CROSS-LAYER OPTIMIZATION STRATEGIES FOR VIDEO TRANSMISSION OVER IEEE 802.11E NETWORKS

(SESSION: SIGNAL PROCESSING)

Simone Milani

Dept. of Information Engineering,
University of Padova, Italy
simone.milani@dei.unipd.it

Pamela Zontone

DIEGM,
University of Udine, Italy
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 performance of the MDC scheme is then improved by a packet classification strategy based on their significance in the multiple description decoding scheme. Experimental results show that both cross-layer optimization algorithms perform well with a small computational effort but different playout delays. As for the MDC packet classification, the PSNR value can be increased up to 2 dB depending on the characteristics of the coded video sequence.

***Index Terms*—** Multiple Descriptions, cross-packet codes, packet classification, cross-layer optimization, IEEE 802.11e.

1. INTRODUCTION

Recent years have witnessed a rapid increment in video applications over wireless networks including on-demand video streaming and videophoning. This growth has increased the research efforts of both academy and industry towards more and more effective robust video transmission schemes. Following this trend, designers have endowed the modules at different protocol layers with new tools and strategies that can be efficiently used to improve the quality of the reconstructed sequence. At the application layer, new robust video coding paradigms, like Multiple Description Coding (MDC) [1], have been designed in order to suit the characteristics of the network over which the video data are transmitted. 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 [2]. 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 (see [3] as an example). However, the quality of the sequence reconstructed at the decoder depends on a clever orchestration of all these elements that permits adapting each tool according to the characteristics of the coded video sequence and the network conditions (see [4]).

To this purpose, a wide variety of cross-layer optimization algorithms have been proposed in the literature involving different layers and techniques. As an example, the algorithms proposed in [5, 6] use Unequal Error Protection (UEP) for the transmitted packets by adapting the FEC coder at transport layer to the characteristics of the coded video signal. In [7] Ksentini *et al.* present 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 strategy is compared in [8] with a cross-layer MDC scheme that assigns a different class of service to each description.

In this paper, we propose some cross-layer (CL) optimization algorithms that were 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 [9]. In particular, at first we present two CL strategies that adaptively combine the MDC scheme in [8] 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 or between the source and the FEC channel coder according to the character-

This work was carried on within the PRIN research project prot. 2005099247 with the support of the Italian Ministry of University and Research (MiUR).

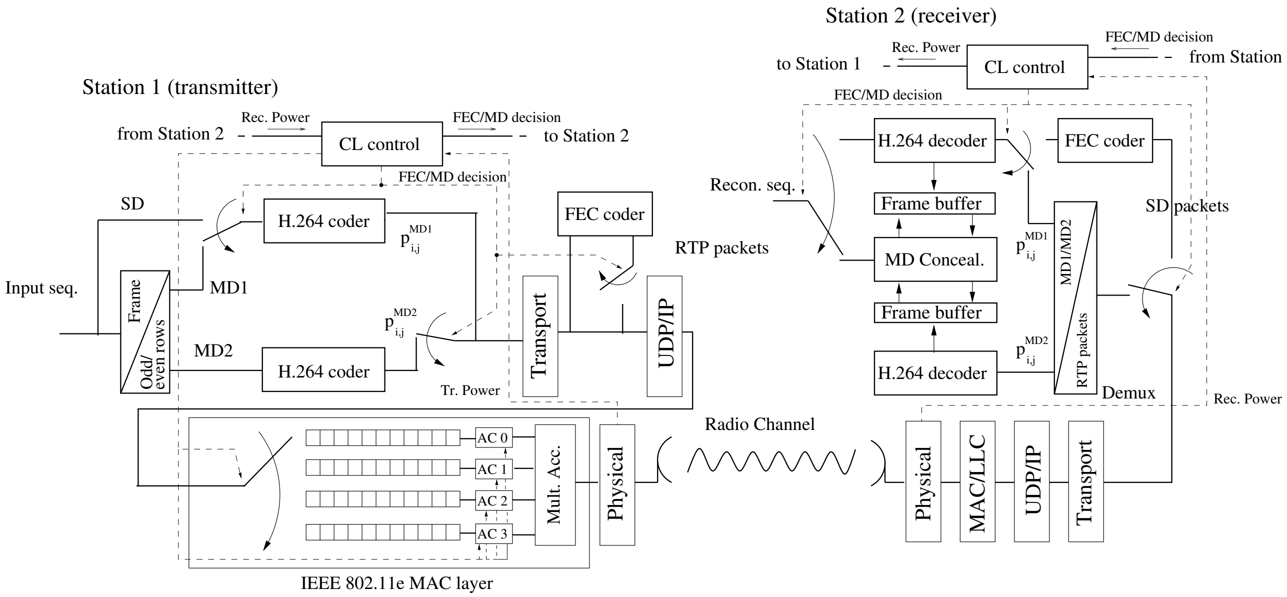


Fig. 1. The adopted Multiple Description scheme.

istics 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 [10], 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 MDC 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 (presented in Section 3.1 and Section 3.2 respectively). Section 3.3 reports the rate control algorithm adopted to combine the SD+FEC and MDC approaches, while Section 4 reports a classification algorithm for MDC RTP packets that significantly improves the performance of the MDC scheme. The experimental results reported in Section 5 show that the proposed algorithms are able to improve the schemes [2, 8] with a negligible computational effort. Conclusions are drawn in Section 6.

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 [10]. 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 [2]. 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.

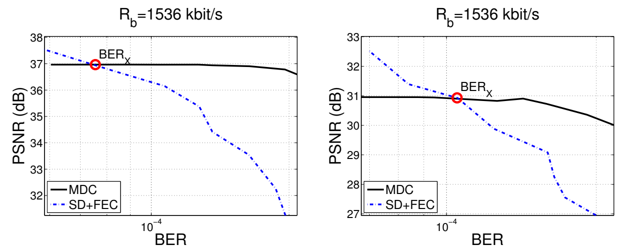
To achieve this, the proposed MDC scheme takes advantage of the possibilities offered by the IEEE 802.11e protocol, which will be described in the following subsection.

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 $[0, \mathrm{CW}]$, where the Contention Window parameter CW is initially set to $\mathrm{CW}_{\min}$ and is doubled every time the transmission fails (up to $\mathrm{CW}_{\max}$). 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 802.11e [3] standard 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 $[\mathrm{CW}_{\min}, \mathrm{CW}_{\max}, \mathrm{RL}]$ 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 [8] for more details). In case the MDC configuration is adopted, Inter coded packets are



(a) `crew` CIF (b) `tempete` CIF

Fig. 2. PSNR vs. BER for the SD+FEC and MDC configuration. The graphs also show the crossing point BER_X which varies according to the transmitted sequence (coded at the same bit rate $R_b = 1536$ kbit/s).

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.

3. CROSS-LAYER OPTIMIZATION OF FEC CHANNEL CODER AND MDC SCHEME

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 the crossing point BER_X, where the curves of PSNR vs. BER for SD+FEC and MDC intersect. For BER values higher than BER_X, the MDC option gives a better quality at the same coding rate, while the SD+FEC configuration is chosen whenever the estimated BER is lower than BER_X. The value of BER_X depends on the characteristics of the coded sequence (see Fig. 2) and on the number of FEC packets added by the SD+FEC scheme. The following subsections will provide further details about the adopted estimation methods.

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.

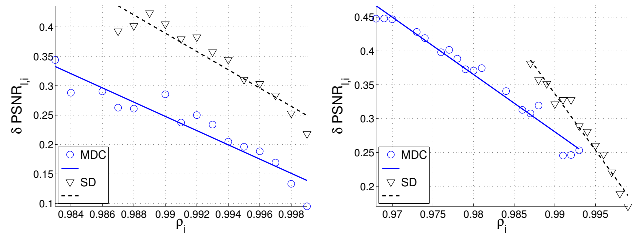
As a consequence, we need a metric to measure the level of spatial and temporal correlation. In the literature, a widely-used parameter to characterize the residual signal energy after a prediction is the Sum of Absolute Differences (SAD), which characterizes the correlation between the original signal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

$$
\begin{aligned}
\mathrm{diff}_t &= \frac{1}{N_x\ N_y} \sum_{x,y=0}^{N_x-1,N_y-1} |I_i(x,y) - I_{i-1}(x,y)| \\
\mathrm{diff}_x &= \frac{2}{N_x\ N_y} \sum_{x,y=0}^{N_x/2-1,N_y-1} |I_i(2\,x,y) - I_i(2\,x+1,y)|\,,
\end{aligned}
\tag{1}
$$

where N_x, N_y are the number of rows and columns for each frame respectively, and $I_i(x,y)$ is the luminance pixel value of the i-th frame at position (x,y).

Before coding, a whole GOP is stored in memory, and the average values $E[\mathrm{diff}_t]$ and $E[\mathrm{diff}_x]$ are computed from the buffered frames and included in an array $\underline{\mathrm{diff}} = (E[\mathrm{diff}_t],\ E[\mathrm{diff}_x])$. The array $\underline{\mathrm{diff}}$ is then classified by means of an LBG vector quantizer [11] 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 5).



(a) `foreman` CIF (b) `mobile` CIF

Fig. 3. $\delta\mathrm{PSNR}_{l,i}$ vs. ρ_i for SD and MDC.

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 diff_t and diff_x on the previously coded data or storing a limited number of frames.

3.2. The ρ-based optimization algorithm

The previous section has presented an optimization method that is based on a vector quantization and implies some play-out 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. In [5] Qu *et al.* propose an Unequal Error Protection (UEP) strategy that increases the number of FEC packets according to the activity level of the original sequence. In [6], the authors parameterize the distortion produced by the loss of one packet as a function of the percentage ρ of null quantized DCT coefficients (called *zeros*). It is possible to find a similar relation for MDC schemes too.

Given the relative PSNR decrement

$$
\delta\mathrm{PSNR}_{l,i} = \frac{\mathrm{PSNR}_i - \mathrm{PSNR}_{l,i}}{\mathrm{PSNR}_i}, \tag{2}
$$

where PSNR_i is the PSNR value for the uncorrupted coded frame and $\mathrm{PSNR}_{l,i}$ is the PSNR value of the concealed frame after a loss, it is possible to linearly relate $\delta\mathrm{PSNR}_{l,i}$ with the percentage ρ 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 mainly evident for sequences with a strong vertical spatial correlation, which allows 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 lower 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 ρ values because of the large amount of small details.

The estimated average quality $\overline{\text{PSNR}}_l^M$ for the configuration M (M = SD+FEC or MDC) can be computed as

$$\overline{\text{PSNR}}_l^M = (1 - p_{\text{loss}}^M)\ \overline{\text{PSNR}} + p_{\text{loss}}^M\ \overline{\text{PSNR}}\left(1 - k_l^M\ \overline{\rho}\right), \quad (3)$$

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

$$\begin{aligned} p_{\text{loss}}^{\text{SD+FEC}} &= p_{\text{loss}} - \sum_{c=0}^{n-1} \binom{N_p - 1}{c} p_{\text{loss}}^{c+1} (1 - p_{\text{loss}})^{N_p - 1 - c} \\ p_{\text{loss}}^{\text{MDC}} &= p_{\text{loss}}(1 - p_{\text{loss}}) + p_{\text{loss}}^2 \ \simeq\ p_{\text{loss}}(1 - p_{\text{loss}}), \end{aligned} \quad (4)$$

where p_{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 $\overline{\text{PSNR}}_l^{\text{SD+FEC}}$. 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 [6]. 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 (see [6]).

In case the computed $\overline{\text{PSNR}}_l^{\text{SD+FEC}}$ value is greater than the $\overline{\text{PSNR}}_l^{\text{MDC}}$ value, the CL control unit assumes that the BER_X point 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.

3.3. A hybrid SD+FEC/MDC rate control algorithm

After the algorithm described in the previous subsection has selected the more appropriate GOP coding setting, the proposed CL scheme needs to implement an accurate rate control algorithm that maximizes the quality of the reconstructed sequence under certain 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

$$T_i = \begin{cases} K_{I,P}\ \dfrac{G}{K_{I,P}\ N_I +\ N_P} & \text{for an Intra frame} \\ \dfrac{G}{N_P} & \text{for an Inter frame} \end{cases} \quad (5)$$

where the N_t, $t = I, P$, 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 [12] 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

$$G \leftarrow G + \frac{R_b\ N}{F_r} \quad (6)$$

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 reported in [12] (omitted here for the sake of conciseness), while the channel coding rate r is converted into the number of columns n for the FEC coder.

In case the CL control unit chooses to adopt the MDC 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 i-th frame of description MDd ($d = 1, 2$) the target number of bits $T_i^{\text{MD}d}$ according to the equation

$$T_i^{\text{MD}d} = \begin{cases} \dfrac{K_{I,P}^{\text{MD}d}\ G}{K_{I,P}^{\text{MD1}}\ N_I^{\text{MD1}} + K_{I,P}^{\text{MD2}}\ N_I^{\text{MD2}} + K_{P,P}^{\text{MD1}}\ N_P^{\text{MD1}} + N_P^{\text{MD2}}} & \text{for Intra frames} \\ \dfrac{K_{P,P}^{\text{MD}d}\ G}{K_{P,P}^{\text{MD1}}\ N_P^{\text{MD1}} + N_P^{\text{MD2}}} & \text{for Inter frames,} \end{cases} \quad (7)$$

where the $N_t^{\text{MD}d}$, $t = I, P$ and $d = 1, 2$, is the number of t-type frames of description MDd in the GOP that are still to be coded, $K_{t,P}^{\text{MD}d}$ is a complexity ratio between t-type frames of description MDd and P-type frames of description MD2 ($K_{P,P}^{\text{MD2}} = 1$).

The target number of allocated bits can be corrected considering the buffer level and possible transmission bit rate variations. As the paper [12] reports, equations (5) and (7) can be modified according to the buffer level and the desired playout delay in order to reduce the probability of buffer underflow at the decoder and to control the effects of delay jitters on the reconstructed sequence. In the same way, it is possible to update the bit budget G for the current GOP in case the average transmission rate R_b varies. Results reported in [12] permits 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 $T_i^{\mathrm{MD}d}$, the CL unit computes a target QP value for the i-th frame of the MDd description in an analogous way with respect to the SD approach. After coding the current MD frames, the number G of remaining bits in the GOP is updated, and the following frames are processed.

When the optimization strategy switches from a MDC coded GOP to a SD+FEC coded GOP, the starting values of rate control parameters are estimated averaging the values obtained from the previous GOP. On the other hand, whenever the CL scheme starts coding a SD+FEC coded GOP after a MDC coded GOP, the values of rate control parameters are initially computed scaling the parameters of the previous GOP in an appropriate way.

4. A CLASSIFICATION ALGORITHM FOR MDC 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 scheme can be significantly improved by taking into consideration the quality relation reported in Fig. 3 and the possibilities offered by the IEEE802.11e architecture.

In [8] 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 [13] the author adopts 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 ρ_i^{MD1} for description MD1 and ρ_i^{MD2} for description MD2, the corresponding packets are assigned to the AC$_i$ queue according to Algorithm 1, where $\overline{\rho}^{\mathrm{MD1}}$ is the average percentage of zeros for description

Algorithm 1 Packet classification using ρ

Naming $p_{i,j}^{\mathrm{MD1}}$ the j-th packet of MD1 for the i-th frame
and $p_{i,j}^{\mathrm{MD2}}$ the j-th packet of MD2 for the i-th frame
if the i-th frame is Intra coded **then**
    $\mathrm{AC}_3 \leftarrow p_{i,j}^{\mathrm{MD1}}$ $\quad \mathrm{AC}_3 \leftarrow p_{i,j}^{\mathrm{MD2}}$
else
    if $\rho_i^{\mathrm{MD1}} < \overline{\rho}^{\mathrm{MD1}}$ **then**
        $\mathrm{AC}_3 \leftarrow p_{i,j}^{\mathrm{MD1}}$
    else
        $\mathrm{AC}_2 \leftarrow p_{i,j}^{\mathrm{MD1}}$
    end if
    if $\rho_i^{\mathrm{MD2}} < \overline{\rho}^{\mathrm{MD2}}$ **then**
        $\mathrm{AC}_2 \leftarrow p_{i,j}^{\mathrm{MD2}}$
    else
        $\mathrm{AC}_1 \leftarrow p_{i,j}^{\mathrm{MD2}}$
    end if
end if

Algorithm 2 Description switching

if $\overline{\rho}^{\mathrm{MD2}} < \overline{\rho}^{\mathrm{MD1}}$ **then**
    MD2 becomes the highest priority description
else
    MD1 remains the highest priority description
end if

MD1 (computed from the previous frames) and $\overline{\rho}^{\mathrm{MD2}}$ 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 ρ value, while AC1 and AC2 are used for Inter packets according to the corresponding ρ 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. In a second optimization procedure, we considered the possibility of switching description MD1 with description MD2 whenever the recovering performance of MD2 is greater than that of MD1, i.e., estimating MD1 from MD2 leads to a better quality in the reconstructed sequence. To this purpose, at the beginning of each GOP, MD1 and MD2 are switched according to the criterion reported in Algorithm 2.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the presented CL algorithms, we adopted the same experimental setting of [8]. The video sequence is transmitted by a video RTP server

Parameter	Value
AC0 ($\mathrm{CW_{min}}$, $\mathrm{CW_{max}}$, RL)	$(31, 1023, 4)$
AC1 ($\mathrm{CW_{min}}$, $\mathrm{CW_{max}}$, RL)	$(31, 1023, 4)$
AC2 ($\mathrm{CW_{min}}$, $\mathrm{CW_{max}}$, RL)	$(15, 31, 8)$
AC3 ($\mathrm{CW_{min}}$, $\mathrm{CW_{max}}$, RL)	$(7, 15, 8)$
Bit rate	11 Mbit/s
Mobile speed	5 m/s
Path loss α	[2.0, 2.3]
Transmission power	75 mW

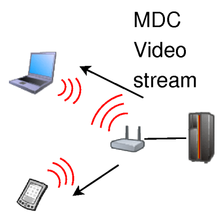
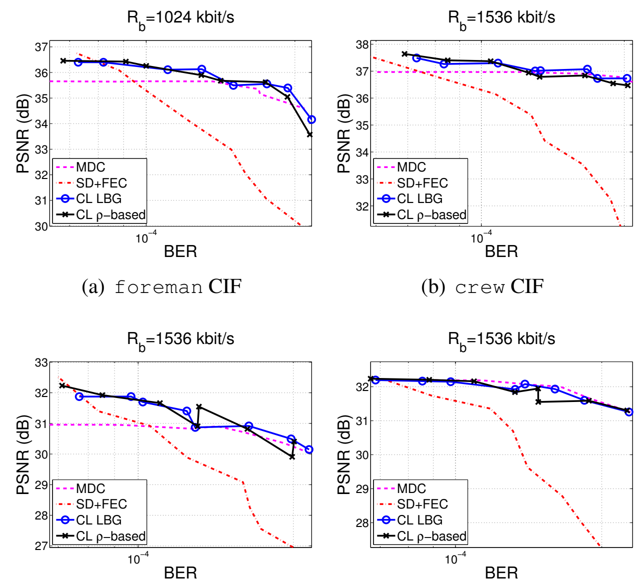


Fig. 4. The adopted network scenario.

in an IEEE 802.11e network implemented using the Omnet++ simulator (see Fig. 4). 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 graphs 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 $0.5 \div 0.6$ 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 performance of the LBG-based CL algorithm and the ρ-based algorithm are quite similar since the differences between the two approaches (see Fig. 5(a) and Fig. 5(b)) are negligible. However, the parameterization introduced with the ρ-based model permits the integration of other optimization and rate allocation techniques (see [6]) which can improve the final performance.

In order to test the packet classification strategy proposed in Section 4, we simulated the transmission of various sequences (coded with fixed QP $\in \{18, 24, 30\}$, GOP IPP...P of 15 frames) varying the channel propagation condition. The plots in Figure 6 report the average PSNR vs. the BER value from a set of 10 trials per point. 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 BER value increases. For the sequence `foreman`, coded with QP $= 24$, the average PSNR is increased by 2 dB with respect to the fixed approach of [8] when the BER is $9 \cdot 10^{-4}$, and it is possible to notice a similar improvement for other sequences too (see the results with BER=$8 \cdot 10^{-4}$ for `table` in Fig. 6(d) and for `paris` in Fig. 6(e)). The use of Algorithm 2 slightly improves the quality of the reconstructed video sequence, as it is shown in Fig. 6(d) for BER=10^{-3}. However, this increment is strictly dependent on the characteristics of the video sequence. For the sequence `news`, coded with QP $= 30$ (see Fig. 6(f)) no significant differences can be appreciated between Alg. 1 and Alg. 1+Alg. 2. This fact is mainly due to the high correlation existing between even and odd lines of each video frame, which makes the two descriptions equivalent.



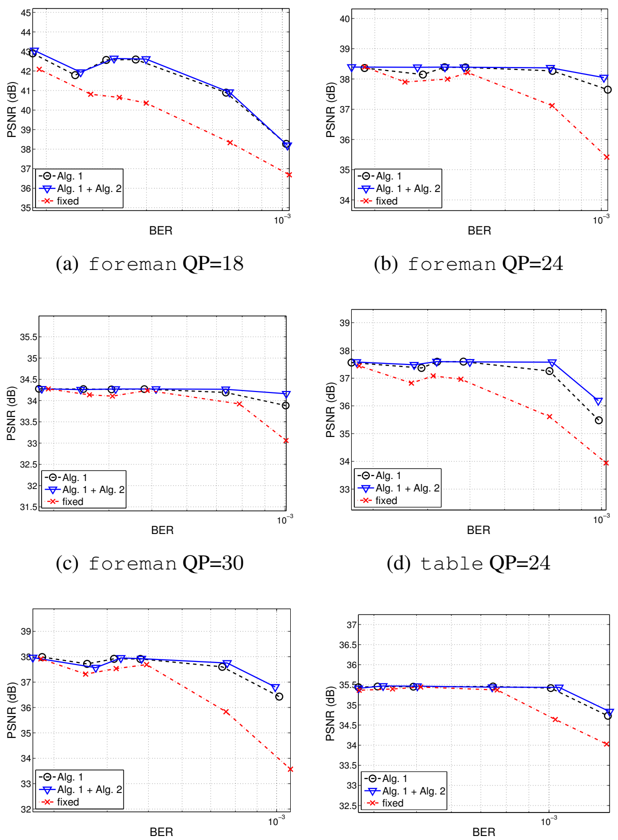
(a) `foreman` CIF (b) `crew` CIF

(c) `tempete` CIF (d) `coastguard` CIF

Fig. 5. Comparison between the performance of the schemes SD+FEC, MDC and the CL optimization algorithms LBG-based and ρ-based. The plots report the PSNR values vs. BER obtained from an average of 10 trials per point.

6. CONCLUSIONS

In this paper we propose a cross-layer control algorithm for a hybrid architecture that combines a classical MDC scheme with a 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 value is always equal or higher (up to 0.6 dB for `foreman`) than the value provided by the best mode between MDC and SD+FEC. The performance of the MDC scheme 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 transmitted with



(a) `foreman` QP=18 (b) `foreman` QP=24

(c) `foreman` QP=30 (d) `table` QP=24

(e) `paris` QP=24 (f) `news` QP=30

Fig. 6. Comparison of different classification algorithm for the CL schemes. The plots report the average PSNR value for the reconstructed sequence from 10 trials vs. the BER value.

a lower QoS level since the quality of the reconstructed sequence is not significantly affected by their loss. Experimental results show that is possible to improve the PSNR value up to 2 dB with respect to a similar approach where packet classification does not take into account the sequence characteristics.

7. REFERENCES

[1] Vivek K. Goyal, "Multiple Description Coding: Compression Meets The Network," *IEEE Signal Processing Mag.*, vol. 8, no. 5, pp. 74–93, Sept. 2001.

[2] J. Rosenberg and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction (RFC2733)," *Internet Draft, Network Working Group*, Dec. 1999.

[3] IEEE 802.11/D13.0, Part 11, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Medium access control (MAC) enhancements for Quality of Service (QoS)," Jan. 2005.

[4] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission," *Proc. IEEE*, vol. 93, no. 1, pp. 135–147, Jan. 2005.

[5] Q. Qu, Y. Pei, and W. Modestino, "An Adaptive Motion-Based Unequal Error Protection Approach for Real-Time Video Transport Over Wireless IP Networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 1033–1044, Oct. 2006.

[6] S. Milani, G. A. Mian, and L. Celetto, "Joint Optimization of Source-Channel Video Coding Using the H.264 Encoder and FEC Codes," in *Proc. of the 13th European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, Sept. 2005.

[7] 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 Commun. Mag.*, vol. 44, no. 1, pp. 107–114, Jan. 2006.

[8] R. Bernardini, M. Durigon, R. Rinaldo, P. Zontone, and A. Vitali, "Real-Time Multiple Description Video Streaming Over QoS-Based Wireless Networks," in *Proc. of International Conference on Image Processing (ICIP 2007)*, San Antonio, TX, USA, Sept. 2007.

[9] PRIN 2005 Research Project, "Reliable Multimedia Transmission over non-Reliable Networks: Advanced Source/Channel Coding Techniques," web site, 2006, http://www.diegm.uniud.it/rinaldo/PRIN/.

[10] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Joint final committee draft (JFCD) of joint video specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)," in *Joint Video Team, 4^{th} Meeting*, Klagenfurt, Germany, July 2002.

[11] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publisher, Norwell, MA, USA, 1991.

[12] S. Milani, L. Celetto, and G.A. Mian, "An Accurate Low-Complexity Rate Control Algorithm Based on (ρ, E_q)-Domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 257–262, Feb. 2008.

[13] I. V. Bajìc, "The Effects of Channel Correlation on the Performance of Some Multiple Description Schemes," in *Proc. of* 10^{th} *Canadian Workshop on Information Theory (CWIT 2007)*, Edmonton, Alberta, Canada, June6–8 2007.