DISTORTION-BASED PACKET MARKING FOR
MPEG VIDEO TRANSMISSION OVER DIFFSERV NETWORKS

Juan Carlos De Martin

IRITI-CNR
Politecnico di Torino
C.so Duca degli Abruzzi 24
I-10129 Torino, Italy
E-mail: `demartin@polito.it`

Davide Quaglia

Dipartimento di Automatica e Informatica
Politecnico di Torino
C.so Duca degli Abruzzi 24
I-10129 Torino, Italy
E-mail: `davide.quaglia@polito.it`

ABSTRACT

We present a distortion-based approach to packet classification for multimedia transmission over differentiated-services packet networks. Instead of sending all traffic as *premium* or relying on a priori data partitioning, packets are individually examined and assigned to different service classes depending on the level of distortion that their loss would introduce at the decoder. Applied to video sequences encoded with the ISO MPEG-2 video coding standard, the proposed distortion-based packet marking scheme outperforms source-transparent techniques and provides substantial and consistent gains in PSNR over the regular best-effort case sending as little as 10% of the packets as premium traffic. Video samples are available at `http://multimedia.polito.it/icme2001/`.

1. INTRODUCTION

Real-time multimedia services require rather stringent Quality of Service (QoS) guarantees. The Differentiated Services (DiffServ) architecture [1] is one of most promising proposals that have recently been made to introduce QoS guarantees over IP networks.

In a DiffServ architecture, packets are assigned to one of a few classes to receive a specific forwarding behavior on nodes along their path. Possible behaviors range from the "virtual wire" case, with low-delay and no packet losses, to the traditional best-effort case, as in the current Internet, with unbounded delay and losses. To each behavior corresponds a cost defined by the provider-user service agreement.

Delay- and losses-sensitive traffic, such as audio or video, in a DiffServ scenario would be interely assigned, cost permitting, to a "premium" class; less time–critical data as best- or quasi-best-effort traffic.

Transmission of multimedia traffic as premium in its entirety clearly delivers very high perceptual quality to end users. Premium bandwidth, however, besides being the most expensive, is also a limited resource. The growth of multimedia traffic over data networks threatens to saturate its availability in corporate as well as in carrier networks rather quickly.

If only a fraction of each flow could be sent as premium and the rest were sent as best-effort, the load on the premium bandwidth would be reduced, thus permitting a higher number of simultaneous streams.

To maximize perceptual quality, the packets marked as premium should be the most perceptually relevant. Current approaches to packet marking, however, are usually source transparent. In [2], for instance, adaptive packet marking delivers soft bandwidth guarantees by randomly marking a certain share of the packets of a flow. Although simple, this approach does not exploit the fact that in speech, audio and video transmission certain packets are more perceptually important than others. Other techniques, instead, mostly proposed for layered video transmission over ATM networks (see, e.g., [3], for a recent survey), distiguish between high-priority and low-priority data, but rely on a priori data partitioning, not on packet-by-packet analysis.

We propose a distortion-based approach to packet marking and we apply it in the specific case of MPEG video transmission. Packets containing video data are individually examined and marked depending on the estimated distortion that their loss would introduce at the decoder and the desired level of perceptual quality of service.

The paper is organized as follows. In Section 2, the distortion-based approach to packet marking for multimedia transmission is explained. In Section 3, the approach is presented for the specific case of the ISO MPEG-2 video coding standard. Results of tests comparing the proposed method to current techniques are presented in Section 4. Finally, conclusions are presented in Section 5.

2. DISTORTION-BASED PACKET MARKING

2.1. Overview

Let us assume that a 1-bit DiffServ architecture is adopted (it is straightforward to generalize this example to the case of more than two classes): video packets are transmitted either on a low-delay, no-losses "virtual wire" (a concept recently proposed by Jacobson et al. [4]) or on regular best-effort network links subject to potentially unbounded delays and packet losses. Figure 1 shows packet classification and marking for such kind of architecture.
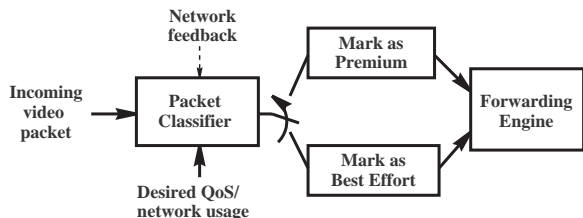

Figure 1: 1-bit packet marker.

The packet classifier examines the incoming video packet and, depending on the desired levels of QoS and network usage, assigns it to either the premium class or the best-effort class. The decision can also be a function, if available, of the current state of the network to further improve performance.

Packet marking for high-quality video (or in general, multimedia) transmission over DiffServ networks is often accomplished by marking as premium the entire flow. Premium bandwidth is, therefore, devoted to real-time transmission, and when no more bandwidth is available, service is denied or degrades without control.

Instead of assigning all packets of a given flow to either the premium class, as increasingly done on corporate networks, or the best-effort class, as is currently the case for most video services offered over the Internet, packet classification and marking can be performed on a *packet-by-packet basis*. Specifically, each individual video packet can be analyzed and assigned to one class or the other depending on its *perceptual importance*. To do so, the packet marker must be capable of decoding the payload and estimate the perceptual impact of the packet at the decoder.

The packet marker may also act as a function of the input video signal itself; in that case, however, packet marking can be accomplished only in the network node originating the flow (in Figure 2, Host 1). Classification based on compressed video alone, instead, may be accomplished at different points in the network (leaf router, border router, elsewhere), as shown in Figure 2.

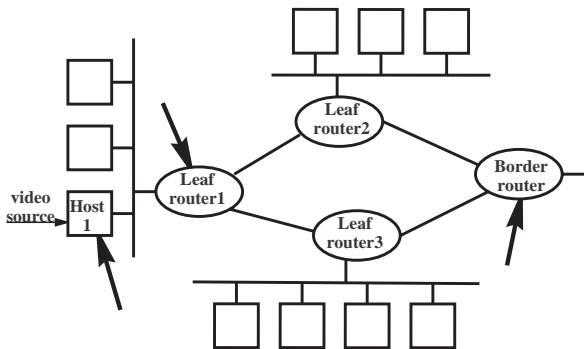From a complexity point of view, distortion-based

Figure 2: Possible placements of a distortion-based packet marker.

marking is best done at the encoder. Packet classification, in fact, can be easily generated as a by-product of the encoding operation at little or no extra cost in terms of computation.

2.2. Distortion-Based Marking

The perceptual importance of a packet can be expressed in terms of the distortion that would be introduced by its loss. The optimal measure of distortion would be to compare video decoded using the correct data and video decoded using the replacement data generated by the concealment technique at the decoder. To do so, the packet marker (whether encoder-based or stand-alone) needs to:

1. decode the video packet and generate the corresponding video sequence, C;

2. replicate the behavior of the decoder in presence of a packet loss and generate a replacement sequence, C';

3. compute a distortion measure between C and C'.

Ideally, a subjective distortion measure should be used. Absent that, the marking algorithm will be based on objective measures that predict reasonably well subjective performance, such as Mean Square Error (MSE) or Peak-Signal-to-Noise Ratio (PSNR). Figure 3 shows the block diagram of the proposed distortion-based marking scheme.
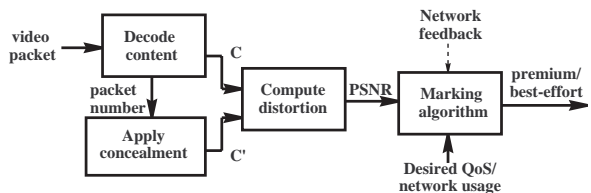

Figure 3: Block diagram of distortion-based packet marking.

Regarding step 2, the generation of the estimates, assumptions about the current state of the decoder need to be made. For low levels of packet losses it probably suffices to assume that the data needed by the concealment technique has been correctly received. More complex models, however, will take into account the probability that data in the past has been lost. This can be accomplished with, for instance, a Markov model of memory up to a few data blocks. In this case, the computation of the distortion would generate a range of values, one for each possible state of the decoder status model. The decision will then be made on the product $p_i D_i$, where p_i and D_i are the probability of being in state i and its associated distortion, respectively.

Moreover, for differentially encoded signals, the distortion measure will also include, ideally, the distortion introduced in future frames by the loss of the current video packet.

3. PACKET MARKING OF MPEG-2 VIDEO

We chose to test the proposed distorsion-based packet marking technique using video sequences coded at constant bitrate with the ISO MPEG-2 video coding algorithm [5]. The reference decoder software was modified to implement a simple concealment technique, described below.

3.1. System Configuration

Each slice of the compressed video bitstream is encapsulated in a different packet. This choice improves decoder resynchronisation after packet loss; slices, in fact, are delimitated by a start code in the bitstream and they are independent as far as differentially-encoded parameters are concerned.

Concealment techniques exploit the redundancy of the compressed bitstream. Concealment can be performed using the temporal correlation between consecutive frames or the spatial correlation between lost pixels and neighboring ones in a picture. In this work, decoder-side concealment is implemented by replacing a missing slice with the slice in the same position of the previous frame, the so-called *anchor slice*. This approach has the advantage of being simple; more sophisticated techniques could be adopted (see [6] for a recent survey).

3.2. Distortion Measure

We adopted the Mean Square Error (MSE) between a slice and its anchor as a measure of its suitability to concealment, that is, of its importance in the packet stream. High MSE values indicate that the concealment technique at the decoder, based on temporal correlation, cannot properly recover the missing information. Low MSE values, instead,

signal slices that are highly correlated with the corresponding slice in the previous frame.

Figure 4 shows the segmentation of a frame into slices (five slices per row) and associated MSE values: the darker a slice, the higher its corresponding MSE value. The dark slices in the lower right corner of the frame identify an object which has moved with respect to the previous frame.
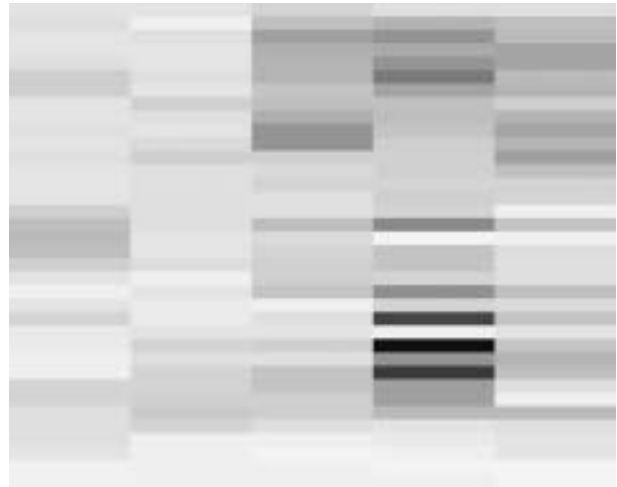

Figure 4: Slice-based MSE map of a frame; the darker a slice, the higher its corresponding MSE value.

3.3. Marking Algorithm

In general, the marking algorithm will depend on the share of traffic that we want to mark as premium and/or the desired level of perceptual quality of service generated at the decoder. Such constraints may be determined at design time, or made dependent on instantaneous network conditions, in which case packet marking would be a function of both the video source and the network status.

In this work, packets containing *sequence headers* or *picture headers* were marked as "premium" without further inspection. Missing header information, in fact, severely, if not completely, degrades decoder output.

The rest of the marking algorithm worked on the principle that slices producing the highest distortion —in our case, the highest MSE values— are to be marked "premium." Such choice could be a function of a given MSE threshold, in which case the premium share would fluctuate in time, keeping perceptual quality approximately constant. Alternatively, the marking algorithm could receive as input the percentage of packets to assign to premium service, creating a constant usage of premium resources and varying levels of quality of service. Cost-function based approaches are also possible in this context. Absent clear guidelines matching absolute MSE values to well-known subjective quality levels, we followed the constant premium–share approach,

which has also the advantage of simplifying the analysis of network usage. If k is the number of slices per pictures corresponding to the desired premium share, then the k slices with the highest MSE were marked as "premium." All other packets were sent as regular best-effort traffic.

4. RESULTS

We conducted formal tests to assess the performance of the proposed marking scheme. The test material was the 720×576 (CCIR-601 resolution) 40-frame standard video sequence known as *Mobile*. The simulation was performed on the sequence concatened with itself 50 times, for a total of 2000 frames, to achieve statistical significance in packet loss conditions. In the experiment, each packet contained exactly one slice; there were five slices per row. Coding was performed using the Main Profile, Main Level, operating at constant bit-rate. For simplicity's sake as well as to estimate the proposed technique without taking into account the effects of temporal redundacy, the sequence was encoded using I-pictures only. Video sequences were encoded using a fully-compliant software encoder known as ISO MPEG-2 Test Model 5 [7].
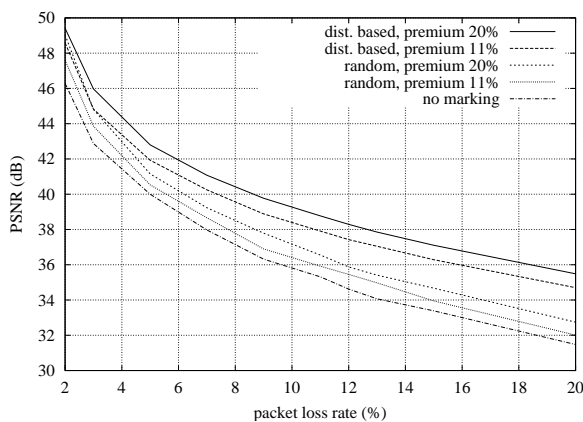
Figure 5: PSNR as a function of marking algoritm, premium share and network conditions.

We tested the performance of the proposed distortion-based marking algorithm against random, source-transparent marking for 11% and 20% premium shares and increasing levels of packet losses. Packet losses were applied only to best-effort packets; losses were uniformly distributed.

Figure 5 shows the performance versus packet loss rates. The "no marking" curve represents the performance of the quasi-best-effort scenario, for which the sequence headers were protected against errors, while the rest went unprotected.

The proposed distortion-based packet marking tech-

nique outperforms the source-trasparent approach by as much as 2.3 dB. With respect to the quasi-best-effort "no marking" case, the gain is an additional 0.5 dB or more.

The results provide clear evidence that delivering as little as 11% of the overall packets over premium bandwidth suffices to achieve substantial and consistent gains in PSNR over the regular best-effort scenario.

5. CONCLUSIONS

A distortion-based approach to marking multimedia content for packet networks offering differentiated services is proposed. Individual packets are classified as either premium or regular depending on the distortion that their loss would introduce at the decoder. A technique for MPEG-2 video was implemented and tested. Experiments showed that distortion-based packet marking clearly outperforms source-transparent techniques, and provides substantially and consistently higher PSNR values than the unprotected case sending as little as 11% of the packets as premium traffic.

6. REFERENCES

[1] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," *RFC 2475*, December 1998.

[2] W. Feng, D.Kandlur, D. Saha, and K. Shin, "Adaptive Packet Marking for Providing Differentiated Services in the Internet," in *Proc. ICNP'98*, Austin, Texas, October 1998, pp. 108–117.

[3] P. Salama, N.B. Shroff, and E.J. Delp, "Error concealment in MPEG video streams over ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1129–1144, June 2000.

[4] V. Jacobson, K. Nichols, and K. Poduri, "The 'Virtual Wire' Per-Domain Behavior," *Internet Draft draft-ietf-diffserv-pdb-vw-00.txt*, July 2000, work in progress.

[5] ISO/IEC 13818-2 MPEG-2 Video Coding Standard, "Generic coding of moving pictures and associated audio information—Part 2: Video," *ISO*, 1995.

[6] 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.

[7] S. Eckart and C. Fogg, "ISO/IEC MPEG-2 software video codec," *SPIE Proceedings*, vol. 2419, pp. 100–109, February 1995.