

PERCEPTUAL CLASSIFICATION OF MPEG VIDEO FOR DIFFERENTIATED-SERVICES COMMUNICATIONS

Fabio De Vito¹, Laura Farinetti¹, Juan Carlos De Martin*

¹ Dipartimento di Automatica e Informatica/*IRITI-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [f.devito|farinetti|demartin]@polito.it

ABSTRACT

We present a distortion-based packet marking technique for transmission of motion-compensated video over Differentiated Services networks. For each macroblock of an MPEG-2 video sequence, the distortion that would be caused at the receiver by its loss is computed. High distortion macroblocks are grouped into perceptually-important slices that can be transmitted as *premium* packets, while lower distortion slices are sent as less expensive, *best-effort* traffic. Firstly, computation of the distortion introduced in the current frame only is compared to exhaustive computation of the distortion introduced in the entire group of pictures (GOP) due to the error propagation. Secondly, allocation of the premium traffic on a frame-by-frame basis is compared to GOP-wide allocation. Results show that GOP-wide allocation of premium traffic is key in using premium bandwidth efficiently, with strong PSNR gains with respect to the other approaches. We also propose a model-based distortion computation technique, which, combined with GOP-level premium traffic allocation, delivers nearly the same performance of the exhaustive approach at a fraction of its complexity.

1. INTRODUCTION

Many emerging multimedia applications require the delivery of audio and video over noisy channels with delay constraints that do not allow data retransmission. Error resilience without retransmission can be achieved by means of several techniques, including error-correcting codes (FEC), multiple description coding and priority-based forwarding techniques [1]. When using these techniques, a single channel can be viewed as partitioned into “virtual” subchannels characterized by different error rates. Each subchannel is used to deliver bits according to their perceptual importance; bit classification is normally done at design time through data partitioning, possibly combined with layered coding, and then kept constant during transmission.

In the context of IP networks, the priority-based forwarding approach called Differentiated Services (Diff-Serv) [2] is one of the most promising proposals to introduce Quality of Service (QoS) guarantees. With this approach, each packet has a label (present in both IPv4 and IPv6 headers) that indicates with which priority each router should handle the packet. The simplest implementation of DiffServ uses a 1-bit label that defines a *premium* class with low delays and no losses, and a *regular* best-effort class.

We propose to mark multimedia data adaptively depending on the desired level of perceptual quality of service and the distortion that their loss would introduce at the decoder. To do so, the marking process replicates the decoder behavior in case of data loss and computes the distortion between the original data and the data generated by the decoder to replace the missing packet. High-distortion, i.e., perceptually important data are sent as premium packets, while the remaining part of the stream is sent as regular traffic. The approach was recently applied to intra-coded MPEG-2 video sequences in [3]. We now propose to extend it to the classification and transmission of generic *motion-compensated MPEG-2 streams*. In this case, due to error propagation, packet losses will generally introduce distortion not only in the current frame, but in future frames as well. We compare the performance of the proposed technique when only the distortion introduced in the current frame is considered, to the performance achieved by taking into account future, GOP-wide distortion as well. We also compare allocation of the premium bandwidth performed on a frame-by-frame basis to higher-delay, GOP-wide allocation. Based on the analysis of the error propagation statistics of several video sequences, we also propose a model to estimate future distortion based on current distortion only. This model allows to obtain performance levels comparable to exhaustive GOP-level distortion computation at a fraction of its delay and complexity.

The paper is organized as follows. Section 2 describes the scenario. Section 3 explains the behavior of four proposed packet marking algorithms. Results about transmis-

sion performance are reported in Section 4. Section 5 describes the proposed model to estimate future distortion, with results about its performance. Finally, conclusions are made in Section 6.

2. SCENARIO DESCRIPTION

We focused on the specific case of MPEG-2 video encoding [4]. The basic element for the description of texture and motion information of each picture is the so-called “macroblock” (MB), which refers to an area of 16×16 pixels. In the MPEG standard, an arbitrary number of consecutive macroblocks belonging to the same row is coded into a “slice.” In the MPEG syntax a slice is the smallest unit which can be decoded independently. When MPEG video is transmitted over a packet network, packets should contain an integer number of slices since, in case of packet losses, a partial slice is generally useless [5]. Frames can be encoded either without any reference to other frames (I-pictures) or as differences with respect to adjacent frames (P- and B-pictures). Adjacent pictures (in coding order) form a logical unit called “group of pictures” (GOP). Each GOP begins with an I-picture and ends with the next I-picture (not included). Because of the temporal prediction, distortion due to packet losses may propagate within the GOP until an I-picture is reached. The reference MPEG-2 decoder [6] was modified to implement a simple concealment technique in which a lost MB is replaced by the MB in the same position in the *anchor* frame; in particular, if the lost MB belongs to an I- or B-picture the anchor frame is the previous frame in *display order*, otherwise, in the case of a P-picture, the anchor frame is the previous frame in *coding order*.

The network architecture considered in this paper manages two classes of traffic, a *premium* class and a *regular*, best-effort class. It is straightforward to generalize this scenario to more than two classes. *Premium* packets have higher priority in the routers queues and are always delivered; *regular* packets, instead, receive a best-effort service, and therefore neither a maximum amount of delay nor the delivery are guaranteed. Figure 1 shows a DiffServ network architecture. The marking algorithm performs macroblock classification; in general this process can be performed outside the encoder although at a higher computational cost. At the receiver the original stream is reconstructed, packet losses are, if needed, concealed, and video data decoded.

3. MARKING ALGORITHMS DESCRIPTION

We propose to express the perceptual importance of a macroblock in terms of the distortion that would be introduced by its loss. Such distortion can be defined as the MSE between the video sequence decoded using the correct data and the video sequence decoded using the replacement data

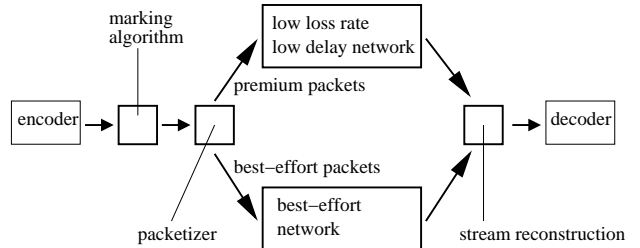


Figure 1: Block diagram of a 2-class DiffServ network architecture.

generated by the concealment technique at the decoder. If we have a constraint on the maximum premium share that can be used, then the problem can be formulated in terms of rate-distortion optimization. Our packet marking algorithm sorts MBs in decreasing order of perceptual importance and marks the most important MBs until the premium share is reached. Adjacent macroblocks belonging to the same class are put in the same slice. Then slices of the same class are grouped into packets. Packets containing *sequence headers* and *picture headers* are always marked as premium since their loss severely degrades decoder output.

According to the way in which *distortion computation* and *macroblock marking* are performed we can identify four main approaches to perceptual classification.

3.1. Frame- vs. GOP-level Distortion Computation

The overall distortion introduced by a macroblock in case of loss is the sum of the distortion introduced in the *current frame* plus the distortion introduced in the *future frames* due to error propagation. *Frame-level distortion computation* takes into account only the first error term, while *GOP-level distortion computation* considers both error terms. Frame-level distortion computation requires lower delay and lower complexity, at the expected cost of lower performance. GOP-level distortion computation, on the other hand, is significantly more complex, since it requires, for each macroblock, decoding and error concealment of all subsequent frames until the end of the current GOP.

3.2. Frame- vs. GOP-level Premium Marking

The marking algorithm can operate with two different scopes: *one frame* or *the complete GOP*. In the former case, a given share of each frame is marked as premium independently of the distortion levels of the other frames of the GOP. In the latter case, instead, the premium share is allocated over the whole GOP; the marking module has complete knowledge of where the highest distortion macroblocks are located in the GOP and can, therefore, allocate the available premium bandwidth where it is most

needed. Figure 2 shows an example of two approaches: with frame-level marking, a constant share of each frame is marked as premium (gray areas), while with GOP-wide marking, premium bandwidth is concentrated to protect high-distortion regions. Frame-level marking is a low-delay, low-complexity scheme that suits interactive transmission scenarios, at the expected price of lower quality performance compared to GOP-level marking. GOP-level marking, on the other hand, although more complex, has the potential of significantly more efficient use of network resources.

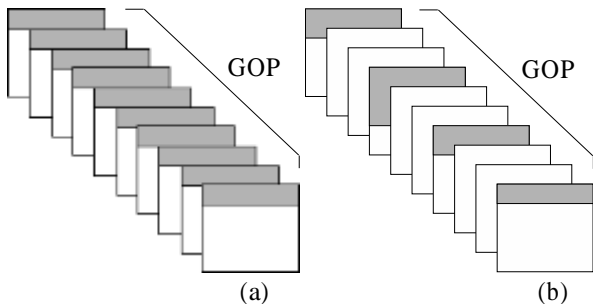


Figure 2: Representation of (a) frame-level and (b) GOP-level premium marking; premium MB's are shown in grey.

4. RESULTS

Figure 3 shows PSNR values corresponding to the four approaches presented above, as a function of packet loss rate. The first three GOP's of the *Mobile* sequence were used; 30% of the overall traffic was marked as premium. PSNR is computed with respect to the error-free sequence. Best-effort packet losses are uniformly distributed, while premium packets are not subject to losses.

The first two curves from the top correspond to GOP-based premium marking, while the third and fourth curves correspond to frame-based premium marking. The freedom to use premium bandwidth where it is most needed within the GOP, instead of allocating it rigidly on a frame-by-frame basis, delivers strong gains in terms of perceptual quality.

Regarding distortion computation, instead, our results show that there is not a great difference in performance between frame- and GOP-level distortion computation.

The regular best-effort case, as in the current Internet, is represented by the lowest curve; as can be seen, the performance gain delivered by DiffServ transmission is always quite significant.

Other simulation results, not reported here, show that absolute PSNR values of all curves increase or decrease as we increase or decrease the premium bandwidth, as expected.

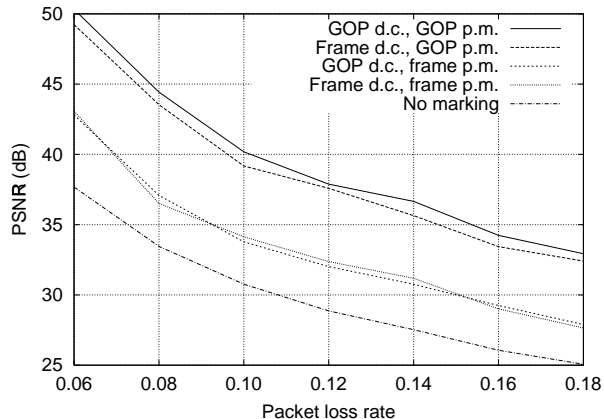


Figure 3: Comparison of the four proposed marking algorithms for increasing packet loss rates: GOP- and frame-level distortion computation (d.c.) and premium marking (p.m.); 30% premium bandwidth.

Table 1: Relative weight of current and future distortion averaged on different video sequences, with standard deviation values.

Frame type	Distortion produced in		Standard deviation (%)
	current frame (%)	future frames (%)	
I	14.8	85.2	2.4
First P	18.9	81.1	2.8
Second P	27.7	72.3	2.9
Third P	65.9	34.1	3.1

5. MODEL OF FUTURE DISTORTION

The total distortion introduced by a macroblock loss is the sum of two contributions: the one produced in the frame the macroblock belongs to, *current distortion*, and the distortion of all macroblocks in the following frames that reference the examined macroblock, *future distortion*. We observed that, for similar frame type and position in the GOP, the ratio between the current and future distortion is quite constant for all observed sequences: current and future distortion of a macroblock seems to depend strongly on the type of frame it belongs to, and weakly on the particular sequence.

Table 1 reports the results of a statistical analysis of several video sequences. B-frames are not considered because they are never referenced by other frames, therefore their future distortion is always zero.

We used these values to build a simple model to estimate the future distortion, with the aim to achieve the benefits — which, although not great with respect to frame-level dis-

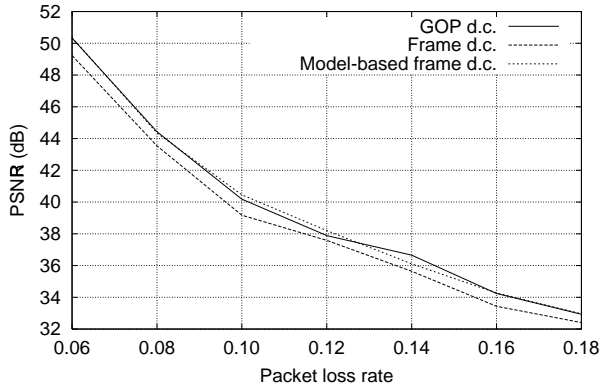


Figure 4: Comparison of three distortion computation (d.c.) algorithms: GOP-level, frame-level and model-based; premium marking is GOP-based in all three cases; premium bandwidth is 30%.

tortion computation, are still measurable— of GOP-based distortion computation at a much lower complexity.

Given the statistics reported above, it is possible to state that the influence of distortion due to prediction depends primarily on the type of frame the macroblock belongs to. Therefore we can define a weighting function $W(type)$ to estimate the total distortion D_{tot} of the i -th macroblock belonging to an I- or a P-picture simply knowing the distortion D_{curr} it introduces in the current frame.

$$D_{tot,i} = W(type) \cdot D_{curr,i}$$

Table 2: Weights for estimating future distortion due to error propagation as a function of frame type.

Frame type	Current distortion	Weight value $W(type)$
I	14.8 %	6.76
First P	18.9 %	5.29
Second P	27.7 %	3.61
Third P	65.9 %	1.52

Table 2 presents the proposed weights; the values are the reciprocal of the distortion produced in the current frame (see Table 1). The computational cost of this approach is minimal. Figure 4 compares the performance of three distortion computation methods: current-frame only, model based, and GOP-level computation. The performance obtained with the model-based algorithm is very close to the one of the GOP-level approach, but at a greatly reduced complexity.

6. CONCLUSIONS

We presented new perception-based techniques for the transmission of motion-compensated compressed video over Differentiated Services packet networks. We compared four distortion-based packet marking algorithms, combinations of, respectively, frame- and GOP-level approaches to *distortion computation* and *premium marking*.

Strong performance gains are associated to GOP-wide premium marking. The freedom to allocate premium bandwidth where it is most needed within the GOP, instead of being rigidly assigned on a frame-by-frame basis, seems to be more important than accurate computation of both current and future distortion. To reduce the complexity and delay of the GOP-level distortion computation, however, an analysis of error propagation in MPEG sequences was carried out. This led to the definition of a simple model to estimate future distortion based on current frame type and current distortion only. In all considered cases, transmitting as little as 30% of the video traffic as premium packets delivered significant PSNR gains with respect to the regular best-effort case.

7. ACKNOWLEDGEMENTS

The authors would like to thank Ing. Enrico Masala and Ing. Davide Quaglia of the Internet Media Group of the Politecnico di Torino for their precious help during the course of this research.

8. REFERENCES

- [1] Y. Wang and Q. Zhu, “Error control and concealment for video communication: a review,” *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
- [2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” *RFC 2475*, December 1998.
- [3] E. Masala, D. Quaglia, J.C. De Martin, “Adaptive Picture Slicing for Distortion-Based Classification of Video Packets,” in *Proc. IEEE Workshop on Multimedia Signal Processing*, Cannes, France, October 2001, pp. 111–116.
- [4] ISO/IEC, “MPEG-2 generic coding of moving pictures and associated audio information,” *ISO/IEC 13818*, 1996.
- [5] R. Hoffman, G. Fernando, V. Goyal, and M. Civanlar, “RTP Payload Format for MPEG1/MPEG2 Video,” *RFC 2250*, January 1998.
- [6] S. Eckart and C. Fogg, “ISO/IEC MPEG-2 software video codec,” *Proc. SPIE*, vol. 2419, pp. 100–118, Apr. 1995.