DELIVERY OF MPEG VIDEO STREAMS WITH CONSTANT PERCEPTUAL QUALITY OF SERVICE

Davide Quaglia[1], *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: [davide.quaglia|demartin]@polito.it

ABSTRACT

Constant levels of perceptual quality of service is what ideally users of multimedia services expect. In most cases, however, they receive time-varying levels of quality of service. This paper describes a technique to deliver nearly constant perceptual quality of service when transmitting video sequences over Differentiated Services IP networks. MPEG video packets are transmitted either as low-loss premium packets or as regular best-effort packets depending on their individual perceptual importance. On a frame-by-frame basis, allocation to the premium class is performed depending on the perceptual importance of each macroblock, the desired level of quality of service and the instantaneous network state. The resulting perceptually-based, time-varying use of premium and best-effort network resources delivers nearly constant quality of service to end users and it yields significant higher PSNR values compared to constant allocation of premium bandwidth.

1. INTRODUCTION

Good perceptual quality of service (QoS) will be a key factor for the success of many upcoming multimedia applications. Lower than expected levels of audio or video quality, in fact, could seriously compromise user adoption of new services, especially if users will be asked to pay more than they currently do for network access.

System design of audio-visual communications systems over IP has usually delivered solutions that, for a given bit rate, guarantee a certain *average* level of perceptual QoS to end users. Although well explained by rate-distortion theory and quite convenient for network planning, this approach, perhaps influenced by the usually fixed capacity of circuit-switching networks, generates time-varying levels of QoS.

A case could, instead, be made in favor of a *constant* perceptual QoS approach. At each time instant, the system will transmit exactly as many bits as necessary to guarantee the desired level of perceptual QoS. Instead of time-varying the perceived quality, potentially objectionable to end users, it will be the *bitrate* to vary with time, a normal traffic condition in packet-switched networks.

In this paper we propose a way to offer constant perceptual QoS transmitting video sequences over Differentiated-Services IP networks. More specifically, we describe a technique to assign MPEG video packets, either to a premium class or to a regular best-effort class so that the instantaneous perceptual QoS is kept as constant as possible. We formulate the marking problem as a rate-distortion optimization problem in which we minimize the premium bandwidth provided that a constraint on the overall distortion is satisfied. The distortion is estimated from the entropy of the video source and the packet loss rate of the network.

The paper is organized as follows. In Section 2, a framework for constant-QoS, perception-based packet marking of MPEG-2 video is presented and a specific algorithm described. In Section 3, transmission results show the behavior in the time of the proposed technique and its overall performance. Finally, conclusions are made in Section 4.

2. PERCEPTION-BASED PACKET MARKING OF MPEG-2 VIDEO

We focused on video sequences coded with the ISO MPEG-2 video coding algorithm [1]. The transmission channel is a 2-class DiffServ IP architecture. The reference decoder software was modified to implement a simple concealment technique, described below.

2.1. System Configuration

Test sequences are encoded using I-pictures only to avoid error propagation in future frames. No rate control is used. The quantization factor is the same for all pictures to obtain a stream of constant quality and constant bitrate. Each macroblock row is divided into a constant number of slices different for each sequence according to its frame width.

0-7803-7304-9/02/$17.00

Each slice of the compressed video bitstream is encapsulated in a different packet according to the IETF Request For Comments 2250 [2]. This choice improves decoder resynchronization after packet loss; slices, in fact, are delimited by a byte-aligned start code in the bitstream and they are independent as far as differentially-encoded parameters are concerned.

In this work, decoder-side concealment is implemented by replacing the lost slice with the slice in the same position in the previous frame. This approach exploits the interframe correlation and has the advantage of being simple. More sophisticated techniques could be adopted; see [3] for a recent survey.

Regarding the network, we adopted a 2-class DiffServ IP architecture, as previously done in [4]. With this approach, each packet has a label (present in both IPv4 and IPv6 headers) which dictates the behavior each router should use to handle it. For simplicity, we limited the classification to two classes, premium and regular; premium packets are transmitted on a low-delay, no-losses "virtual wire," while regular packets are transmitted on a best-effort Internet-like channel with potentially unbounded delays and losses. In our work packet losses are modeled using a Gilbert model [5] that captures the temporal dependency of packet losses in the Internet. Gilbert model parameters P and Q can be determined based on packet loss rate and average burst length. In this work, P and Q were determined using network simulation data for a transmission scenario such as the one described in this paper.

2.2. Optimal allocation of the premium bandwidth

The marking algorithm allocates the premium share depending on 1) the desired perceptual quality of service at the decoder and 2) the packet loss rate of the network. The first parameter is specified for the whole sequence as the maximum PSNR decrease (in dB) admitted with respect to the error-free sequence (PSNR is considered with respect to the original uncompressed frames); the second parameter can be obtained from the receiver (e.g., using the RTCP Receiver Report). In this work the packet loss rate is kept constant during the whole transmission, but the marking algorithm, as described below, can adapt the premium share according to changes in network conditions, since the allocation is performed on a frame-by-frame basis.

In our approach, packets containing *sequence headers* or *picture headers* are marked as "premium" without further inspection. In fact, if a sequence or picture header is corrupted, the decoder may skip the whole sequence or picture respectively, leading to severe loss of perceptual quality at the decoder. The other packets of the frame are marked according to the distortion that their loss would introduce at the decoder for a given level of packet loss rate. The goal is to minimize the premium share provided that the distortion of the received frame does not exceed the desired quantization distortion by a given threshold. This problem can be formulated as follows:

$$\min_{m \in M} R(m), \text{ subject to } D(m) \leq D_{max} \qquad (1)$$

where m is a marking pattern for the slices of the current frame, M is the set of all possible marking patterns, $R(m)$ is the premium bitrate caused by pattern m, $D(m)$ is the distortion of the received frame with respect to the original uncompressed frame, and D_{max} is the bound on the distortion of the received frame with respect to the original uncompressed frame. If N is the number of slices in the frame, then m can be defined as the vector $(x_1, x_2, ..., x_N)$, where x_i is 1 if the i-th slice is marked as premium, and 0 otherwise. Let $d_i(x_i)$ be the distortion of the portion of frame coded in the i-th slice; if this slice is marked premium, then it will not be lost and the distortion is caused by the video coding process only; on the other hand, if it is not marked, then it may be lost with probability p and the distortion will depend on the effectiveness of the concealment strategy for this specific slice. Let $\hat{d}_i$ be the distortion of the i-th slice due to coding noise (i.e., the MSE between the coded slice and the original uncompressed one) and let $\tilde{d}_i$ be the distortion due to error concealment (i.e., the MSE between the coded slice in the previous frame and the original uncompressed one in the present frame); then we define the distortion of the i-th slice in function of x_i as follows

$$d_i(x_i) = \begin{cases} \hat{d}_i & \text{if } x_i = 1 \\ p\tilde{d}_i + (1-p)\hat{d}_i & \text{if } x_i = 0 \end{cases} \qquad (2)$$

Equation (2) expresses the expected distortion of both regular and premium slices. Therefore the initial problem (1) can be reformulated as follows

$$\min_{m \in M} \sum_{i=1}^{N} x_i \text{ subject to } \sum_{i=1}^{N} d_i(x_i) \leq K \sum_{i=1}^{N} \hat{d}_i \qquad (3)$$

where K is the multiplicative increase of the distortion corresponding to an additive reduction of the PSNR (in dB) at the decoder. We also assume that all slices have the same number of bits.

Equation (3) controls the end-to-end distortion of the video sequence by the allocation of the premium bandwidth according to the source entropy and the channel conditions. In Equation (2), in fact, $\hat{d}_i$ depends on the effectiveness of the intra-coding compression on frame i, $\tilde{d}_i$ depends on the temporal correlation between frame i and frame $i-1$ (useful during error concealment), and p models the network state. Given Equation (3), then the constrained minimization in (1) can be converted into an equivalent unconstrained problem by merging rate and distortion through the Lagrangian multiplier λ. The unconstrained problem becomes the determination, for each frame, of the marking

strategy which gives the minimum of the Lagrangian cost function expressed as

$$J(\lambda) = R(m) + \lambda D(m). \quad (4)$$

For a given multiplier λ, the minimization of the cost function (4) results in a solution $m^*(\lambda)$ with associated rate $R^*(\lambda)$ and distortion $D^*(\lambda)$. It was demonstrated that if, for a given λ_c, $D^*(\lambda_c)$ happens to coincide with the targeted distortion bound D_{max} then $m^*(\lambda_c)$ is the solution of the constrained problem (1).

2.3. Marking Algorithm

Starting from (3), by shifting the rate-distortion curve of the amount $\sum_{i=1}^{N} \hat{d}_i$ we obtained the following Constant Quality (CQ) algorithm:

```
mark sequence header as premium
for (each frame in the sequence)
  mark picture header as premium
  start with all slices as regular
  while (p∑_{x_i=0}(d̃_i − d̂_i) > (K − 1)∑_{i=1}^{N} d̂_i)
      mark the regular slice w/ highest (d̃_i − d̂_i)
  end
end
```

The algorithm takes as parameters the packet loss rate p and the multiplicative increase K of the distortion corresponding to the maximum additive reduction of the PSNR (in dB) allowed at the decoder with respect to the error-free sequence (PSNR is considered with respect to the original uncompressed frames). In general K is given for the whole sequence, while p can change on a frame-by-frame basis (even if in these experiments it was kept constant for simplicity's sake).

3. RESULTS

We tested the performance of the proposed constant-QoS marking scheme on well-known video sequences covering different kinds of video material, from almost static "talking heads" to strong motion sequences. The simulation was performed on the sequences concatenated with themselves for a total of 1500 frames, to achieve statistical significance in packet loss conditions. In the experiment, there were eleven slices per row and each packet contained one slice. Video sequences were encoded using a fully-compliant software encoder known as ISO MPEG-2 Test Model 5 [6]. Coding was performed using the Main Profile, Main Level; for simplicity's sake, as well as to estimate the proposed technique without taking into account the effects of temporal error propagation, the sequences were encoded using I-pictures only. Table 1 reports, for each sequence, the format, the original length and the average encoding quality; since we are interested in keeping the quality constant, the standard deviation on the frame PSNR before transmission is also reported in brackets.

Sequence	Format	Frames	Quality (PSNR)
Foreman	QCIF	100	34.33 dB (0.95)
News	QCIF	100	34.20 dB (0.07)
Garden	CIF	250	31.99 dB (0.29)
Mobile	CIF	250	30.51 dB (0.09)

Table 1: Features of the video sequences used for the test. The number of frames refers to the original length of each sequence. The PSNR values refer to the coding distortion with Q=8. The standard deviation of the PSNR values is reported in brackets.

We marked the test sequences using the CQ algorithm with packet loss rate $p = 0.05$ (5%) and 1 dB of maximum decrease of PSNR with respect to the error-free sequence (PSNR is considered with respect to the original uncompressed frames). The parameter p is constant during the whole transmission. Then we marked the same sequence so that each frame had a constant premium share equal to the average premium share obtained with the CQ algorithm for the whole sequence. For this case we used an algorithm (referred as Constant Share, or CS) that marks all sequence and picture headers and, for each frame, a constant number of slices chosen among those with the highest $\tilde{d}_i$, as described in [4]. The same loss pattern was then applied to both the CQ and the CS marked streams. The loss pattern was generated using a Gilbert model with 5% average packet loss rate and 1.3 average burst length (P=0.04, Q=0.77).

Sequence	CQ	CS	Premium
Foreman	33.47 (1.14)	32.97 (1.9)	30%
News	33.56 (0.99)	33.43 (1.63)	3%
Garden	31 (0.7)	30.96 (1.22)	33%
Mobile	29.53 (0.46)	29.48 (0.75)	9%

Table 2: Average PSNR (standard deviation in brackets) for the CQ and CS marking algorithms with 5% packet loss rate. The PSNR is measured with respect to the original uncompressed frames. The overall premium share (varying in time for CQ, constant for CS) is also reported.

Table 2 reports the average PSNR for the CQ and CS approaches with 5% packet loss rate. The PSNR is measured with respect to the original uncompressed frames. The overall premium share (common to both CQ and CS) is also reported. The values for CQ show that the constraint on the max decrease of PSNR is satisfied for all sequences except *Foreman*. The standard deviation (reported in brackets) is significantly smaller for CQ than for CS. The premium

share depends on the sequence and, in particular, on the correlation between successive frames; in fact, the highest premium share has been allocated for *Garden* and *Foreman*, which contain camera panning.

Figure 1 shows the premium traffic allocation for the *Foreman* sequence as a function of frame number. The algorithm concentrates the premium share where the concealment technique fails; using temporal concealment, the premium share is high in frames which exhibit an innovation of the video information with respect to the past frames (e.g., all sequences show a high premium share in the first frame). In particular, the premium share of the *Foreman* sequence increases after frame 50 for a scene change with camera panning.

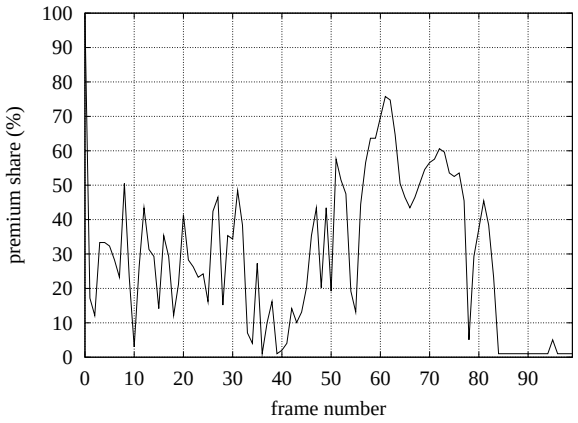

Figure 1: Evolution in time of the premium share allocation for the *Foreman* sequence with 5% packet loss rate and 1 dB maximum allowed decrease of PSNR.

Figure 2 compares the PSNR profile for both the CQ and the CS algorithms as a function of frame number for the *Foreman* sequence. PSNR is computed with respect to the original uncompressed frames. The CQ marking strategy performs better than CS especially in high entropy regions (e.g., frames 55-70) where error concealment fails. The constraint on the max decrease of PSNR is not satisfied for all frames.

4. CONCLUSIONS

We presented a technique to deliver MPEG-2 video sequences with nearly constant perceptual quality in a DiffServ IP framework. Video packets are transmitted either as premium, or as regular, best-effort packets, depending on the perceptual importance of each individual packet, instantaneous network conditions and the desired level of quality of service. For equal levels of transmitted premium traffic, the perceptual quality of the video sequence obtained employing the proposed technique is not only nearly constant in time, but is

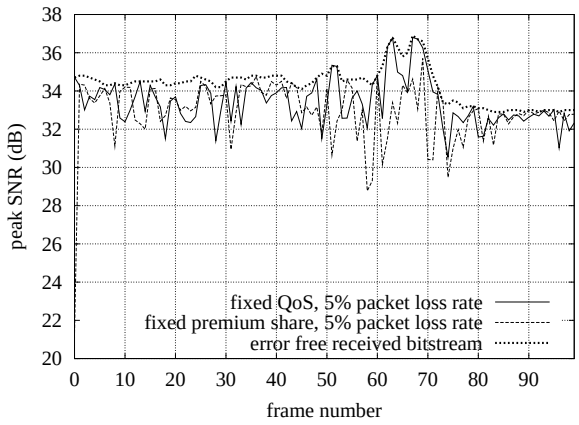

Figure 2: PSNR with respect to the original uncompressed frames for the *Foreman* sequence as a function of the marking algorithm (the premium bandwidth is 30% and the packet loss rate is 5%).

also on average better than the quality obtained by transmitting a constant share of each frame as premium. The approach is well suited to address the time-varying nature of many communications channel, both wireline and wireless, a contribute towards the implementation of ubiquitous multimedia applications.

5. REFERENCES

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

[2] R. Hoffman, G. Fernando, V. Goyal, and M. Civanlar, "RTP Payload Format for MPEG1/MPEG2 Video," *RFC 2250*, January 1998.

[3] 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.

[4] 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.

[5] W. Jiang and H. Schulzrinne, "Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality," in *Proceedings of NOSSDAV*, 2000.

[6] S. Eckart and C. Fogg, "ISO/IEC MPEG-2 software video codec," *Proc. SPIE*, vol. 2419, pp. 100–118, Apr. 1995.