P2P SCALABLE VIDEO STREAMING USING DATA PRIORITY AND FEC-BASED NONCOOPERATIVE MULTIPLE DESCRIPTION

Marco Gaggio$^\flat$*, Simone Milani*$^\flat$*, Livio Lima*$^\sharp$*, Pierangelo Migliorati*$^\sharp$*, Giancarlo Calvagno*$^\flat$

$\flat$ University of Padova, Dept. of Information Engineering, Padova, Italy
$\sharp$University of Brescia, Dept. of Information Engineering, Brescia, Italy
e-mail: {gaggioma,simone.milani}@dei.unipd.it, livio.lima@ing.unibs.it
calvagno@dei.unipd.it, pierangelo.migliorati@ing.unibs.it

ABSTRACT

Recent years have assisted the proliferation of peer-to-peer (P2P) communications in streaming audio/visual contents over widely-distributed networks. This widespreading has also fostered the need for efficient video coding and transmission strategies. The paper present an MD coding and packet classification strategy that improves a traditional Multiple Description scheme basde on a scalable video coder and FEC codes via effective Unequal Error Protection and QoS classification. The proposed strategy uses a low-complexity packet prioritization strategy based on dependencies between video coding units to estimate the distortion associated to their loss. These data will be then used to tune the protection level of video packets and optimize the classification via Game Theory based approach. The paper compares the performance of the different configurations of the proposed approach with some state-of-the-art solutions (like srTCM) showing a quality improvement up to 6 dB.

Index Terms— P2P video, multiple description, SVC, game theory, classification, cross-layer.

1. INTRODUCTION

Recent years have assisted the proliferation of peer-to-peer (P2P) communications in streaming audio/visual contents over widely-distributed networks. The flexibility and versatility of P2P protocols permit optimizing the delivering strategies making the different nodes download and upload contents of common interest. The possibilities offered by this kind of communication facilities have, therefore, fostered a significant amount of research in order to adapt and optimize traditional video coding and transmission strategies to these new protocols.

A first set of solutions considered the possibility of coding the transmitted video stream using a scalable video coder (e.g., SVC [1]). Scalable coders are effective in terms of rate-distortion performance and generates different streams with different importance. Unfortunately, these are strictly dependent and need to be decoded in a given order. As a matter of fact, the most important information in the compressed data needs to be repeated in different P2P streams in order to ensure its correct delivery.

Multiple Description (MD) architectures [2] has recently proved an effective coding and packetizing strategy for P2P video communications since the video stream to be transmitted is divided into independently-coded correlated streams (called "descriptions"). Each MD substream can be independently decoded with respect to the others, and the perceived quality of the signal reconstructed at the end terminal only depends on the number of correctly-received substreams (since each one is equally-significant). This coding strategy proves to be extremely effective in P2P networks since the topologies and the capacities of the different connected nodes is time-varying, and therefore, the number of connections can significantly change.

Experimental results also have shown that the effectiveness of such approaches depends on how MD packets are routed and transmitted. [3]). However, in presence of congestions and packet losses, the final performance is significantly affected by the protection and labelling operated on the different packets.

This work aims at combining these three strategies in a common flexible framework that permits providing high quality multimedia experiences to the end users at a reasonable computational cost. The input video stream is coded using the SVC coder. For each of the generated packet, a priority level is assigned by a low-complexity packet prioritization strategy based on dependencies between video coding units to estimate the distortion associated to their loss. These data will be then used to tune their protection level in a FEC-based MD scheme. Then, the different MD packets are assigned to different Quality-of-Service (QoS) classes according to a distributed optimization strategy based on Game Theory (GT) [4]. Different packetization strategies have been considered and tested.

In the following, Section 2 presents the adopted MDC

Thanks to XYZ agency for funding.

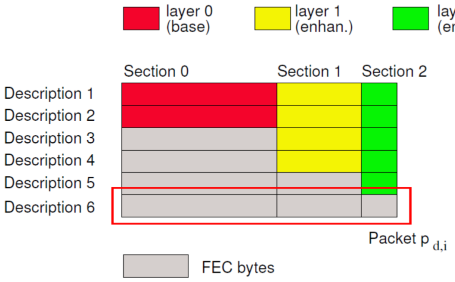


Fig. 1. Example of MDC-FEC method using static error capacity

scheme and Section 3 describes the classification strategy for SVC packets. Section 4 presents how packets are organized in the MD coding matrix and how QoS classes are assigned to the different descriptions. Experimental results (Section 5) and conclusions (Section 6) will conclude the paper.

2. FEC-BASED MULTIPLE DESCRIPTION

The adopted MD scheme uses the information generated by a scalable video coder and encapsulate it into a $R \times C$ matrix. The scalable video coder generates N_L hierarchical bit streams (named *layers*) and encapsulates them into different sets of packets. In scalable coding different packet streams present a strong dependence, and as a matter of fact, the quality of the reconstructed video sequence at the decoder strongly depends on which packet streams it correctly receives. More precisely, the coder generates a *base layer*, which can be independently decoded and provides the minimum quality level for the reconstructed sequence, and additional *enhancement layers* such that each of them refines the coding distortion of the previous layer. One of the latest scalable video coders is defined within the coding standard H.264/SVC [1], which permits generating a layered bit stream that can be scaled in time, resolution, and coding distortion.

Puri and Ramchandran [5] have shown that it is possible to generate an MDC stream using a scalable video coder and FEC channel codes. Data are included in a $R \times C$ matrix along its rows and additional FEC bytes are generated along the columns and encapsulated into packets. More precisely, the columns of the coding matrix are grouped into ordered sections, and the layer L of the scalable video stream ($L = 0, ..., N_L - 1$) is included into the L-th section of the matrix, as shown in Fig.1, given the coding parameters $R_{L,S}$ and $R_{L,C}$. As a result, the video data fill the first $R_{L,S}$ rows of the matrix, and additional bytes are generated in the remaining $R_{L,C} = R - R_{L,S}$ rows of section L using the FEC code (e.g., for layer 0 in Fig. 1 $R_{L,S} = 2$ and $R_{L,C} = 4$).

Then, the MDC system includes the data of each row into a separate packet that constitutes an independent description. At the decoder, the more descriptions are received, the more enhancement layer can be correctly recovered and the better the quality of the decoded sequence is.

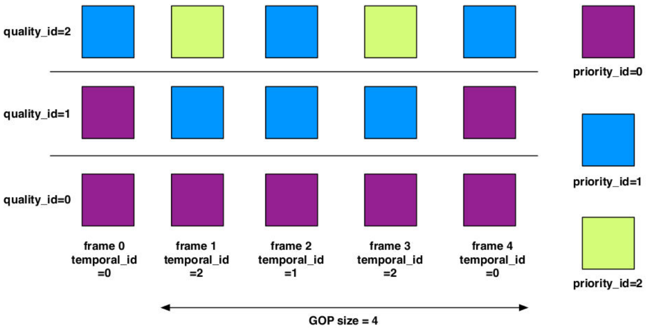


Fig. 2. Example of priority assignment in a GOP with 4 frames. Note that for *priority_id* is assigned according to *temporal_id* and *quality_id*.

The example of Fig.1 shows a 6 descriptions MDC where the base layer can be recovered whenever two descriptions (packets) are received, layer 1 can be obtained with 4 packets, and layer 2 requires 5 descriptions to be recovered.

The width of the matrix (and as a consequence, the width of each section) is determined by the size of the packets, the values $R_{L,S}$, and the MTU size (which is set to 1000).

3. PACKET CLASSIFICATION

Scalable extension of H.264/SVC standard [1] offers native mechanisms for packet classification. Video sequences is processed in Group of Pictures (GOP) with typical size of 8 or 16 pictures for GOP. Temporal scalability is achieved by hierarchical B-frames decomposition within each GOP. Each frame (Intra, B-type and P-type) is than encoded in order to provide quality scalability, where two approaches can be adopted: Coarse Grain Scalability (CGS) or Medium Grain Scalability (MGS). In this work we only consider MGS scalability, with the additional functionality called “MGS vector”, that enables the split of transformed DCT coefficients into multiple data-unit. This approach provides high flexibility in rate adaptation.

Coded video data are organized into Access Unit (AU), where each AU contains the data for a single picture. Within an AU, data are distributed into NALUs, each one identified by the following fields of the NALU header: *dependency_id* for the spatial resolution, *temporal_id* for the temporal level, and *quality_id* for the quality level. NAL unit is the elementary data-unit for transmission or adaptation purposes. These three fields of NALU header provide a first mechanism for SVC content classification. Additionally, SVC offers a native mechanism for NALU classification following a priority map, by using the 6-bits *priority_id* field of the NALU header. An example of the relation between *dependency_id*, *temporal_id*, *quality_id* and *priority_id* is shown in Figure 2. Note that for

I frames *priority_id*$= 0$ is assigned to *quality_id*$= 0, 1$, while for less important frames *priority_id*$= 0$ is assigned to *quality_id*$= 0$ only. It is also possible to see that a higher average priority is given to *temporal_id*$= 1$ with respect to *temporal_id*$= 2$.

Even if in SVC the use of this field is not normative, an optimal generation of priority map in addition to priority-driven adaptation mechanisms could increase the decoding performance when the available rate is lower than that required to transmit the full SVC stream. In this work we adopt the approach proposed in [6], that is briefly explained in the following. Compared to other approaches in literature, this approach combines computational advantages since adopts a distortion model with high flexibility in problem description by the use of Integer Linear Programming.

Let us assume that each NAL unit is described by a binary variable $x_{t,q}$, its rate contribution $r_{t,q}$ and its contribution $c_{t,q}$ over the GOP distortion reduction, estimated using a particular distortion model. Given a maximum available GOP rate R, lower than that of the full GOP stream, we want to identify the set of NAL units that maximize the GOP distortion reduction. Let us call this problem SP(R). The problem is described with the following ILP model:

$$\text{(ILP):} \qquad Z = \max \mathbf{cx}$$
$$\text{subject to} \begin{cases} A\mathbf{x} \geq b & \\ \mathbf{x} \geq 0 & \text{integer} \end{cases} \tag{1}$$

The unknown $\mathbf{x} = \{x_{t,q}\}^T$ is a vector of binary variables, one for each NALU, that indicates if the NALU has to be maintained ($x_{t,q} = 1$) or discarded ($x_{t,q} = 0$). The vector $\mathbf{c} = \{c_{t,q}\}$ represents the distortion contributions. Consequently, the objective function $\mathbf{cx}$ represents the overall GOP distortion reduction:

$$\mathbf{cx} = \sum_{i=0}^{N-1} \sum_{q=0}^{Q} x_{i,q} c_{i,q} \tag{2}$$

where N is the GOP size, while Q is the maximum value of *quality_id*.

A is the constraint matrix. Our problem presents two types of constrains. Within each picture i, SVC standard defines that a NALU with *quality_id*=x can be decoded only if all the NALU with *quality_id*$< x$ are available. This set of constraints can be represented as:

$$x_{i,q} - x_{i,q+1} \geq 0 \qquad \forall i = 0, ..., N-1 \tag{3}$$

Since we defined a maximum available rate R, we have a further rate constrain expressed as:

$$\sum_{s=0}^{S} \sum_{i=0}^{N-1} \sum_{q=0}^{Q} x_{i,q} r_{i,q} \leq R \tag{4}$$

The solution of problem (1) gives a binary map $\mathbf{x}$ that represents the set of NAL units that maximize the GOP distortion reduction for a particular rate R. Now, given the solution for this sub-problem SP(R), with the following algorithm it is possible to generate the priority map for the GOP:

1. estimation of distortion vector $\mathbf{c}$
2. choice of a set of K rates $R_0, ..., R_K$, starting from the rate of SVC base layer (R_0) to the rate of the full SVC stream (R_K)
3. resolution of K subproblems $SP(R_k)$, obtaining K solution vectors $\mathbf{x}_k$
4. let $\mathcal{N}_k$ the set of NALUs with related binary variable equal to 1 in $\mathbf{x}_k$. The *priority_id* value equal to k is assigned to the NALUs that belong to the set $\{\mathcal{N}_k - \mathcal{N}_{k-1}\}$, with $\mathcal{N}_{-1} = \emptyset$

4. STREAMING VIDEO OVER P2P NETWORKS

In this section will be presented several methods of MDC-FEC matrix implementation through a priori evaluation of the importance (in terms of distortion) of each layer (see section 3). Thereafter, different classification methods at the routing nodes are considered as well.

4.1. MDC-FEC matrix

Initially, the video source was encoded in a H.264/SVC stream, composed by N layers each with own priority $p \in P = \{1, 2, 3, ..., np\}$, where np is the number of priorities (see sec. 3). From this set-up, M MDC descriptions need to be created. The matrix constructor routine associates to $p \in P$ a certain error capacity function $f(p)$ to compute the number of redundant FEC symbols ($NFEC$) so that $NFEC(p) = f(p)$. Note that if $p_1 \geq p_2$, $p_1, p_2 \in P$, then $NFEC(p_1) \geq NFEC(p_2)$. With these assumptions the layer with priority p will be split in $M - NFEC(p)$ rows of the associated MDC-FEC matrix, while the remaining $NFEC(p)$ rows will contain FEC bytes. An example of layer partition in the MDC-FEC matrix is presented in work [4].

The composition of the MDC-FEC matrix can be performed according to the following 4 strategies.

- *Priority*. For each GOP, $|P|$ matrices are created. Each matrix contains all the packets with the same priority p. Furthermore, each layer is protected with $NFEC(p)$ redundancy symbols.
- *Layer*. For each GOP, N matrices are created. In the same matrix all packets of the same layer were collected (i.e. the matrix M_0 contains all the base layer

packets of the GOP, the matrix M_1 contains all enhancement layers 1 and so on). Furthermore, each layer was protected with $NFEC(p)$.

- *Frame*. For each GOP, $GOPSize$ matrices have been created (the $GOPSize$ variable indicates the number of frames in each GOP). In the same matrix all layers of the single frame were distributed. Each layer was protected with $NFEC(p)$ redundancy symbols.
- *Fixed error capacity*. For each GOP, $GOPSize$ matrices are created like in the "Frame" method, but each layer is protected with a *fixed a priori* error capacity (e.g., the base layer is protected with $NFEC = 8$, the enhancement layer 1 is protected with $NFEC = 7$ and so on). This assignment of redundant symbols will remain constant for each frame of all the coded sequence.

4.2. Packet classification in routing nodes

The need of providing both protection against packet losses and limitations to the delivery packet delay has brought to the definition of several packet classification strategies at routing nodes:

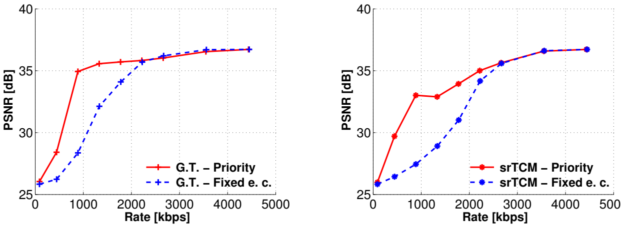
- *Game Theory (GT)*. The priority of each packet is determined considering the Nash equilibria [4]. Moreover, the utility function used by the peer G_d is

$$f_{G_d}(c_i) = \sum_{L=0}^{N-1} P\{\text{Loss layer L}|c_i\} \cdot D(L) \qquad (5)$$

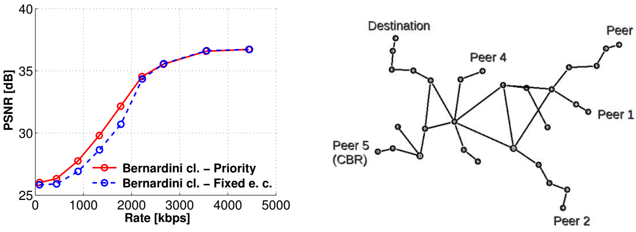
  where $D(L)$ is the GOP distortion caused by the loss of layer L-th (see work [4]) and c_i is the vector of the priorities assigned by all the peers at the instant i-th;
- *Bernardini classification*. The packet classification is fixed and constant for each stream [7];
- *srTCM*. The priority of each packet is assigned using the standard srTCM [8].

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of FEC-based MDC combined with UEP and GT-based QoS classification, several simulations have been run (using the network simulator NS2) comparing the different approaches of MDC-FEC matrix and packet classification discussed in section 4. In particular, three video streams had been encoded (relative to the sequences `foreman`, `news`, and `crew`) using an H.264/SVC encoder, splitting the sequences into GOP with GOP Size $= 8$, and organizing the stream in 8 quality (MGS) layers per frame. Priorities p (see section 3) vary in the range $[1, \ldots, 8]$. The MDC-FEC matrix (see Section 4) generates



(a) GT packet classification (b) srTCM packet classification



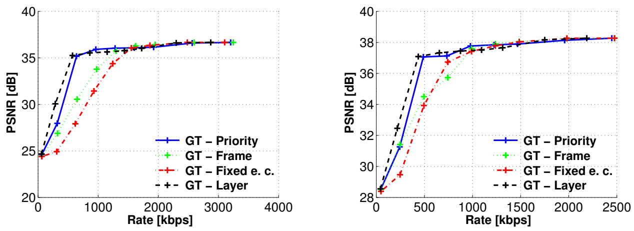
(c) Bernardini *et al.* classification (d) Gt-itm network

Fig. 3. Comparison between "Priority" and "Fixed error capacity" for sequence `crew`. Figure (d) displays one the random generated Gt-itm networks.

12 descriptions for each matrix. For each layer in every matrix, it was assigned a corrective capacity according to the strategies explained in Section 4. Finally, descriptions are transmitted on a P2P network (simulated in NS2 via a random GT-itm network as Fig.3d shows). Four source nodes transmits a set of 3 descriptions each and one additional node streaming a CBR packet flow to add extra interfering traffic in the network. One receiver node downloads all the streams and evaluates the PSNR for different congestion levels. In our tests GT-itm was used since our aim was to simulate the performance of the transmission once the overlay network has been defined and the peers start transmitting. As a matter of fact, the definition of a random physical network suits well this purpose.

Performances (for sequence `crew`) are compared in Fig.3, using GT packet classification in Fig.3a, srTCM classification in Fig.3b and Bernardini *et al.* classification in Fig.3c. Moreover, for each classification approach, FEC-MDC matrix "Priority" and "Fixed error capacity" has been compared. In these tests, we consider srTCM classification, combined with the approach FEC-MDC matrix "Fixed error capacity", the state-of-the-art since it combines the original approach in [5] with the standards [9] and [8].

As the graphs show, the matrix organizing strategy "Priority" allows to improve the performance with respect to "Fixed error capacity". This improvement is about 1 dB in case Bernardini *et al.* classification is chosen (Fig. 5c), 5 dB with srTCM classification (Fig. 5a), while in the case of GT classification of up to 6 dB (Fig. 5b). From the results, we observe that an adaptive packet classification based on an network conditions analysis (by means Game Theory or srTCM)



(a) `foreman` (b) `news`

Fig. 4. Comparison between four FEC-MDC composition technique using GT classification

bring an significantly performance improvement compared to a undisciplined classification (see G.T. Fig.3(a) or srTCM Fig.3(b) against Bernardini cl. Fig.3(c)). However, a Game Theory based classification with Nash equilibria is better than the simple standard srTCM (Fig.3a against Fig.3b). This is due to the fact that the Nash equilibria evaluate the optimal priority condition on a forecast of the future network state.

In the following, we compared the different filling strategies in Sec. 4 using the GT packet classification approach. This analysis has been performed both for the sequence `foreman` (Fig.4a) and `news` (Fig.4b).

With the same classification technique (in this case applying GT classification) MDC-FEC “Fixed error capacity” presents a bad performance compared to MDC-FEC “Layer” and MDC-FEC “Priority”. In the best case, MDC-FEC “Fixed error capacity” presents a performance close to the MDC-FEC “Frame”. In particular, the gap between the performance of MDC-FEC “Fixed error capacity” and MDC-FEC “Layer” is approximately 6 dB for the sequence `foreman` and about 4 dB for the sequence `news`. These results remain almost unchanged with the comparison between MDC-FEC “Priority” and MDC-FEC “Fixed error capacity”. Moreover, comparing MDC-FEC “Fixed error capacity” with MDC-FEC “Frame” the gap is approximately $1 - 2$ dB both in `foreman` sequence and in the `news` sequence. The same gap can be noticed by comparing the technique MDC-FEC “Priority” with MDC-FEC “Layer”.

These result suggest that by splitting the layers of the same frame in different MDC-FEC matrices, the probability of frame loss decrease in the condition of “burst” packet loss (typical behavior in presence of congestion in routing nodes). However, beyond a certain margin the performance can be considered equals (see the case of “Priority” against “Layer” in Fig. 4).

6. CONCLUSIONS

In this paper we have seen how the FEC-based MDC technique is suitable for video streams generated by the H.264/SVC coder provided that accurate protection levels and packet classification are operated. This technique converts the scalable stream into a MD stream meanwhile offering protection against packet loss. Experimental results show that this solution benefits from an Unequal Error Protection (UEP) of the different packets tuned conformingly to a packet prioritization. In addition, the performances utterly improves in case, together with the UEP, packets in the MDC-FEC matrix of different layers are organized properly. We also observed that final quality greatly benefits from an adaptive packet classification based on an network conditions estimates (by means Game Theory or srTCM). Future works will be devoted to extend the proposed approach to the transmission of 3D video signals.

7. REFERENCES

[1] Schwarz, H. and Marpe, D. and Wiegand, T., “Overview of the Scalable Video Coding extension of the H.264/AVC standard,” *IEEE Transaction on CSVT*, vol. 17, no. 9, pp. 1103–1120, 2007.

[2] S. Milani and G. Calvagno, “A Game Theory Based Classification for Distributed Downloading of Multiple Description Coded Videos,” in *Proc. of the IEEE ICIP 2009*, Cairo, Egypt, Nov. 24 – 28, 2009, pp. 3077 – 3080.

[3] X. Wu, B. Ma, and N. Sarshar, “Rainbow network problems and multiple description coding,” *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4565 – 4574, Oct. 2008.

[4] S. Milani, M. Gaggio, and G. Calvagno, “3dtv streaming over peer-to-peer networks using fec-based noncooperative multiple description,” in *Proc. of ISCC 2011*, June 28–July 1, 2011, pp. 13 –18.

[5] R. Puri and K. Ramchandran, “Multiple description source coding using forward error correction codes,” in *Proc. of Asilomar Conf. on SSC 1999*, Pacific Grove, CA , USA, Oct.24 – 27, 1999, vol. 1, pp. 342 – 346.

[6] Livio Lima, Massimo Mauro, Tea Anselmo, Daniele Alfonso, and Riccardo Leonardi, “Optimal rate adaptation with integer linear programming in the scalable extension of h.264/avc,” in *Proc. of ICIP 2011*, sept. 2011, pp. 1625 –1628.

[7] 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 ICIP 2007*, San Antonio, TX, USA, Sept. 2007, pp. 245 – 248.

[8] J. Heinanen and R. Guerin, “A Single Rate Three Color Marker,” RFC 2697, Sept. 1999.

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