ADAPTIVE PACKET CLASSIFICATION FOR CONSTANT PERCEPTUAL QUALITY OF SERVICE DELIVERY OF VIDEO STREAMS OVER TIME-VARYING NETWORKS

Davide Quaglia§, *Juan Carlos De Martin**

§Dipartimento di Automatica e Informatica/*IEIIT-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: `[davide.quaglia|demartin]@polito.it`

ABSTRACT

This paper describes a technique to deliver video streams with constant perceptual quality of service (QoS) over time-varying packet-erasure channels that support differentiated classes of service. During compression, the encoder estimates the distortion introduced at the decoder by each video packet both in case it is received and in case it is lost. Concealment and error propagation due to inter-frame prediction are taken into account. During transmission, an optimization algorithm assigns packets to different service classes according to the estimated distortion, the current channel status and a constraint on the desired quality at the receiver. This technique is compared with other video packet classification approaches in the specific case of a DiffServ IP network implementing the Assured Forwarding scheme. Network simulations show that the proposed technique delivers higher and more constant levels of perceptual QoS than traditional approaches. Moreover, the technique is characterized by reactiveness to congestions and fairness in the use of network resources.

1. INTRODUCTION

Assuring sufficiently high and stable levels of perceptual Quality of Service (QoS) is one of the key factors that will decide the success of the next generation of multimedia services. The best-effort model of traditional IP networks, however, combined with the challenges of the radio channel in the case of wireless transmission, do not favor meeting the stringent packet loss and delay requirements of successful audio and video transmission.

In the context of the well-known Rate-Distortion (RD) optimization, several have addressed the specific problem of minimizing the perceived distortion at the decoder given a constraint on the use of network resources [1]. In [2] an optimal algorithm is proposed for the inter/intra-mode switching during the video coding process. In [3] the optimization problem is addressed considering a number of delivery techniques but without taking into account the effect of concealment. The new video coding standard H.264 [4] also proposes the estimation of the end-to-end distortion at the encoder through the simulation of various error patterns. This simulation, however, does not take into account error propagation due to temporal prediction.

This work was supported in part by CERCOM, the Center for Wireless Multimedia Communications, Torino, Italy, http://www.cercom.polito.it.

In this work we adopt the RD optimization approach, but instead of minimizing distortion for a given transmission rate, we aim at minimizing the use of network resources while assuring a *constant level of perceptual quality of service* at the receiver. The main challenge of this optimization problem is the accuracy in the estimation of end-to-end distortion to which the constraint is applied. In fact, the relationship between end-to-end distortion and its causes cannot be described by simple models. The problem has been previously studied in [5] and also in [6], where a hybrid MMAX/MMSE criterion is used.

More specifically, we describe a technique to assign MPEG video packets to either a high-QoS class, or a regular, best effort class, so that the instantaneous perceptual QoS is kept as constant as possible. We formulate the packet assignment problem as a rate-distortion optimization problem in which we minimize the use of high-QoS bandwidth provided that a constraint on the overall distortion is satisfied. The expected distortion is estimated based on the instantaneous packet loss rate of the network and the entropy of the video source. We tested the proposed Adaptive Type Of Service (ATOS) technique in the specific case of a Differentiated Services IP network implementing the Assured Forwarding scheme. In [7], it was demonstrated the effectiveness of assigning video packets to different classes of service. The key features and original contributions of the proposed technique are: independence of coding standard, prediction scheme (i.e., temporal prediction, inter-layer prediction), and concealment strategy, end-to-end distortion estimation, awareness of the prediction dependencies, and an RD framework driven by the source entropy, the instantaneous network conditions and the desired level of perceptual quality of service.

The paper is organized as follows. Section 2 presents the distortion-based approach for the classification of video packets. Section 3 describes the Adaptive TOS (ATOS) algorithm for the delivery of video streams with constant perceptual QoS. Simulation results and comparison with other techniques are reported in Section 4. Finally, conclusions are drawn in Section 5.

2. PERCEPTUAL PACKET CLASSIFICATION

2.1. Distortion-based packet classification

In a video bitstream not all bits are perceptually-equally important and, for this reason, several data-partitioning techniques, usually combined with some form of unequal error protection, have been developed. The problem has two main aspects: 1) the estimation of the perceptual importance of each bitstream element, and 2) the allocation of resources (high-QoS bandwidth, forward error correction, etc.) to protect them from channel errors. This Section deals with the former issue, while the latter will be addressed in the next Section.

Classic data partitioning identifies the elements of the compressed bitstream that are statistically more important and then consistently protects them. This work proposes, instead, to estimate the importance of each individual video packet according to the distortion it generates at the decoder, as in [8]. If a packet arrives at the receiver, the distortion associated to it is solely due to the source coding process. If a packet is lost, it will cause a distortion that depends on the replacement data generated by the error concealment technique plus, in case of inter-frame prediction, the distortion due to error propagation in future frames.

Figure 1 shows the block diagram of the proposed classification process. The encoder takes the uncompressed video source and produces a set of data units (i.e., for MPEG, slices) which will be sent on the network as video packets. Let B_i be the size of the i-th data unit. The packet classifier computes the *encoding distortion*, denoted as $\hat{D}_i$, by comparing the decoded pixels with the corresponding pixels in the uncompressed video source; the encoding distortion measures the effectiveness of the source coding for the given data unit. The *concealment distortion*, denoted as $\tilde{D}_i$, is computed by simulating the concealment process on each data unit; the concealment distortion measures the effectiveness of the error concealment on the given data unit. Regarding the simulated recovery of the data unit, assumptions about the current state of the receiver need to be made, as in [2]. For low levels of packet losses it suffices to assume that the data needed by the concealment technique has been correctly received.

In the context of rate-distortion optimization and of IP networks supporting the Differentiated Services (DiffServ) Model, as in this paper, the RD optimization algorithm takes into account the set of values $\{B_i, \hat{D}_i, \tilde{D}_i\}$, the current packet loss rate and the desired value of perceptual QoS and then sets the *ToS field* in each IP packet to indicate the class to which the packet belongs.

From a complexity point of view, $\{B_i, \hat{D}_i, \tilde{D}_i\}$ can be easily generated as a by-product of the encoding operation at little or no extra cost in terms of computation. The rate-distortion optimization, instead, is performed during transmission since it takes into account the current channel status. In Section 3 an optimization algorithm, whose complexity grows linearly with the number of the data units, is presented.

2.2. Extension to motion-compensated video sequences

In case of inter-frame prediction, the distortion introduced by a lost data unit not only affects the corresponding pixels in the current frame but also propagates to future frames that use such pixels for prediction. For this reason $\tilde{D}_i$ should take into account both present and future distortion terms. The error on a generic pixel k also affects all those pixels that use pixel k as prediction. The pixel-level concealment distortion, denoted as $\tilde{d}_k$, may be roughly approximated assuming that the distortion associated to each predicted pixel is the same as the distortion on pixel k. Mathematically, the distortion in the current frame is, therefore, multiplied by $\alpha_k + 1$ where α_k is the number of predicted pixels depending on pixel k (known at encoding-time). The concealment distortion $\tilde{D}_i$ of the i-th data unit is the sum of the distortion contributions of its pixels. This definition of $\tilde{D}_i$ can be straightforwardly extended to all types of prediction.

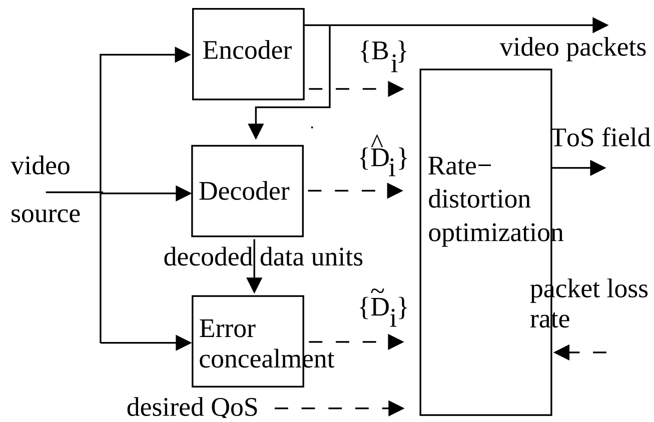

Fig. 1. Block diagram of the proposed classification technique.

3. OPTIMAL CONSTANT-QUALITY NETWORK-ADAPTIVE CLASSIFICATION

The proposed technique aims at keeping the perceived quality as constant as possible. It is based on the assumption that the transmission channel can be partitioned into “virtual” sub-channels, each characterized by a different packet loss rate. This Section describes a method to use the most reliable classes to deliver the perceptually most important packets.

For simplicity’s sake we deal with two classes, a *premium* class with low packet loss rate and a *regular* best-effort class with potentially unbounded losses. The process of assigning each packet to one of the two classes is called *packet marking*. We refer to the proposed marking technique as the Adaptive Type Of Service (ATOS) algorithm.

Each video source uses the ATOS algorithm to allocate the premium bandwidth depending on 1) the desired perceptual quality of service (PQoS) at the decoder and 2) the packet loss rate of the network. The desired PQoS is specified for the whole sequence as the maximum PSNR decrease (in dB) with respect to the error-free sequence (PSNR is computed with respect to the original uncompressed frames); the instantaneous packet loss rate can be obtained from the receiver (e.g., through RTCP Receiver Report packets). The allocation of the premium share is performed for each “group of pictures” (GOP) to take into account the error propagation due to inter-frame prediction. This choice increases the algorithmic delay; we assume, however, that this delay can be tolerated in non-interactive applications. Using periodically updated information on packet loss rate, the marking algorithm adapts the use of premium share to changes in network conditions.

Packets containing *sequence headers* or *picture headers* are marked as premium without further inspection. In fact, if these headers are corrupted, the decoder may skip the whole sequence or picture respectively, leading to severe loss of perceptual quality. The other packets of the GOP 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 use of premium share provided that the expected distortion at the receiver does not exceed the encoding distortion by a given threshold. This problem can be formulated as follows:

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

where m is a marking pattern, M is the set of all possible patterns, $R(m)$ is the number of premium bits, $D(m)$ is the distortion of the received, and concealed, frames of the GOP with respect to the original uncompressed ones. If N is the number of data units in the GOP, then m can be defined as the vector $(x_1, x_2, ..., x_N)$, where x_i is 1 if the i-th data unit is marked as premium, and 0 otherwise. Let p_0 and p_1 be the packet loss rates of the regular and premium class respectively. Let $\hat{D}_i$ and $\tilde{D}_i$ be the encoding and concealment distortions of the i-th data unit as described in Section 2; then the expected contribution to distortion of the i-th data unit in function of x_i can be defined as follows

$$D_i(x_i) = \begin{cases} p_0\tilde{D}_i + (1-p_0)\hat{D}_i & \text{if } x_i = 0, \\ p_1\tilde{D}_i + (1-p_1)\hat{D}_i & \text{if } x_i = 1. \end{cases} \quad (2)$$

Therefore the initial problem (1) can be reformulated as follows

$$\min_{m \in M} \sum_{i=1}^{N} B_i x_i \text{ subject to } \sum_{i=1}^{N} D_i(x_i) \leq K \sum_{i=1}^{N} \hat{D}_i, \quad (3)$$

where B_i is the size of the i-th data unit and K is the multiplicative increase of the distortion corresponding to an additive reduction of the PSNR (in dB) at the decoder. In general p_0 and p_1 vary with time according to channel conditions, while K is a system-defined parameter that remains constant. Equation (3) controls the end-to-end distortion of the video sequence through GOP-level allocation of the premium bandwidth according to source entropy and channel conditions. In fact, in (2) $\hat{D}_i$ and $\tilde{D}_i$ depend on the effectiveness of the source coding and error concealment respectively, while p_0 and p_1 model the channel state.

Given (3), the constrained minimization in (1) can be converted into an equivalent unconstrained problem by merging rate and distortion through the Lagrangian multiplier λ, as shown in (4).

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

This problem is solved using a fast iterative approach known as the bisection algorithm [1], whose complexity grows linearly with the number of data units in the GOP.

4. SIMULATION RESULTS

The NS-2 package [9] was used to simulate the transmission of MPEG-2 video sequences over a Differentiated Services IP network implementing the Assured Forwarding (AF) scheme [10]. With reference to Section 3 the AF class represents the premium class, while the best-effort (BE) class represents the regular class. We assumed that the packet loss rate p_1 of the AF class is negligible, an assumption confirmed by the simulation results.

In the simulated network scenario ten sources transmit video packets to their corresponding receivers. All flows transit through the same bottleneck which has a bandwidth of 6 Mb/s. The other links are oversized in bandwidth not to impact on the results. Each video source produces RTP/UDP packets which are marked according to one of the marking techniques under investigation.

We used the *Foreman* video sequence. The sequence format is QCIF (176 × 144 pixel) and contains 100 frames. Each GOP consists of an I-picture followed by nine P-pictures. To keep the encoding quality constant we used a fixed quantization factor. The encoder is a modified version of the Test Model 5 Encoder by the MPEG Software Simulation Group. For decoding, we used the MPEG-2 Reference Decoder version 1.2 by the same authors; the decoder was modified to implement a simple concealment technique, i.e., lost areas are replaced with the corresponding ones in the previous frame.

The following four test cases are considered: 1) ATOS classification, 2) Fixed-AF-share classification, 3) 100% best-effort (BE) traffic, and 4) Prescient BE sources. Each test case consists of ten video sources implementing the same classification technique. Sources begin to transmit at different times to avoid potential synchronization problems. The simulation length is 600 s for all test cases (the original video sequence is concatenated with itself 150 times). Sources transmit (except in test case 4) at an average bitrate of 620 kb/s, which corresponds to a PSNR of 35.6 dB, with standard deviation of 0.68. To vary the network state during the simulations, one of the ten sources (referred below as source 1) stops to transmit for 150 s at time 100 s and 350 s. When source 1 is active, the aggregate bitrate exceeds the bottleneck capacity, causing an overall packet loss rate of approximately 3%.

In Test Case 1, parameter K of Equation (3) was set to 1.26, corresponding to 1 dB of maximum PSNR decrease (computed with respect to the original uncompressed frames). Test Case 2 illustrates the behavior of an implementation of the a priori data partitioning approach. Within each GOP the technique always marks what experimentally we found to be the most important bitstream elements, that is, the I-picture, the first P-picture and all headers (see Figure 2). In Test Case 3 all traffic is assigned to the BE class. In Test Case 4, Prescient Best-Effort, the source bitrate is chosen a priori to avoid packet drops. Each source was thus coded with a higher quantization factor to obtain the target bitrate of 600 kb/s, corresponding to a PSNR of 34.5 dB (standard deviation of 1.3).

Figure 2 shows the allocation of the AF bandwidth within a GOP for both an ATOS and a Fixed-AF-share source. The AF share is about 30% for both sources. For this specific example, the distortion-based classification algorithm counterintuitively protects the first and the second P-frames more than the I-frame itself. Figure 3 compares the allocation of the AF bandwidth as a function of time for both an ATOS and a Fixed-AF-share source. The ATOS source, as desired, uses AF bandwidth as network conditions demands (i.e., mostly during congestions), then releases it.

Table 1 compares the PSNR performance of all tested classification techniques. PSNR standard deviations are also shown since they are a good indication of how stable is the perceptual quality of service. For the ATOS technique the decrease in PSNR with respect to the error-free sequence is less than 1 dB (34.8 versus 35.6 dB), as desired. Moreover, ATOS outperforms the Fixed-AF-share approach, achieving a substantially higher (+2.3 dB) as well as more constant PSNR. The results for the ATOS technique are indeed quite close to those obtained with the prescient ideal approach of Test Case 4.

Table 2 compares five typical video sources for Test Case 1. The PSNR values, all quite similar, suggest that the ATOS algorithm leads to a fair use of networks resources among competing video streams. Source 1 experiences a higher average byte loss rate since, by simulation design, while it is active, the network is always congested. Higher loss rates lead, as expected, to a larger use of AF bandwidth to achieve the desired level of PQoS.

Video transmission assigning all packets to the AF class was also simulated. As expected, however, when 100% of the traffic is

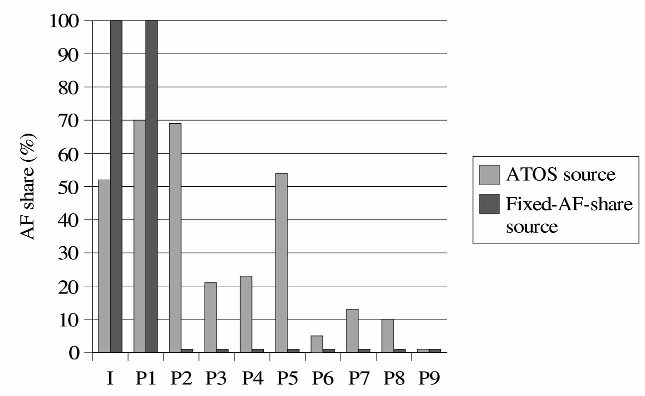

Fig. 2. Allocation of the AF bandwidth within a GOP for an ATOS source and a Fixed-AF-share source; the average AF share is about 30% for both sources.

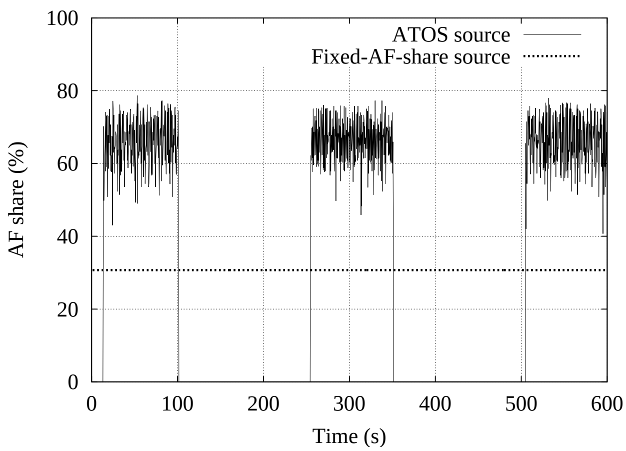

Fig. 3. Allocation of the AF bandwidth as a function of time for an ATOS source and a Fixed-AF-share source.

sent as premium, the DiffServ mechanism collapses, resulting in regular best-effort network behavior (i.e., Test Case 3).

5. CONCLUSIONS

We presented a new algorithm to deliver video streams with constant perceptual quality of service. The Adaptive Type Of Service (ATOS) algorithm protects video packets according to their perceptual importance, the current channel status and the desired level of perceptual QoS. The distortion introduced by each video packet is estimated during encoding, taking into account concealment and inter-frame error propagation. A rate-distortion optimization is performed during transmission by a low-complexity algorithm. The ATOS technique was tested in the specific case of a DiffServ IP network implementing the Assured Forwarding scheme. In this case, perceptually important packets were assigned to the AF service class. Network simulations demonstrated that the ATOS algorithm delivers significantly higher and more constant PSNR values than traditional non-adaptive approaches. Moreover, it leads to fairness and prompt reactivity to network congestions.

Table 1. Performance of classification techniques.

Test case	AF share	Byte loss rate	PSNR (dB) Avg.	PSNR (dB) St.Dev.
1. ATOS	30.8%	1.9%	34.8	1.3
2. Fixed-AF-share	34.6%	2.0%	32.5	4.2
3. 100% BE	0%	4.1%	30.8	3.9
4. Prescient BE	0%	0.0%	34.5	1.3

Table 2. Behavior of five ATOS sources for test case 1.

Source	AF share	Byte loss rate	PSNR (dB) Average	PSNR (dB) Stand. Dev.
1	60.1%	4.0%	33.9	1.9
2	30.8%	1.9%	34.8	1.3
3	30.5%	1.9%	34.7	1.3
4	28.6%	2.1%	34.5	1.7
5	29.8%	2.1%	34.6	1.4

6. REFERENCES

[1] A. Ortega and K. Ramchandran, "Rate-distortion Methods for Image and Video Compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, November 1998.

[2] R. Zhang, S. L. Regunathan, and K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, June 2000.

[3] P.A. Chou and Z. Miao, "Rate-Distortion Optimized Streaming of Packetized Media," *IEEE Transactions on Multimedia*, submitted 2001.

[4] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Joint Final Committee Draft (JFCD) of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)," *Doc. JVT-D157*, July 2002.

[5] D. Quaglia, J.C. De Martin, "Delivery of MPEG Video Streams with Constant Perceptual Quality of Service," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, Lausanne, Switzerland, August 2002, vol. 2, pp. 85–88.

[6] S.Y. Lee and A. Ortega, "Optimal Rate Control for Video Transmission over VBR Channels Based on a Hybrid MMAX/MMSE Criterion ," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, Lausanne, Switzerland, August 2002, vol. 2, pp. 93–98.

[7] J. Kim J. Shin and C. Kuo, "Quality-of-Service Mapping Mechanism for Packet Video in Differentiated Services Network," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 219–231, June 2001.

[8] E. Masala, D. Quaglia, and 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.

[9] UCB/LBNL/VINT, "Network Simulator – ns – version 2," *URL: http://www.isi.edu/nsnam/ns*, 1997.

[10] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group," *RFC 2597*, June 1999.