REAL-TIME MULTIPLE DESCRIPTION VIDEO STREAMING OVER QOS-BASED WIRELESS NETWORKS

R. Bernardini, M. Durigon, R. Rinaldo, P. Zontone

DIEGM, University of Udine,
Via delle Scienze 208
33100 Udine - Italy
{bernardini,durigon,rinaldo,zontone}@uniud.it

A. Vitali

ST Microelectronics
Via C. Olivetti 2
20041 Agrate Brianza, Italy
andrea.vitali@st.com

ABSTRACT

We consider the problem of robust video streaming over networks that support QoS differentiation, such as the 802.11e wireless network infrastructure. We consider the benefits obtained matching the H.264 Data Partitioning (DP) mode with such a QoS-based interface. We compare this solution with an innovative scheme which combines a Multiple Description (MD) coding framework with a QoS-based network. Results are reported using both a simple IID channel model and a more realistic wireless network, simulated using the OmNet++ network simulator.

***Index Terms*—** Video coding, Wireless LAN, Transport Protocols

1. INTRODUCTION

Transmitting video over wireless channels poses a set of new problems, due to the high packet loss ratios that are common over mobile radio channels [1]. On the other hand, multimedia streaming, and video streaming in particular, can be designed considering the fact that not all the encoded elements are strictly necessary to allow for an acceptable signal reconstruction at the decoder side. Using layered coding techniques [2] or scalable algorithms, we can identify the most important parts of the video stream the decoder should correctly receive in order to reconstruct an *acceptable* version of the original stream. To increase the overall performance such scalable or prioritized streams should be coupled with a channel that provides different interfaces with different Quality of Service (QoS) characteristics. These interfaces can be virtually created using a single physical layer and different amounts of Forward Error Correction (FEC) redundancy. Optimizing the amount of FEC redundancy to apply to the elements of a scalable stream is the well-known basic framework of the *Priority Encoding Transmission* (PET) scheme [3]. Due to the importance that multimedia streaming has in today network applications, the most recent networking standards have begun to explicitly introduce some form of QoS in the supported traffic. In particular, the 802.11e medium access control protocol is an emerging standard providing quality of service to the 802.11 family of wireless standards [4].

At the application level, the recent H.264/AVC has a set of advanced features to allow the transmission of video over a wide range of bit rates, and also over unreliable networks. In particular the layered streams obtained using the Data Partitioning feature of H.264 are very well matched with the QoS support offered by today networks. The match between the QoS support allowed by the lower layers of the protocol stack - i.e., the MAC layer in the 802.11e standard, and the application level packets produced by the video encoder is usually seen as a cross-layer approach [5].

In this paper we consider the problem of video streaming over a wireless link. We concentrate on the last hop of the communication network, i.e., from the wireless Access Point (AP) to the mobile user. We investigate the benefits of cross-layer design and compare two approaches with the classical non-layered solution. The first approach has been suggested in [5], and considers a scheme where the H.264/AVC layered stream and the 802.11e QoS support are used together. In [5], the performance is evaluated only on the basis of the residual packet error rate at the decoder. We consider here a specific concealment strategy and give explicit rate-distortion figures for the overall system performance. We then propose the second approach, which employs a simple multiple description (MD) coding scheme, and which uses the 802.11e QoS support to deliver with higher priority one description to the video client. We show that, when the highest priority channel offered by 802.11e is very reliable, the layered H.264/AVC solution is the best one in terms of received video quality. However, when the highest priority channel experiences even a very low packet loss probability, the MD solution outperforms the other scheme, both from an objective and from a subjective point of view. These results are obtained using simple channel models for the 802.11e interfaces. They are then confirmed using a more realistic simulation with the *OmNet++* network simulator.

2. QOS IN THE 802.11E STANDARD

802.11 WLAN can be interpreted as a wireless version of Ethernet, supporting best effort service [4]. The basic 802.11 MAC protocol is referred to as the *Distributed Coordination Function* (DCF); it operates according to the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) access strategy. Applying the DCF the access to the transmission media is individually decided by each station. When a station has a MAC Protocol Data Unit (MPDU) to deliver, it listens to the shared radio channel to detect if there are other transmissions in progress on the wireless medium. To further decrease the probability of collision - which may occur if more stations try to initiate their transmissions at the same time, the DCF applies a Collision Avoidance (CA) mechanism. After detecting the medium as idle, before initiating a transmission, stations keep sensing the medium for an additional random time called *backoff time*. A station initiates its transmission only if the medium remains idle for this additional random time. The duration of this backoff time in number of time slots is randomly selected between 0 and a Contention Window (CW) parameter, initially set to the minimum value CW_{min}. In the original version of the 802.11 standard [6], all the stations use the same value for CW_{min}, which leads to the same medium access priority for all stations.

Considering the high packet loss probability common in the wireless scenario, the 802.11 standard also introduces an acknowledgement (ACK) mechanism. For each successful reception of an MDPU packet, a receiving station immediately acknowledges the frame reception by transmitting an ACK frame back to the transmitting station. If this ACK is not received the transmitter of the original MPDU concludes that the packet was not delivered successfully and may repeat the transmission. The CW of a transmitting station is doubled when a transmission fails, up to a maximum value defined by CW_{max}. The basic set of parameters that are used by each station according to the DCF procedure are: CW_{min}, CW_{max}, RETRY_LIMIT. RETRY_LIMIT is the maximum number of consecutive transmissions of the same packet a station can try if it does not receive any valid ACK message from the receiver.

The 802.11e standard allows the support of differentiated QoS classes inside the 802.11 standard; the simplest and most common medium access control strategy provided by 802.11e is known as Enhanced Distributed Channel Access (EDCA). The basic idea of 802.11e is to allow multiple *backoff processes* - i.e., multiple queues - operating in parallel within one 802.11e station. Each backoff process is referred to as Access Category (AC). ACs are prioritized using AC-specific contention parameters, that form the so called EDCA parameter set. There are four ACs per station; thus four backoff entities exist in every 802.11e station. The EDCA parameter set specifies CW_{min} [AC], CW_{max} [AC], RETRY_LIMIT [AC], AIFS [AC], AC $= 0, ..., 3$[1]. By properly selecting these parameters ACs can implement different QoS classes.

3. H.264/AVC VIDEO OVER AN 802.11E LINK

The recent video coding standard H.264/AVC jointly proposed by ISO/IEC and ITU-T, improves compression efficiency and offers many new techniques to increase the robustness to packet erasures [7, 8]. One of these techniques is known as Data Partitioning (DP). DP assignes different syntax elements of the coded stream to three different partitions; partition A contains all control and header information, as well as any data related to the motion compensation process. Partition B contains intra-related information, whereas inter-related information - i.e., prediction errors - is assigned to partition C [9].

Decoding of partitions B and C is possible only if partition A is correctly decoded - partition A can be viewed as the base layer, while partitions B and C can be viewed as the enhancement layers in an equivalent layered scheme. Such a partitioned compressed stream is effectively sent over a channel that provides different QoS services, giving more priority to partitions A and B and allowing a worse equivalent channel for partition C. Such an approach has been followed in previous works on the transmission of layered video streams over a 802.11e wireless network [5]. As we do in this paper, if partition C only is lost or damaged during transmission, an effective concealment method can be implemented in inter slices using the motion vectors coded inside the DP A slice. We can reconstruct the corresponding macroblocks using only the prediction signal, without the prediction errors lost with DP C.

However, other coding schemes can be coupled with the availability of different QoS interfaces offered by today networks. In particular *Multiple Description* (MD) schemes create multiple representations of the original multimedia sequence [10]. Descriptions give independent representations of the original stream - as opposed to layered coding; by receiving more than one description the quality of the reconstructed signal is improved. In the following we refer to a very simple yet effective MD scheme based on spatial subsampling, which can be though as a particular implementation of the frame-based MD scheme presented in [11]. We create two descriptions by subsampling the original video stream by a factor of two along the columns. Each description is separately compressed using a standard H.264/AVC encoder and then it is sent to destination possibly on a different channel (Figure 1). Lost slices from one description can

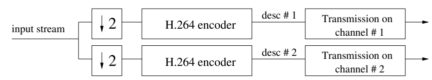

Fig. 1. In the MD scheme descriptions are obtained via spatial subsampling along columns; then each description is independently processed by an H.264/AVC video coder.

be easily recovered if the other description is correctly received. In particular, using the strong spatial correlation among descriptions, a simple bilinear interpolation process can be used to recover lost information. If both descriptions are lost the standard H.264/AVC concealment capabilities can be used. A key point is that each concealed frame is copied into the corresponding decoder frame buffer, Figure 2. This prevents error propagation from reference frames due to interframe coding.

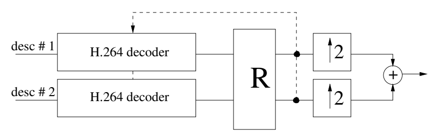

Fig. 2. The MD decoder copies concealed descriptions into the H.264 frame buffer in order to avoid error propagation due to interframe coding.

In the following section, experimental results refer to a Single Description (SD) scheme compressed using the H.264 coder with data partitioning (DP-SD scheme). We transmit the Sequence Parameter Set (SPS) and the Picture Parameter Set (PPS) packets [8] of the compressed stream using the highest priority AC, i.e., AC 3. These packets contain information that describes parameters relative to the entire sequence, or to a group of consecutive frames in the sequence. They must be correctly received in order to allow for a correct decoding process. DP A and DP B slices - i.e., motion vectors and intra coded slices, respectively - are sent using AC 2. DP C slices are sent using AC 1. In the MD scheme descriptions are coded without DP; SPS and PPS slices are still sent using AC3. Description 1 is sent using AC 2, while description 2 is sent over AC 1. The experiment setup is summarized in Table 1. The comparison is performed compressing MD descriptions such that they involve the same bandwidth occupation of the SD scheme in both access categories. We observe that in the MD scheme all received information is useful to improve the reconstructed video quality. This is not the case with the SD scheme: lost information in DP A - the base layer - makes unusable also the corresponding DP B and DP C packets, even if they are correctly received. A key point we try to investigate is to understand how sensible the SD scheme is to losses in DP A, and how it compares with a scheme based on MD coding.

[1]AIFS defines the *arbitration interframe space*, which is the time the medium must be idle before starting downcounting the backoff counter.

Table 1. Experiment setup

class	priority	DP	MD	bitrate
AC0	very low			
AC1	low	DPC (inter)	descr. 2	540 kbps
AC2	high	DPA (motion vect) DPB (intra)	descr. 1	460 kbps
AC3	very high	SPS-PPS	SPS-PPS	

4. EXPERIMENTAL RESULTS

In order to compare SD and MD schemes in different channel conditions we start by adopting a simple IID model for all access categories, with different packet loss probabilities. We consider 13 seconds of the Foreman CIF video sequence at 30 fps, coded at 1 Mbps. Using the H.264 data partitioning mode, the SD bandwidth occupation is about 460 kbps in AC 2 and 540 kbps in AC 1. As said before, in order to perform a fair comparison, description 1 of the MD scheme is coded at 460 kbps, while description 2 is coded at 540 kbps. We use the JM10.2 reference software implementation of the H.264/AVC codec with rate controller. Sequences are coded with GOP of the type IPPP..., with one intra refresh frame per second. Slices have a maximum size of 400 bytes. As said before, in the SD scheme, when only DP C is lost, motion vectors decoded from DP A are used to perform the error concealment process. When DP A is lost, or both descriptions that cover the same spatial region are lost in the MD scheme, the concealment procedures perform a simple macroblock replacement, copying the corresponding decoded macroblock from the previous reference frame.

In Figures 3-5 we depict the average quality of the reconstructed sequence taking the average over a large number of independent simulations. The quality is expressed in terms of the mean luminance PSNR. In these figures we assume that the packet loss probability in AC 3 is $P_{AC3} = 0$ - i.e., we assume that SPS and PPS packets are always correctly received. We further consider $P_{AC2} = 0$ in Figure 3, $P_{AC2} = 0.01$ in Figure 4, and $P_{AC2} = 0.03$ in Figure 5, while P_{AC1} is depicted along the x-axis of the figures.

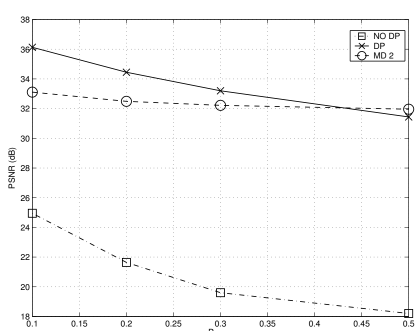

Fig. 3. Reconstructed video quality using an IID channel model, with $P_{AC3} = 0$, and $P_{AC2} = 0$.

A first observation is that using the DP-SD mode achieves a significant performance improvement compared with the SD mode without DP - in the latter case we transmit all video packets over AC1. When the high priority channel used for the DP-SD video streaming is very reliable - i.e., $P_{AC2} = 0$, Figure 3 - the DP-SD stream offers a significant gain also over the proposed MD scheme, at least for small values of P_{AC1}. In fact the DP mode of the H.264 standard implies a simple reordering of coded information; it is not afflicted by other significant coding losses. On the other hand, MD schemes realize a trade-off between coding efficiency and error robustness [10]. In the proposed MD scheme coding inefficiency is paid because two descriptions are independently coded, decreasing the efficiency of the H.264 coding engine.

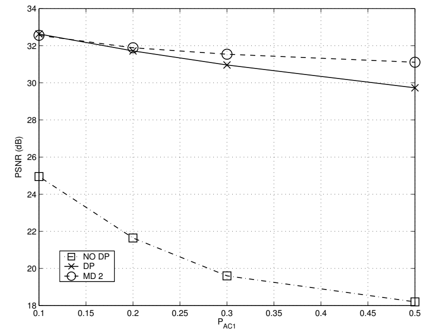

Fig. 4. Reconstructed video quality using an IID channel model, with $P_{AC3} = 0$, and $P_{AC2} = 0.01$.

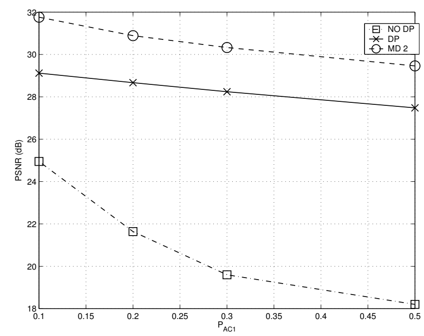

Fig. 5. Reconstructed video quality using an IID channel model, with $P_{AC3} = 0$, and $P_{AC2} = 0.03$.

However the robustness to packet erasures of the MD scheme becomes clear for increasing values of P_{AC1} in Figure 3, and especially when the high priority channel is not so reliable as shown in Figure 4 and Figure 5. Even for the very small values of P_{AC2} considered in these figures, the MD scheme significantly outperforms the SD scheme. This is mainly due to the fact that all correctly received information in the MD scheme is useful to improve the quality of the reconstructed sequence. The description sent over the higher priority channel allows for a basic video quality, which is improved by the second description. However, losses in the higher priority description can be still effectively recovered with the other description, if correctly received. On the other hand, the SD scheme is heavily affected by losses in DP A, which make useless related information in DP B and DP C.

These results are confirmed by simulations of video streaming over an 802.11b wireless infrastructure at 11 Mbps, carried out with the OmNet++ network simulator [12]. Different QoS interfaces are obtained according to the 802.11e MAC extensions. We use the EDCA access strategy discussed in Section 2; we adopt the AC parameters of Table 2, as proposed in [5]. The proposed network configuration is depicted in Figure 6. The video server is placed in the

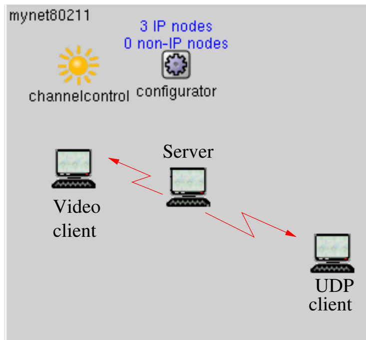

Fig. 6. OmNet++ wireless network infrastructure for SD and MD video streaming.

middle of a rectangular area covered by the wireless infrastructure. The video client - served respectively by the DP-SD stream or by the MD stream - moves in the allowed region according to the *RandomWPMobility* model offered by the simulator. The central server also sends an UDP stream at 1 Mbps to a mobile UDP client. This stream simulates the background traffic which may be present in a loaded wireless network. The physical channel model adopted for the simulation is provided in [13].

In Figure 7 we depict the reconstructed video quality for the MD and the DP-SD scheme, for different channel conditions. Channel

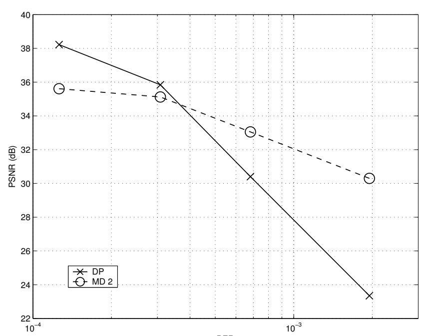

Fig. 7. Reconstructed video quality using the *OmNet++* simulator.

conditions are expressed in terms of the mean Bit Error Rate (BER) experienced by the video client towards the video server. The channel behaviour is changed by increasing the equivalent path loss α adopted in the channel model. These results confirm those discussed before using the IID channel model. When the channel experienced by the video client has low packet loss probabilities, the DP-SD scheme has very good performance. When the BER of the channel increases, again the MD scheme significantly outperforms the SD counterpart, both from an objective - i.e., PSNR - and from a subjective point of view.

5. CONCLUSIONS

The availability of a QoS-based network can significantly improve the quality of a video streaming application. One way to use this type of network is to use a layered coding scheme, such as the DP mode offered by the H.264 standard. If the highest priority channel is extremely reliable, this scheme provides very high quality video.

Table 2. EDCA parameter set adopted in simulations

AC	AIFS (μsec)	CW_{min}	CW_{max}	RETRY_LIMIT
AC3	50	7	15	8
AC2	50	15	31	8
AC1	50	31	1023	4
AC0	70	31	1023	4

However, if the highest priority channel experiences even a small packet loss probability, we found that a very simple scheme based on MD coding and a channel with QoS support can significantly outperform the classical solution.

6. REFERENCES

[1] B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," *International Conference on Image Processing (ICIP 2002)*, September 2002.

[2] T. Stockhammer and M. Bystrom, "H.264/avc data partitioning for mobile video communication," *IEEE International conference on Image Processing (ICIP 2004), Republic of Singapore*, October 2004.

[3] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Sel. Areas Comm.*, vol. 18, June 2000.

[4] S. Mangold, S. Choi, G. R. Hiertz, O. Klein, and B. Walke, "Analysis of ieee 802.11e for qos support in wireless lans," *IEEE Wireless Communications*, pp. 40–50, December 2003.

[5] 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*, pp. 107–114, January 2006.

[6] I. technology Telecommunications, information exchange between systems Local, and metropolitan area networks Specific requirements, "Part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications," *IEEE 802.11 WG, ref. no. ISO/IEC 8802-11:1999(E) IEEE Std. 802.11*, 1999.

[7] "Itu-t rec. h.264 — iso/iec 14496-10 avc," tech. rep., Joint Video Team of ITU-T and ISO/IEC JTC 1, March 2003.

[8] T. Wiegand, G. J. Sullivan, G. B. ntegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, July 2003.

[9] S. Wenger, "H.264/avc over ip," *IEEE Transactions on Circuits and Systems for Video Technology*, July 2003.

[10] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, September 2001.

[11] R. Bernardini, M. Durigon, R. Rinaldo, and A. V. P. Zontone, "Frame-based multiple-description video coding with extended orthogonal filter banks," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 3282–3296, August 2005.

[12] "http://www.omnetpp.org."

[13] "http://typo3.cs.uni-paderborn.de/en/research-group/research-group-computer-networks/projects/chsim.html."