex prediction structures, and at the same time the variance of motion vector magnitudes rapidly increases.

In Figure 4 the relationship between *M* and the traffic prioritization strategy which minimizes the distortion is investigated. Similarly to Figure 2, for each GOP of every video sequence the *M* value and the difference in PSNR between the case of loss of the temporal enhancement layer and the case of loss of the spatial layer is shown.

Figure 4 shows that two main clusters of GOPs exist. The first cluster includes GOPs located in the leftmost area of the graph (low *M* values) and with negative PSNR difference: since they require better protection of the spatial enhancement layer, we refer to them as *static*. The second cluster includes GOPs located in the rightmost area of the graph (high *M* values) and with positive PSNR difference: since they require better protection of the temporal enhancement layer, we refer to them as *dynamic*. From Figure 4, a reasonable *M* value to discriminate between static and dynamic clusters of GOPs is in the 0.2–0.4 range. Analogously to Figure 2, in Figure 4 part of the tested sequences have been omitted for picture clarity. However, sequences were carefully selected not to alter the trend which is present in the comprehensive picture.

Given the preliminary analysis presented above, the implementation of a content-adaptive traffic prioritization strategy is briefly described in the following. For each GOP, the base layer is handled separately, i.e., it is packetized in compliance with the RTP standard [17] and mapped to the highest priority AC3 class, while enhancement layers with temporal index ranging from 0 to 2 are mapped to AC2. The algorithm then computes the GOP motion complexity index *M* using Eq. (1) and, by comparison against a threshold value, it classifies the GOP either as static or dynamic. If the GOP is classified as static, NALUs in the spatial enhancement layer are given priority over those of the temporal enhancement layer, and vice versa for dynamic GOPs. Then, after NALU packetization, packets are considered in decreasing priority order. Half of them are assigned to AC1, and
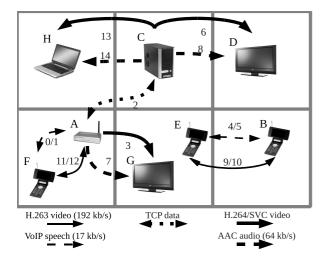
Figure 5: The simulated network topology, showing the analyzed video flow (i.e., flow #3) and all the interfering background traffic.

	AIFSN	CWmin	CWmax	TXOP limit (ms)
AC3	2	7	15	3.264
AC2	2	15	31	6.016
AC1	3	31	1023	0
AC0	7	31	1023	0

Table 3: Parameters of the four ACs, from the highest (AC3) to the lowest (AC0) priority, set as suggested in the IEEE 802.11e standard [14]. Unless otherwise noted, values represent the number of slots. Each slot lasts 20 μs.

the remaining ones are assigned to AC0. However, any other traffic share could be used, provided that higher importance packets are mapped to a better-service AC. Moreover, although the described implementation is designed for 802.11e networks, the approach can be applied to any DiffServ-like network.

Note also that the H.264/SVC standard also provides the priority ID (PRID) field in the NALU header, whose purpose is to signal the priority of the NALU. The standard does not specify any algorithm to compute the value of this field, thus the PRID field could be initialized using a value proportional to the importance of the NALU as determined by the proposed algorithm. In this case, media-aware network nodes, for instance those participating in an ad hoc wireless home networking scenario, could use such a value to determine which packets should be discarded first in case of network congestion to optimally control the degradation of video quality.

6. Simulation setup

Figure 5 illustrates the simulated network topology, which represents a typical domestic environment where different types of hosts are located in the various rooms of a building and operate within the same collision domain. The network is operated in ad hoc mode to allow any transmitter to directly send data to its destination, thus avoiding data duplication on the wireless channel as it would happen when nodes communicate in the presence of an access point. The NS network

simulator [12] and the 802.11e EDCA extension developed at the Berlin Technische Universität [22] are used to simulate all the details of the link layer behavior of each node. In the simulations, four ACs are considered. Their parameters have been set as summarized in Table 3, according to the suggestions of [14], in order to implement four different priority levels.

The exact meaning of each parameter can be found in [14] and it is summarized in the following. The number of slots of the Arbitration InterFrame Space (AIFSN) controls the amount of time a station must wait before beginning the channel contention procedure after the wireless medium becomes idle. The minimum and maximum values of the Contention Window (CWmin and CWmax) specify the limits of the random value, computed each time a new channel contention is performed, that determines the backoff duration, i.e., the time a station must wait before starting a transmission attempt. A random value is used to minimize the probability that two stations start transmitting at the same time. Finally, the transmission opportunity (TXOP) limit indicates the maximum amount of time a station has the right to initiate frame exchange sequences onto the wireless medium without restarting the channel contention procedure.

The priority of each AC is mainly controlled by the AIFSN parameter. The highest priority level AC3 is aimed at voice communications, that require low latency, thus a low AIFSN value is used. Since voice data units are generally short, the TXOP limit is low. For video communications, the AC2 level provides a latency similar to voice communications, but in case of contention among voice and video, the former takes priority since its CWmax is equal to the CWmin of the video traffic. However, the size of video data units (e.g., a frame) is usually larger than voice data units, therefore video has the right to keep the channel for a longer time (a higher TXOP value) once it has been acquired. The AC1 and AC0 are aimed at best effort and background traffic, respectively. They have higher AIFSN

values, thus a higher fixed latency before starting channel contention, as well as they do not have the right to keep the channel once acquired (TXOP=0): they must contend again to allow traffic in AC3 or AC2 to immediately take priority if present. Moreover, the CWmin of AC0 and AC1 is equal to the CWmax of AC2, therefore in case of contention with AC2 or AC3 traffic, the latter always takes priority.

Host A is a gateway which provides Internet access to all the other hosts in the building. Hosts G and D are digital TV sets which receive an H.264/SVC and an AAC audio stream each. Host F is a videoconferencing device which communicates with a remote host located in the Internet by sending and receiving VoIP traffic and low bitrate H.263 video. Hosts B and E are videophones which communicate among themselves, each generating traffic whose characteristics, in type and bandwidth, are similar to the ones of node F. Host C is a PC which exchanges data with a host on the Internet via a TCP connection, and it additionally acts as a domestic media server which streams AAC audio and H.264/SVC video to hosts D and H.

In such a scenario data flows with different bandwidth and delay requirements coexist. For example, a VoIP call requires limited bandwidth, although it loads the network with a high number of small packet which have tight maximum delay and jitter requirements. Videoconferencing traffic demands more bandwidth than VoIP and, similarly, requires timely packet delivery. Maximum delay requirements for streaming of pre-recorded contents are less stringent, albeit a minimum bandwidth is required to ensure a smooth playback. Therefore the four VoIP streams are delivered using the highest importance AC3 class provided by the 802.11e standard. Traffic class AC2 encompasses the four H.263 video streams and the 50% perceptually most important traffic, i.e., approximately all the base layer of the three H.264/SVC flows. AC1 encompasses the three AAC audio streams and the 25% of the SVC enhancement traffic, i.e., the most important part of the enhancement data. Finally, the background TCP traffic and the remaining H.264/SVC enhancement traffic are assigned to the lowest importance AC0.

Since we aim at investigating the performance of the proposed traffic prioritization algorithm in relation with the QoS capabilities offered by 802.11e, we prefer not to account the effect that the noise on the channel would have on the video transmission, thus no link error model is simulated. Nevertheless, in the considered scenario packet losses are caused by the limited size of the transmission queues (50 packets), together with collisions in channel access happening due to the heavy traffic sus-

tained by the network, whose channel bandwidth is set to 11 Mb/s. Transmission queues also implement, as suggested in the 802.11 standard, a timeout mechanism which drops packets after 0.5 s of stay in the queue. The network congestion level depends on the interfering background traffic and the bitrate of each test video, which varies from sequence to sequence as reported in Table 1.

Several simulation sets are considered. First, each sequence is transmitted using the content-adaptive traffic prioritization algorithm described in Section 5 and then, for comparison purposes, using both the a priori traffic prioritization strategy introduced in Section 3, which privileges the spatial layer (PRI-S), and a similar one which gives priority to the temporal as opposed to the spatial enhancement layer (PRI-T). Then, the influence of the *M* parameter is investigated and an optimal value is experimentally determined. For completeness, an analogous study is performed considering *nmv* rather than *M* as the parameter which differentiates between static and dynamic sequences. Finally, the performance of the proposed content-adaptive traffic prioritization algorithm is compared to the FEC-based UEP strategy proposed in [9]. In all simulations, fairness in the comparison of the various traffic prioritization strategies is ensured, for each sequence, by an identical allocation of video traffic among the 802.11e classes.

7. Results

7.1. Comparison with fixed prioritization strategies

Table 4, 5 and 6 present the simulation results for each combination of test sequence and traffic prioritization strategy. The H.264/SVC video under test is the stream #3 flowing from nodes A to G. For each scenario, the global byte loss rate (fraction of bytes lost on sent), the loss rate of spatial and temporal enhancement information and the PSNR of the reconstructed video are reported. Table 6 also reports the GOP classification produced by the content-adaptive algorithm: GOPs of a given sequence may be classified all as static (S), all as dynamic (D) or some as static and some as dynamic (S/D).

For the proposed content-adaptive prioritization strategy the byte loss rate ranges from as few as 1.87% (Container) up to nearly 20% (Football), depending on the network congestion level. The bitrate of each H.264/SVC video stream is varied, ranging from 66 kb/s (Akiyo) up to about 570 kb/s (Football). Losses mainly affect the low priority AC0 class, but when the network is heavily loaded, e.g., in the case of the Soccer sequence, also a few packets in the AC1 class are lost.

	Byte loss rate [%]	Spatial layer loss [%]	Temporal layer loss [%]	PSNR [dB]
Akiyo	1.65	0.00	17.12	**37.33**
Bus	15.69	0.00	72.11	**24.63**
Coastguard	8.16	0.00	80.50	**28.79**
Container	1.63	0.00	34.02	**33.59**
Flower	18.76	0.13	82.09	**22.08**
Football	19.31	0.21	67.14	**26.88**
Foreman	7.49	0.00	41.15	**32.09**
Irene	3.89	0.00	20.54	**34.51**
Mobile	18.14	0.12	82.62	**22.50**
Mother	2.32	0.00	17.29	**36.60**
News	4.62	0.00	45.91	**34.50**
Paris	7.91	0.00	76.09	**31.38**
Silent	3.98	0.00	27.18	**33.00**
Soccer	11.28	0.00	53.08	**30.07**
Students	2.69	0.00	30.36	**34.25**
Tempete	9.78	0.00	65.17	**29.49**

Table 4: Performance of the PRI-S fixed prioritization strategy which privileges the spatial layer as opposed to the temporal one.

	Byte loss rate [%]	Spatial layer loss [%]	Temporal layer loss [%]	PSNR [dB]
Akiyo	1.52	14.55	2.65	**37.33**
Bus	16.36	94.76	16.90	**26.93**
Coastguard	8.11	82.02	4.41	**28.86**
Container	1.18	17.55	0.23	**33.56**
Flower	19.14	95.77	7.99	**24.29**
Football	21.15	95.43	32.63	**27.95**
Foreman	7.64	49.61	14.49	**32.27**
Irene	4.04	24.89	6.07	**34.58**
Mobile	19.78	95.72	8.78	**23.30**
Mother	2.41	18.04	2.96	**36.61**
News	3.22	29.10	0.94	**34.38**
Paris	7.52	72.45	2.60	**30.61**
Silent	3.37	25.14	7.78	**33.11**
Soccer	11.47	66.87	14.36	**31.44**
Students	2.18	23.60	0.90	**34.28**
Tempete	10.21	81.18	15.56	**29.62**

Table 5: Performance of the PRI-T fixed prioritization strategy which privileges the temporal layer as opposed to the spatial one.

As it can be surmised from the table, the content-adaptive traffic prioritization strategy always assigns more protection to the enhancement layer identified as the most perceptually important one on a GOP basis. Therefore, the byte loss ratio for the temporal and spatial enhancement layers depends on the layer which is deemed the most important one. For instance, in the case of the very static Akiyo sequence packet losses only affect the temporal enhancement layer, while in the case of the very dynamic Bus sequence packet losses mainly affect the spatial enhancement layer. Conversely, the a priori strategy presented in Table 4 protects better the spatial layer, thus data losses mainly affect the temporal layer, and vice versa in Table 5. In the

Sequence	Type	Byte loss rate [%]	Spatial layer loss [%]	Temporal layer loss [%]	PSNR [dB]	Gain vs PRI-S [dB]	Gain vs PRI-T [dB]
Akiyo	S	2.00	0.00	20.79	**37.33**	**0.00**	**0.00**
Bus	D	15.24	95.43	11.80	**27.04**	**2.41**	**0.11**
Coastguard	S/D	8.04	56.83	26.74	**28.81**	**0.02**	**-0.05**
Container	S	1.87	0.00	39.41	**33.58**	**-0.01**	**0.02**
Flower	D	17.45	94.28	1.00	**24.94**	**2.86**	**0.65**
Football	D	19.70	94.88	27.54	**28.33**	**1.45**	**0.38**
Foreman	S/D	6.49	46.81	9.77	**32.51**	**0.42**	**0.24**
Irene	S	4.52	0.00	23.99	**34.43**	**-0.08**	**-0.15**
Mobile	D	19.10	95.23	2.38	**23.20**	**0.70**	**-0.10**
Mother	S	2.88	0.00	21.60	**36.58**	**-0.02**	**-0.03**
News	S	4.36	0.00	43.48	**34.52**	**0.02**	**0.14**
Paris	S/D	8.10	12.79	65.43	**31.28**	**-0.10**	**0.67**
Silent	S	3.87	0.00	26.29	**33.01**	**0.01**	**-0.10**
Soccer	D	10.42	66.76	9.88	**31.73**	**1.66**	**0.29**
Students	S	2.95	0.00	33.27	**34.24**	**-0.01**	**-0.04**
Tempete	D	9.04	79.11	9.12	**29.84**	**0.35**	**0.22**
Average					**31.34**	**0.61**	**0.14**

Table 6: Performance of the proposed content-adaptive strategy. The last two columns compare the performance with the PRI-T and PRI-S fixed prioritization strategies.

case of the Akiyo sequence, the PSNR performance is the same for all the techniques despite the fact that the byte loss rate slightly varies. This is due to the fact that the loss is in the background, and it can be effectively concealed. Therefore, such a variation in the byte loss rate has negligible effects on the PSNR performance. Moreover, since the bitrate of the Akiyo sequence is low, the byte loss rate can be easily influenced even by few packet losses.

The content-adaptive strategy greatly improves the quality of the received video in the case of markedly dynamic sequences. The Bus, Flower and Soccer videos show the most noticeable quality improvement, with PSNR gains up to 2.86 dB with respect to the PRI-S technique, despite nearly all the spatial enhancement layer data were lost. For these sequences, the visual inspection of the reconstructed videos shows less perceivable error concealment artifacts in comparison with the PRI-S technique. For sequences classified as static, e.g., Akiyo or Container, the content-adaptive traffic prioritization strategy provides, as expected, approximately the same performance as the PRI-S technique.

In the case of the Paris sequence, a slight performance degradation is recorded because two GOPs (the two rightmost points of Paris in Figure 4) exceed in motion the M threshold, set to 0.30 in these experiments, thus they are classified as dynamic even though they present

a negative PSNR difference value. The visual analysis revealed that, unfortunately, in those GOPs a very fast and localized motion activity is present while the rest of the scene is static. In these conditions, the M index tends to be high, nevertheless the best strategy would be to give priority to the spatial layer even for those GOPs. Therefore, in this particular case the PSNR performance is negatively affected, albeit minimally (0.1 dB).

Figure 6 provides a visual comparison between the proposed content-adaptive and the PRI-S techniques. Picture sharpness for the content-adaptive technique is reduced, yet the quality is strongly preferable to the one provided by the PRI-S technique. Moreover, the playback of the video is smoother and provides a better visual experience. The shadow effects seen in Figure 6 are due to the TD concealment algorithm used to reconstruct the frame when the temporal enhancement layer is missing. Such algorithm computes the missing frame as the average of the previous and the subsequent frames in the video sequence. If objects quickly move in the scene, there might appear shadow effects since objects might be in very different positions in the two frames used for the averaging operation. However, note that if the video sequence is played back at full speed, i.e., 30 fps, the perception of the shadow effect is minimized and the performance of the TD concealment technique appears similar to other techniques available

Figure 6: Example of the video quality achieved by the PRI-S technique (left) and the proposed content-adaptive traffic prioritization technique (right). Flower (top) and Football (bottom) sequences.
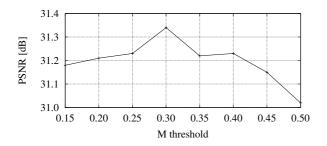

Figure 7: PSNR performance as a function of the M threshold for the proposed content-adaptive technique.

Sequence	M-based	nmv-based
Akiyo	37.33	37.32
Bus	27.04	27.21
Coastguard	28.81	28.81
Container	33.58	33.58
Flower	24.94	24.77
Football	28.33	28.12
Foreman	32.51	32.26
Irene	34.43	34.57
Mobile	23.20	23.50
Mother	36.58	36.62
News	34.52	34.56
Paris	31.28	30.58
Silent	33.01	33.12
Soccer	31.73	31.48
Students	34.24	34.28
Tempete	29.84	29.67
Average	31.34	31.28

Table 7: PSNR performance (dB) of the proposed content-adaptive strategy which discriminates between static and dynamic GOPs using a threshold on either M or nmv (M threshold=0.30, nmv threshold=0.06).
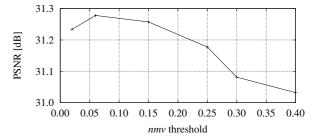

Figure 8: PSNR performance as a function of the nmv threshold for the proposed content-adaptive technique.

in the JSVM software.

For completeness, Table 5 reports the performance obtained by a fixed prioritization strategy (PRI-T), which privileges the temporal layer instead of the spatial layer. The last column of Table 6 reports the PSNR performance gain of the proposed technique compared to the data in Table 5. Performance is similar for a number of sequences, but others, such as Paris, present significant gains. Therefore, the use of an adaptive strategy such as the proposed one can improve the performance by automatically selecting the most appropriate prioritization strategy. With the considered set of video sequences, the average PSNR improvement with reference to the fixed prioritization strategies is 0.61 dB and 0.14 dB, compared to the spatial and temporal layer prioritization, respectively.

Another experiment evaluates the performance of the proposed adaptive strategy when the threshold value used to discriminate between static and dynamic GOPs is varied. Results are shown in Figure 7. For each threshold value on M, simulations are run, as previously described, for each video sequence. The plotted PSNR value is the average over all the considered video sequences. Figure 7 shows that indeed an M threshold equals to 0.30 is the optimal choice.

Analogously, the nmv value could be used to discriminate between static and dynamic GOPs, instead of using the M value. Figure 8 presents the PSNR performance of the proposed content-adaptive technique as a function of the nmv threshold, showing that the maximum performance is achieved when the nmv threshold is equal to 0.06. Table 7 compares the PSNR per-

Level	FEC-1	FEC-2
#1: Base layer	k=32	k=32
#2: TID=0 and TID=1	k=48	k=40
#3: TID=2 and TID=3	k=56	k=48
#4: TID=4	k=60	k=56

Table 8: Parameters of the FEC codes used in the simulations. n is equal to 64 in all simulations.

Sequence	FEC-1	FEC-2	Proposed technique
Akiyo	34.86	35.23	37.33
Bus	21.67	21.89	27.04
Coastguard	23.91	25.67	28.81
Container	31.28	31.89	33.58
Flower	20.81	21.57	24.94
Football	23.87	23.88	28.33
Foreman	27.76	29.06	32.51
Irene	30.85	31.73	34.43
Mobile	21.39	22.96	23.20
Mother	34.33	34.84	36.58
News	29.11	31.39	34.52
Paris	24.82	27.81	31.28
Silent	30.17	30.99	33.01
Soccer	25.29	26.15	31.73
Students	31.69	32.42	34.24
Tempete	25.41	26.59	29.84
Average	27.33	28.38	31.34

Table 9: PSNR performance (dB) of the proposed content-adaptive strategy compared to two FEC-based UEP strategies.

Figure 9: FEC-based packetization scheme.

formance, for each sequence, of the proposed content-adaptive technique when it employs either M or nmv to discriminate between static and dynamic GOPs. Both the M and nmv threshold values are set to the optimal values as previously determined. Considering the average over all the sequences, there is a slight advantage in terms of PSNR performance if the M value rather than the nmv value is used. Besides, the M-based technique is strongly preferable due to the lower complexity of computing M, as explained in Section 5.

7.2. Comparison with FEC-based UEP

A second set of simulations investigates the performance of the proposed content-adaptive strategy in comparison with an FEC-based UEP strategy. In particular, we refer to the algorithm proposed in [9]. A different protection level is applied depending on the video layer. Four levels are considered: base layer, TID equals to 0 and 1, TID equals to 2 and 3, and TID equals to
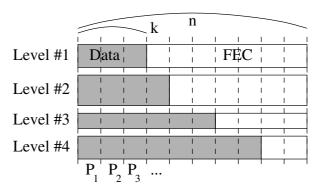
4. Levels are protected with error correcting codes with decreasing strength. In particular, we simulated a maximum distance separable (MDS) code, which is able to recover all packets in the block if at least k packets out of n are correctly received. The code is applied by considering one GOP at a time. As proposed in [9], a different FEC code is applied for each level, and packets are created by including data and FEC from all levels, as shown in Figure 9. Since it is not possible to determine, a priori, the optimal amount of redundancy for each level due to the difficulty in estimating the packet loss probability of each 802.11e service class, we experimented with various sets of codes; the code parameters are presented in Table 8. The n value is equal to 64 for all sets and levels. Since the video stream is already protected by FEC, simulations assumed that packets are transmitted in the AC1 class. Finally, note that, in order to fairly compare the FEC-based technique with the proposed one, we ensured that for each sequence the total bitrate, including the FEC overhead, is equal to the total bitrate used in the case of the content-adaptive prioritization strategy.

The PSNR performance is presented in Table 9. For convenience of comparison, the last column reports the PSNR performance of the proposed content-adaptive technique presented in Table 6. The performance gain between the proposed technique and the UEP strategy based on the FEC-1 set of codes ranges from 1.81 to 6.46 dB, achieved for the Mobile and Paris sequences, respectively. The average over all the sequences is 4.01 dB. The gain is slightly reduced considering the FEC-2 set of codes. In this case, it ranges from 0.24 to 5.58 dB, achieved for the Mobile and Soccer sequences, respectively, whereas the average is 2.96 dB.

8. Conclusions and future work

In this paper we proposed a content-adaptive traffic prioritization strategy for H.264/SVC communications over 802.11e wireless networks. By means of an extensive analysis on a large set of test video sequences, we show that a relationship exists between the level of motion in the video sequence and the relative perceptual importance of the temporal and spatial enhancement layers. We exploit such relationship to design a low complexity content-adaptive traffic prioritization algorithm. On the basis of simple features extracted from the compressed bitstream, the algorithm classifies each part of a video sequence either as static or dynamic, and selects the optimal traffic prioritization strategy accordingly. Simulations in a realistic 802.11e ad hoc wireless network scenario show that the proposed strategy consistently outperforms, particularly for dynamic video sequences, the traditional a priori approach, based on fixed layer prioritization, either spatial or temporal, the former being the classical error propagation minimization approach. A second set of simulations also shows how to optimally choose the value of the threshold involved in the proposed content-adaptive algorithm. Finally, further comparisons demonstrate that the proposed technique also outperforms an FEC-based UEP strategy.

Future work will be devoted to investigate the possibility of using a joint optimization approach to determine the mapping of the SVC layers to the 802.11e ACs, as well as to study how to use potentially available auxiliary information about the importance of each layer, e.g., provided by hint-tracks, to maximize the performance in the case of a non-live streaming scenario.

References

[1] ITU-T & ISO/IEC, Advanced video coding for generic audiovisual services — Annex G, H.264 & 14496-10 (Nov. 2007).

[2] 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 (9) (Sept. 2007) 1103–1120. doi:10.1109/TCSVT.2007.905532.

[3] A. Albanese, Priority encoding transmission, IEEE Trans. Inf. Theory 42 (6) (1996) 1737–1744.

[4] S. H. Kang, A. Zakhor, Packet scheduling algorithm for wireless video streaming, in: Proc. Packet Video Workshop, Pittsburgh, PA, 2002.

[5] Y. Cho, D.-K. Kwon, C.-C.J. Kuo, R. Huang, C. Lima, J.D. Black, J.J. Pan, Video streaming over 3G networks with GOP-based priority scheduling, in: Proc. Intelligent Information Hiding and Multimedia Signal Processing, 2006, pp. 201–204.

[6] A. Ksentini, M. Naimi, A. Guéroui, Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture, IEEE Communications Magazine 44 (1) (2006) 107–114.

[7] K. Gao, W. Gao, S. He, P. Gao, Y. Zhang, Real-time scheduling on scalable media stream delivery, in: Proc. Intl. Symp. on Circuits and Systems, Vol. 2, 2003, pp. 824–827.

[8] J. Xu, X. Yang, S. Zheng, L. Song, Group-of-pictures-based unequal error protection for scalable video coding extension of H.264/AVC, Optical Enginnering Letters 48 (6) (2009) 060502–1–3.

[9] A. Naghdinezhad, M.R. Hashemi, O. Fatemi, A novel adaptive unequal error protection method for scalable video over wireless networks, in: Proc. IEEE Intl. Symp. on Consumer Electronics, Dallas, TX, 2007, pp. 1–6.

[10] C.-W. Lee, C.-S. Yang, Y.-C. Su, Adaptive UEP and packet size assignment for scalable video transmission over burst-error channels, EURASIP Journal on Applied Signal Processing, Article ID 10131, 9 pages (2006). doi:10.1155/ASP/2006/10131.

[11] E. Akyol, A. M. Tekalp, M. R. Civanlar, Content-Aware Scalability-Type Selection for Rate Adaptation of Scalable Video, Article ID 10236, 11 pages (2007). doi:10.1155/2007/10236.

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

[13] IEEE, Wireless LAN medium access control (MAC) and physical layer (PHY) specification, 802.11 (1999).

[14] IEEE, Wireless LAN medium access control and physical layer specifications amendment 8: Medium access control quality of service enhancements, 802.11e (2005).

[15] ITU-T & ISO/IEC, Advanced video coding for generic audiovisual services, H.264 & 14496-10 (May 2003).

[16] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: A transport protocol for real-time applications, RFC 3550 (Jul. 1999).

[17] S. Wenger, Y. K. Wang, T. Schierl, RTP payload format for SVC video, draft-ietf-avt-rtp-svc-20 (Oct. 2009).

[18] A. Fiandrotti, D. Gallucci, E. Masala, E. Magli, Traffic prioritization of H.264/SVC video over 802.11e ad hoc wireless networks, in: Proc. of the IEEE ICCCN 17th International Conference on Computer Communications and Networks, 2008.

[19] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, D. Singer, RTP payload format for H.264 video, RFC 3984 (Feb. 2005).

[20] H.264/SVC reference software JSVM 9.4, http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm (2007).

[21] Y. Chen, K. Xie, F. Zhang, P. Pandit, J. Boyce, Frame loss error concealment for SVC, Journal of Zhejiang University SCIENCE A 7 (5) (2006) 677–683.

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