Cross-Layer Joint Optimization of FEC Channel Codes and Multiple Description Coding for Video Delivery over IEEE 802.11e Links

Simone Milani*, Giancarlo Calvagno*
Dept. of Information Engineering (DEI),
University of Padova, Padova, Italy
E-mail: {simone.milani,calvagno}@dei.unipd.it
*Corresponding author

Riccardo Bernardini, Pamela Zontone
DIEGM,
University of Udine, Udine, Italy
E-mail: {riccardo.bernardini,pamela.zontone}@uniud.it

Abstract: This paper presents two cross-layer optimization strategies based on the IEEE 802.11e standard that enable a robust video transmission using adaptively Forward Error Correction (FEC) channel codes at transport layer and a Multiple Description Coding (MDC) architecture. The first approach classifies the characteristics of the sequence to be coded and selects the most appropriate coding mode according to the channel conditions obtained through a cross-layer signalling protocol. The second relies on a parametric model of the distortion which is estimated during the coding operations. The performances of these schemes are then improved by a packet classification strategy based on their significance in the decoding process. Experimental results show that both cross-layer optimization algorithms perform well with a small computational effort but different playout delays. The packet classification strategy improves the performance of the mode switching scheme obtaining a PSNR increment of 1.5 dB at max.

Keywords: Multiple Descriptions; cross-packet codes; packet classification; cross-layer optimization; IEEE 802.11e

Reference to this paper should be made as follows: Milani, S., Calvagno, G., Bernardini, R., Zontone, P. 'Cross-Layer Optimization Strategies for Video Transmission Over IEEE 802.11e Networks', Int. J. Internet Protocol Technology

Biographical notes: Simone Milani (MA ‘02, Ph.D. ‘06) is Research Assistant at the Department of Information Engineering at the University of Padova, Padova, Italy (url: http://www.dei.unipd.it/ simmil). Giancarlo Calvagno is Assistant Professor at the Department of Information Engineering at the University of Padova, Padova, Italy. Riccardo Bernardini is Researcher at the DIEGM Department the University of Udine, Udine, Italy. Pamela Zontone (MA’04, Ph.D. ’08) is Research Assistant at the DIEGM Department the University of Udine, Udine, Italy. Their research work mainly concerns video and image compression, source and channel coding, video transmission over unreliable networks. The work was developed under the PRIN research project “Reliable Multimedia Transmission over non-Reliable Networks: Advanced Source/Channel Coding Techniques,” as part of a joint effort of the Universities of Genova, Padova and Udine (url: http://www.diegm.uniud.it/rinaldo/PRIN/).
1 Introduction

The advent of wireless multimedia communications has brought to evidence that the traditional coding schemes and transmission protocols are not adequate for video delivery over time-varying channels. The occurrence of information losses, the delays in accessing the transmission medium, and the varying bandwidth that characterize wireless channels severely degrade the quality of the video sequence reconstructed at the receiver, and as a consequence, designers have endowed the modules at different protocol layers with new tools and strategies that improve the robustness of the video transmission. At the application layer, new video coding paradigms, like Multiple Description Coding (MDC), have been designed in order to suit the characteristics of the network over which the video data are transmitted (Goyal, 2001). At the transport layer, Forward Error Correction (FEC) codes have been introduced in order to allow the recovery of the lost information whenever the amount of lost data does not overcome a certain threshold (Rosenberg and Schulzrinne, 1999). At the network and MAC layers, different Quality-of-Service (QoS) classes have been defined in order to handle and transmit the packets with different priorities and loss probabilities (IEEE 802.11e standard, 2005). Despite each of these techniques can be independently optimized thanks to the modularization of traditional protocol stacks, the quality of the sequence reconstructed at the decoder is significantly improved by a joint orchestration of all these elements that permits adapting each tool according to the characteristics of the coded video sequence and the network conditions (Katsaggelos et al., 2005). In fact, the modularization of traditional protocol stacks permits great flexibility and interoperability among heterogeneous networks and devices, but at the same time, it could lead to significant inefficiencies considering the issues raised by the transmission of video sequences over wireless communication systems (Kawadia and Kumar, 2005).

During the last years a considerable research effort has been made to investigate efficient cross-layer (CL) solutions that grant a high Quality-of-Service (QoS) level to the end user by allowing a synergetic interaction between different protocol layers. The main goal of these architectures is to improve the performance of the transmission by jointly tuning the parameters of each layer according to holistic algorithms. As an example, the algorithms proposed respectively by Qu et al. (2006) and by Milani et al. (2005) apply an Unequal Error Protection (UEP) to the transmitted packets by adapting the FEC coder at transport layer to the characteristics of the coded video signal. Other solutions rely on the possibility of retransmitting a packet at network layer. A first approach based on a feedback channel that signals whether the information has been lost or not was introduced by Girod and Färber (1999). Later Zhai et al. (2006) presented a retransmission policy that can adaptively choose whether to retransmit or to protect the current packet by adding redundant information in the packet stream. The performance of the transmission can be further improved by employing QoS-aware transmission protocols that differentiate service classes. As an example, Ksentini et al. (2006) proposed a strategy based on the IEEE 802.11e MAC protocol that provides a different Quality-of-Service level to each packet according to the coded syntax elements. This algorithm was compared by Bernardini et al. (2007) with a cross-layer MDC scheme that assigns a different class of service to each description.

In this paper, we present two CL strategies that adaptively combine the MDC scheme proposed by Bernardini et al. (2007) with cross-packet FEC channel coding according to the transmission rate and the channel conditions. The proposed approaches partition the available bandwidth either between the two MDC coded descriptions (MDC option) or between the source and the FEC channel coder (SD+FEC option) according to the characteristics of the coded sequence and the power level measured at both the transmitter and the receiver. The first approach classifies the input sequence according to its characteristics and needs to buffer a whole Group-of-Pictures (GOP) before coding the video segment using the H.264/AVC video coder (JVT, 2002), while the second solution adopts a parametric model to estimate the channel distortion and does not require any buffering. Then, the final performance of the MDC scheme is further improved by a cross-layer packet classification strategy that assigns the coded RTP packets to different QoS classes according to their significance in the decoding process and the description they belong to.

In the following, Section 2 provides a general overview of the CL architecture, while Section 3 describes the adopted optimization strategies to switch from the MDC to the SD+FEC options (presented in Section 3.1 and Section 3.2 respectively). Section 4 reports two classification strategies for MDC and SD+FEC RTP packets respectively that significantly improve the quality of the sequence reconstructed at the decoder, while Section 5 reports the rate control algorithm adopted to combine the SD+FEC and MDC approaches. The experimental results reported in Section 6 show that the proposed algorithms are able to improve the schemes by Rosenberg and Schulzrinne (1999) and by Bernardini et al. (2007) with a negligible computational effort. Conclusions are drawn in Section 7.

2 The basic architecture

The adopted CL architecture combines a two-descriptions MDC scheme with a FEC coder operating on RTP packets and an IEEE 802.11e MAC layer unit (see Fig. 1). At the application layer, a single frame can be either coded directly as a single description (SD) or divided into odd and even rows of pixels, which are coded separately by two independent H.264/AVC coders (JVT, 2002). At the MAC layer, each packet is assigned to a different Access Category (AC) defined within the standard IEEE 802.11e according
to different criteria. In the following, the building blocks of the proposed CL scheme are presented in detail.

2.1 SD+FEC source-channel coding

In case the CL control unit has chosen the SD option, the input frames are coded by an H.264/AVC video coder into a fixed number of RTP packets without any previous subsampling. At the transport layer the FEC coder includes the video RTP packets related to a single frame in the columns of a matrix (one per column) and generates $n$ redundant columns applying a Reed-Solomon (RS) code on the rows of bytes in the filled matrix (Rosenberg and Schulzrinne, 1999). The adopted channel coding approach generates $n$ FEC columns that are packetized into $n$ additional RTP packets and included in the original packet stream produced by the H.264/AVC coder.

At the receiver, the adopted approach is able to recover up to $n$ lost RTP packets per coding matrix. In order to equalize the amount of used bandwidth and the quality of the coded video stream, the number $n$ of additional FEC packets is initially fixed and equals the redundancy $r$ introduced by the MDC scheme with respect to the SD approach. However, the recovering performance can be improved by modulating the loss probability for different packets by increasing or decreasing the number $n$ of additional FEC columns according to their significance in the decoding process (i.e., according to their coding type or the characteristics of the coded signal). This issue will be discussed with more details in Section 3.2.

2.2 MD source coding

In case the MD option is adopted, the packet streams generated by each H.264/AVC coder are directly transmitted to the lowest levels of the protocol stack without any FEC channel coding stage. At the receiver, in case both H.264/AVC decoders correctly get the packets of the corresponding description, the coded frame can be reconstructed without any further quality loss. In case some parts of one description are missing, the MD concealment unit estimates the lost rows through a simple bilinear interpolation involving the rows of the other description, which have been correctly decoded and are interleaved with lost ones. This estimation produces a degraded version of the coded frame, which is however, in most cases, a good reconstruction of the missing parts thanks to the spatial correlation. Note that in case both descriptions get lost, it is necessary to adopt the same error concealment techniques that are adopted for SD coded packets that can not be recovered by the FEC redundant information, and therefore, the probability that this condition is verified must be minimized by an appropriate packet scheduling.

To achieve this, the proposed MDC scheme takes advantage of the possibilities offered by the IEEE 802.11e protocol, which will be described in the following subsection.

2.3 The IEEE 802.11e protocol

The basic 802.11 MAC protocol, named Distribution Coordination Function (DCF), operates according to the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) access strategy. Before a station starts transmitting a queued MAC Protocol Data Unit (MPDU), the channel has to remain available for a random time interval, called backoff time, that varies in the interval $[0, CW]$, where the Contention Window parameter $CW$ is initially set to $CW_{min}$ and is doubled every time the transmission fails (up to $CW_{max}$). Whenever the packet is not correctly acknowledged by the receiver, the station retransmits it.
until the maximum number RETRY\_LIMIT (RL) of trials is reached.

In order to support different levels of QoS, the IEEE 802.11e standard (IEEE 802.11e standard, 2005) introduces the Enhanced Distributed Channel Access (EDCA) strategy, where multiple backoff processes (up to four) are allowed by distinguishing multiple packet queues within the same wireless station (see Fig. 1). Each queue is referenced by an Access Category (AC) label, which can be characterized by a different set of parameters $[\text{CW}_{\text{min}}, \text{CW}_{\text{max}}, \text{RL}]$ and a different priority. It is possible to distribute the RTP packets among the different queues in order to optimize the quality of the sequence reconstructed at the decoder.

As for the SD+FEC setting, the packets relative to Intra (I-type) frames are mapped to the Access Category with the highest priority (i.e., AC3), while the packets of Inter (P-type) frames are mapped to AC2 (see Bernardini et al. (2007) for more details). In case the MDC configuration is adopted, Inter coded packets are sent to AC2 or AC1 depending on whether they belong to description 1 or 2.

The physical layer (PHY) of the transmitting station reports the transmission power at the antenna to the CL control unit. In a similar way, the PHY module of the receiver communicates the power of the received signal to the CL unit of the station, which will report this measure to the transmitter via control packets. Hence, the transmitter will be able to compute a Bit Error Rate (BER) estimate of the last transmission session. According to this value, the CL control unit sets the coding parameters and chooses the most appropriate protection strategy, as it will be presented in the next section.

2.4 Cross-layer optimization of the different units

As Fig. 1 shows, the orchestration of the whole CL architecture is performed by the CL control unit, which selects the source coding technique (either SD+FEC or MDC), assigns the target bit rates to the source and channel coders, and chooses the ACs at MAC layer. At the beginning of each GOP, the optimization algorithm estimates whether the MDC coding scheme or the SD scheme with additional FEC packets (SD+FEC) is more appropriate for the current channel conditions. Previous works have shown that for high Bit Error Rate (BER) values the MDC option gives a better quality for a given coding rate, while the SD+FEC configuration proves to be more effective at low BER values. Unfortunately, the research activity has not reported so far a practical approach that permits detecting the most appropriate video coding solution between SD+FEC and MDC given certain network conditions. Fig. 2 show how the PSNR vs. BER curves for the SD+FEC and the MDC approaches intersect at the crossing point $\text{BER}_X$. Comparing the estimated threshold with the actual BER measurement from the network permits detecting the best coding choice for the current segment of video.

Moreover, it is possible to take advantage of the QoS differentiation provided by the MAC layer protocol in order to improve the quality of the reconstructed sequence. In case of congestions or contentions in channel access, an accurate classification of video packets permits transmitting the most important information for the decoding process with the highest priority and discarding those data that can be easily estimated from the received ones. As a result, the quality of the reconstructed sequence improves even if the amount of lost data increases provided that the CL control unit assigns the most suitable AC to each packet.

In the following, we will present different CL optimization strategies for packet classification and mode switching between SD+FEC and MDC. In the end these approaches will be merged together into a common CL classification strategy that maximize the quality of the reconstructed sequence.
3 CL optimization of SD+FEC and MDC schemes

3.1 The LBG-based estimation of BER$_{X}$

The robustness of SD+FEC and MDC schemes is strictly related to the spatial and temporal correlations existing in the original video sequence. Since the FEC coder permits recovering the lost information in case the number of lost packets is lower than a given threshold, at high BER values the FEC code proves to be ineffective and the lost parts of the current frame have to be estimated from the previous ones exploiting the temporal correlation. The MDC coding permits estimating the lost information from the correctly-received descriptions, and therefore, at high loss percentages the quality degradation results lower with respect to the SD+FEC coder. Whenever the temporal correlation among adjacent frames is lower than the spatial one, a better concealment can be obtained using the MDC configuration for small BER values too. As an example, the graph related to the sequence crew in Fig. 2(a) shows that the crossing point BER$_{X}$ is smaller than the one for the tempete sequence (see Fig. 2(b)). In fact, the crew sequence presents a low temporal correlation, and therefore, the spatial correlation involved in the MDC error concealment proves to be more effective in estimating the lost information for smaller BER values with respect to the sequence tempete. On the other hand, the strong temporal correlation between adjacent frames of the sequence tempete allows the SD error concealment unit to effectively estimate the lost information at higher BER values.

As a consequence, we need a metric to measure the level of spatial and temporal correlation. In the literature, a widely-used parameter to characterize the residual signal energy after a prediction is the Sum of Absolute Differences (SAD), which characterizes the correlation between the original signal and the predicted one. In our scheme, we considered two SAD metrics

\[
\text{diff}_x = \frac{1}{N_x N_y} \sum_{x,y=0}^{N_x-1, N_y-1} |I_t(x, y) - I_{t-1}(x, y)|
\]

\[
\text{diff}_x = \frac{2}{N_x N_y} \sum_{x,y=0}^{N_x/2-1, N_y-1} |I_t(2 x, y) - I_t(2 x + 1, y)|,
\]

(1)

where $N_x, N_y$ are the number of rows and columns for each frame respectively, and $I_t(x, y)$ is the luminance pixel value of the $t$-th frame at position $(x, y)$. The SAD value $\text{diff}_t$ parameterizes the temporal correlation between adjacent frames, while the term $\text{diff}_x$ measures the spatial correlation existing between odd and even rows. In this way, it is possible to estimate how effective can be the error concealment algorithms related to the MDC coding approach and the SD coding, respectively.

Before coding, a whole GOP is stored in memory, and the average values $E[\text{diff}_t]$ and $E[\text{diff}_x]$ are computed from the buffered frames and included in an array $\text{diff} = (E[\text{diff}_t], E[\text{diff}_x])$. The array $\text{diff}$ is then classified by means of an LBG vector quantizer (Gersho and Gray, 1991) into one of 16 classes, each corresponding to a different packet loss probability which is mapped into the crossing point BER$_{X}$ taking into account the average length of RTP packets. Then the optimization procedure selects the most appropriate coding configuration considering the obtained BER$_{X}$ and the measured BER value provided by the PHY layers.

The classes have been created from an extensive set of training sequences in order to tune the classifier using a significantly heterogeneous amount of data (further details about LBG classification are given in Section 6).

One of the disadvantages of this method is the need of buffering a whole GOP, which implies some delay that could be unsuitable for interactive applications. It is possible to overcome this problem by computing $\text{diff}_t$ and $\text{diff}_x$ on the previously coded data or storing a limited number of frames.

3.2 The $\rho$-based optimization algorithm

The previous section has presented an optimization method that is based on a vector quantization and implies some playout delay at the decoder. In this section, we present an alternative strategy for coding mode selection.

Previous works have shown that for SD-based video coders the distortion produced by the loss of a single packet strongly depends on the characteristics of the coded information. Qu et al. (2006) proposed an Unequal Error Protection (UEP) strategy that increases the number of FEC packets according to the activity level of the original sequence. A possible alternative was suggested by Milani et al. (2005) parameterizing the distortion produced by the loss of one packet as a function of the percentage $\rho$ of null quantized DCT coefficients (called zeros). It is possible to find a similar relation for MDC schemes too.

Given the relative PSNR decrement

\[
\delta\text{PSNR}_{t,i} = \frac{\text{PSNR}_t - \text{PSNR}_{t,i}}{\text{PSNR}_t},
\]

(2)

where PSNR$_t$ is the PSNR value for the uncorrupted coded frame and PSNR$_{t,i}$ is the PSNR value of the concealed frame after a loss, it is possible to linearly relate $\delta\text{PSNR}_{t,i}$ with the percentage $\rho$ associated with the information included in the lost packet. The graphs reported in Fig. 3 show that this characterization is possible both for the SD H.264/AVC coder and for the MDC scheme, but the slope and the intersection of the SD line denote a more significant quality decrement for the single description case. This difference is more evident for sequences with a strong vertical spatial correlation, which permits the error concealment unit of the MD decoder to accurately estimate the lost information. From Figure 3, it is possible to notice that the quality loss of the MDC scheme is much smaller than the quality loss of its SD counterpart for the sequence foreman (see Fig. 3(a)), but in the case of the sequence mobile the adoption of MDC does not permit an improvement of the transmission performance at low $\rho$ values because of the large amount of small details.
The estimated average quality $\text{PSNR}_i^M$ for the configuration $M$ ($M = \text{SD+FEC}$ or MDC) can be computed as

$$\text{PSNR}_i^M = (1 - p_{\text{loss}}^M) \text{PSNR} + p_{\text{loss}}^M \text{PSNR} (1 - k_i^M \rho),$$

(3)

where $p_{\text{loss}}^M$ is the probability of losing a single packet relative to the configuration $M$, and $\text{PSNR}$ and $\rho$ are the average PSNR and $\rho$ values of the previous GOP. The parameter $k_i^M$ is estimated from a set of training sequences and depends on the coding mode $M$. The loss probabilities can be expressed as

$$p_{\text{loss}}^{\text{SD+FEC}} = p_{\text{loss}} - \sum_{c=0}^{n-1} \binom{N_p - 1}{c} p_{\text{loss}}^{c+1} (1 - p_{\text{loss}})^{N_p - 1 - c}$$

$$p_{\text{loss}}^{\text{MDC}} = p_{\text{loss}} (1 - p_{\text{loss}}) + p_{\text{loss}}^2 \simeq p_{\text{loss}} (1 - p_{\text{loss}}),$$

(4)

where $p_{\text{loss}}$ is the packet loss probability estimated from the BER value and the average packet length. The parameter $N_p$ is the total number of video and FEC packets for a single frame in the SD+FEC configuration. Note that for the MDC configuration we assume that, in case a packet is lost, the corresponding packet in the other description has been correctly received since the probability of losing both packets is small.

As it was mentioned in Section 2.1, the number of additional FEC packets can be changed in order to improve the effectiveness of the protection scheme. The presented modelization also permits an efficient modulation of the protection level since for each frame the parameter $n$ can be varied in order to maximize the PSNR estimate $\text{PSNR}_{i,\text{SD+FEC}}$. In addition, previous works have also proposed low-complexity parameterization techniques that appropriately tunes the FEC coder according to the percentage of zeros in the coded frame to be transmitted (Milani et al., 2005). Therefore, despite the aim of the presented cross-layer strategy is to identify the best coding solution between SD+FEC and MDC with predefined configurations, it is possible to improve the transmission performance including effective packet classification algorithms and joint source-channel coding techniques in the final algorithm.

In case the computed $\text{PSNR}_{i,\text{SD+FEC}}$ value is greater than the $\text{PSNR}_{i,\text{MDC}}$ value, the CL control unit assumes that the BER value for the current sequence is higher than the current BER value, and therefore, the SD+FEC configuration is selected. Otherwise, the CL unit employs the MDC scheme.

Note that in this approach no frame buffering is needed since the parametric model is computed from the characteristics of frames that were previously coded.

The performance of the proposed scheme can be improved classifying efficiently the coded video packets and assigning them to the different ACs standardized within the IEEE 802.11e protocol.

4 Classification algorithms for RTP packets

In the previous algorithms, MDC coding is employed in the video transmission whenever the reported BER value is higher than an estimated threshold or the expected distortion related to the SD+FEC scheme is higher than that related to MDC. However, the performance of the MDC and the SD+FEC schemes can be significantly improved by taking into consideration the quality relation reported in Fig. 3 and the possibilities offered by the IEEE802.11e architecture.

4.1 A classification algorithm for MDC RTP packets

In the approach proposed by Bernardini et al. (2007) and in the algorithm described in Section 3, one description is assigned to the AC2 class and the other to the AC1 class, made exception for Intra packets that are assigned to AC3. The assignment is fixed and proves to be effective since it is able to grant the necessary diversity between the loss statistics of each MD stream. In order to improve the performance of MDC schemes Baji (2007) adopted an unbalanced MD-FEC scheme and assigns to the most reliable channel the description with the highest bit rate. These assignment techniques do not take into consideration the changing characteristics of the MD video signal along time. In our research, we have designed and tested a classification strategy that distributes the packets among the different ACs according to the description they belong to and according to the percentage of zeros for the coded signal. More specifically, given the $i$-th frame and the percentage of zeros $\rho_i^{MD1}$ for description MD1 and $\rho_i^{MD2}$ for
description MD2, the corresponding packets are assigned to the AC\textsubscript{i} queue according to Algorithm 1, where $\overline{\rho}_\text{MD1}$ is the average percentage of zeros for description MD1 (computed from the previous frames) and $\overline{\rho}_\text{MD2}$ is the average percentage of zeros for description MD2. In this way, the highest priority AC is reserved for Intra frame packets and Inter packets from description MD1 with a low $\rho$ value, while AC1 and AC2 are used for Inter packets of description MD2 according to the corresponding $\rho$ values. Hence, the highest QoS level is assigned to the packets that contain the most critical information for the decoding process, while the lowest priority is given to the least important data. Note also that the proposed algorithm provides a different average QoS level to the packet streams belonging to different descriptions. In fact, the MDC paradigm proves to be effective whenever it is possible to estimate the lost rows from the other ones. Giving the same priority to both descriptions would reduce the probability of having at least one description correctly received.

### 4.2 A classification algorithm for the SD+FEC option

A procedure similar to Algorithm 1 can also be applied to the packets created by the SD+FEC configuration. The priority of the transmitted data must be inversely proportional to the percentage of null DCT coefficients since the lower is the value of $\rho$ the higher is the relative PSNR loss $\delta$PSNR. Assuming that the $p_{i,j}^{\text{SD}}$ is the $j$-th coded packet for the $i$-th SD frame, the CL control unit assigns to video RTP packet $p_{i,j}^{\text{SD}}$ the class AC\textsubscript{i} according to the rule expressed by Algorithm 2. The parameter $\overline{\rho}_\text{SD}$ is the average percentage of zeros for SD coded frames and $\sigma_\rho^\text{SD}$ the corresponding standard deviation. In this way, those packets whose loss does not significantly affect the quality of the sequence reconstructed at the decoder are assigned to the classes with the lowest priorities. As a consequence, they can be easily discarded by the system in case the congestion level of the network increases too much.

### In the following, we will show how these classification strategies, together with the mode switching algorithm of Section 3, can significantly improve the quality of the reconstructed sequence.

### 5 A hybrid SD+FEC/MDC rate control algorithm

The algorithm described in the previous subsections selects the most appropriate GOP coding setting and packet classification for the coded video sequence. However, the proposed CL scheme needs to control the produced bit according to given bandwidth constraints.

In case the SD+FEC option has been selected, the CL control unit assigns to the $i$-th frame of the GOP a target number of bits $T_i$ according to the equation

$$T_i = \begin{cases} K_{I,P} \frac{G}{N_I} + N_P & \text{for an Intra frame} \\ K_{I,P} \frac{G}{N_P} & \text{for an Inter frame} \end{cases}$$

where $N_I$, $t = I, P$, is the number of $t$-type frames in the GOP that are still to be coded, $K_{I,P}$ is a complexity ratio between I-type frames and P-type frames (see Milani et al. (2008) for more details). The parameter $G$ denotes the number of bits available in the current GOP, which is updated at the beginning of each GOP via the equation

$$G \leftarrow G + \frac{R_b N}{F_r}$$

where $R_b$ is the target bit rate, $N$ is the number of frames in the GOP, and $F_r$ is the frame rate. After coding each frame or field, the number of available bits $G$ is updated.

In a second step, the CL unit partitions $T_i$ into $T_i^S = T_i/(1+r)$ and $T_i^C = r T_i^S$, where $r$ is the coding rate for the FEC coder. The target value $T_i^S$ is then associated to a Quantization Parameter (QP) value adopting the algorithm by Milani et al. (2008) (omitted here for the sake of
frames of description MD  
In the same way, it is possible to update the bit to control the effects of delay jitters on the reconstructed  
duce the probability of buffer underflow at the decoder and  
solution is chosen. The average values  
into the number of columns  
for the current MD frames, the number  
was obtained for the sake of conciseness and  
worker we omit these details for the sake of conciseness, while the channel coding rate  
configuration, an analogous rate allocation algorithm is employed keeping separate frame counters, complexity parameters and ratios for each description. However, in this case, the whole target bit number is assigned to the video source coder since no redundant FEC packets are added. For the MDC configuration, the rate allocation algorithm assigns to the  
it is possible to update the bit budget  
for the average transmission rate  
resulted by varying changing the Signal-to-Noise Ratio (SNR)  
and control the effects of delay jitters on the reconstructed sequence. In the same way, it is possible to update the bit budget for the current GOP in case the average transmission rate  
stated to be coded,  
where the , , is the number of  
mentioned above  
results reported by Milani et al. (2008) permit inferring that the proposed strategy works well in these critic conditions too. However, in the present work we omit these details for the sake of conciseness and we let the reader refer to the cited papers for any additional information. 
Given , the CL unit computes a target QP value for the  
frame of the MDC description in an analogous way with respect to the SD approach. After coding the current MD frames, the number  
the CL scheme starts coding an SD+FEC coded GOP after an MDC coded GOP, the values of rate control parameters are initially computed scaling the parameters of the previous GOP in an appropriate way. 
After coding the current frame, the generated RTP packets are classified using either Algorithm 1 in case the MDC option is chosen or Algorithm 2 in case the SD+FEC solution is chosen. The average values  
then updated using  
variables, equations (5) and (7) can be modified according to  
for the SD+FEC option and by equations  
for the MDC option. As for the parameters  
the optimization algorithm performs a linear regression considering the percentages of zeros, the PSNR values, and the energies of the residual signals for the previously coded frames.

6 Experimental results

In order to evaluate the performance of the presented CL algorithms, we adopted the same experimental setting of Bernardini et al. (2007). The video sequence is transmitted by a video RTP server in an IEEE 802.11e network implemented using the Omnet++ simulator (see Fig. 4). The network is made of an access point that streams the coded video sequences to two terminal nodes. The BER is varied changing the Signal-to-Noise Ratio (SNR) at the end terminal in order to increase the percentage of packets lost in each run. In the simulation setting, we considered a reference value for the SNR of 12 dB (at 50 m from the transmitting node with  
and in Figures 6, 7, and 8 we report the average PSNR value of the reconstructed sequence against the difference between the reference value and the average current one on the x-axis. The quality of the reconstructed figures is measured using the Peak Signal-to-Noise Ratio (PSNR) between the received sequence and the original one acquired by the camera; however, other additional quality metrics have been
we also adopted the sequences tempete, football
we used the sequences tion algorithm, in the training phase of the LBG classifier
formance (Loke et al., 2006). As for the CL optimiza-
be used to obtain a more complete evaluation of the per-
and the video quality metric VQM (Xiao, 2000) which will
the Structure Similarity Index SSIM (W ang et al., 2004)
to the human perception. More precisely we have adopted
computed in order to provide a quality evaluation closer
to the human perception. More precisely we have adopted the Structure Similarity Index SSIM (W ang et al., 2004) and the video quality metric VQM (Xiao, 2000) which will be used to obtain a more complete evaluation of the performance (Loke et al., 2006). As for the CL optimization algorithm, in the training phase of the LBG classifier we used the sequences foreman, news, bus, mobile, city, football, table, soccer, paris, while in the test phase we also adopted the sequences tempete and crew in addition. We simulated the transmission of 120 frames coded in GOPs of 15 frames with IPP...P structure. The network setting is reported in Fig. 4. Experimental results show that the LBG-based algorithm performs very well on both training and test sequences. The experimental results reported in Fig. 5 show that the optimization algorithm is able to obtain the best PSNR value between the MDC and the SD+FEC configurations. Moreover, the PSNR value is about 0.5±0.6 dB better than the higher PSNR between MDC and SD+FEC for BER values close to the crossing point BER\(_x\) (see Fig. 5(a)). In this case, the optimization algorithm is able to switch from one configuration to the other according to the characteristics of the coded video sequence. The performances of the LBG-based CL algorithm and of the \(\rho\)-based algorithm are quite similar since the differences between the two approaches (see Fig. 5(b) and 5(a)) are negligible. However, the average PSNR value for the \(\rho\)-based approach can be significantly improved by a more accurate estimation of the model parameters (such as \(k_l^M\)). The experimental results reported in Fig. 5 are obtained using a fixed \(k_l^M\) obtained from a set of training sequences, and sometimes the adopted model parameters do not accurately suit the characteristics of the transmitted video. As an example, the performance of the \(\rho\)-based algorithm slightly decreases for the coastguard and tempete sequences (see Fig. 5(c) at BER=10\(^{-4}\) and Fig. 5(d) at BER=0.5 10\(^{-4}\)). Better results can be obtained adapting the \(k_l^M\) value according to characteristics of the video signal (e.g. using temporal and vertical correlations). In addition, the parameterization introduced with the \(\rho\)-based model allows the integration of other optimization and rate allocation techniques (see Milani et al. (2005)) which can improve the final performance.

In order to test the packet classification strategies proposed in Section 4, we performed the same testing at the same bit rates using Algorithm 1 for the MDC approach and Algorithm 2 for the SD+FEC approach. The plots in Figures 6 and 7 report the average PSNR vs. SNR loss from a set of 10 trials per point for the sequences crew and tempete. It is possible to notice that the Algorithm 1 makes possible to improve the PSNR value for sequences coded at different quality levels whenever the SNR value increases. For the sequence crew, the average PSNR is increased by 1 dB with respect to the fixed approach of Bernardini et al. (2007) when the SNR loss is \(-6\) dB, and it is possible to notice a similar improvement for other se-

Figure 5: Comparison between the performance of the schemes SD+FEC, MDC and the CL optimization algorithms LBG-based and \(\rho\)-based. The plots report the PSNR values vs. BER obtained from an average of 10 trials per point. The speed of mobile nodes is 5 m/s and the transmission power \(P_\text{Tx}\) is 95 mW.
sequences too (see the results for tempete we have a 0.5 dB PSNR gain with SNR = −6 dB). This improvement is possible since those packets that can be easily recovered by the MDC error concealment algorithm are classified with the lowest priorities allowing the system to discard them in case of congestions in the network. On the other hand, the classification performed by Algorithm 2 does not differentiate significantly from the fixed classification approach since the use of additional FEC packets permits recovering most of the lost information. However, the proposed solution obtains a slight gain for the sequence tempete (0.5 dB with SNR loss −6 dB) since the low temporal correlation makes the sequence more sensitive to information losses. Therefore, a classification strategy that discards packets according to their influence on the final quality of the reconstructed sequence proves to be an effective solution.

In the end, Fig. 8 reports some simulation results for a common approach that combines the CL ρ-based mode switching strategy with the packet classification approaches of Algorithms 1 and 2. It is possible to notice that the classification strategy permits improving all the quality metrics with respect to the fixed approaches. As an example, the PSNR value at −6 dB of SNR loss is 1.5 dB higher for the CL algorithm with respect to the best PSNR value between MDC and SD+FEC approaches (see Fig. 8(a)). The quality of the reconstructed sequence also proves to be higher considering the other metrics too (see Figures 8(b) and 8(c)). As for the sequence tempete, a similar gain can be obtained (see Fig. 8(d), 8(e), and 8(f)).

7 Conclusions

In this paper we propose a cross-layer control algorithm for a hybrid architecture that combines a classical MDC scheme with an SD coder where some redundant packets are added in order to protect the data stream (SD+FEC). In the optimization strategy the characteristics of the image are extracted and each GOP is adaptively coded using either the MDC option or the SD+FEC one. Two strategies are adopted to select the best mode. The first one is based on an LBG classification of the frames to be coded, while the second one relies on an accurate parametric modelization of the distortion. Both approaches prove to be quite competitive since the resulting PSNR values are always equal or higher (up to 0.6 dB for foreman) than the values provided by the best mode between MDC and SD+FEC. The performance of these schemes is enhanced by a cross-layer packet classification algorithm that relies on assigning the highest-priority service classes to the packets of frames with a low percentage of null quantized transform coefficients, while the others are transmit-
Figure 8: Performance of different CL optimization strategies for the sequence crew and tempete. The plots report the average PSNR value for the reconstructed sequence from 10 trials vs. the average SNR decrement. The mobile speed is 0 m/s and the transmitting power is 95 mW.

Experimental results show that is possible to improve the PSNR value up to 1.5 dB with respect to a fixed approach where the source coding choice and the packet classification strategy do not adapt to the characteristics of the coded sequence.

REFERENCES


