A NOVEL COST-DISTORTION OPTIMIZATION FRAMEWORK FOR VIDEO STREAMING OVER DIFFERENTIATED SERVICES NETWORKS

Fan Zhai, Carlos E. Luna, Yiftach Eisenberg,
Thrasyvoulos N. Pappas, Randall Berry, and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
E-mail: {fzhai, carlos, yeisenbe, pappas, rberry, aggk}@ece.northwestern.edu

ABSTRACT

This paper presents a novel framework for streaming video over a Differentiated Services (DiffServ) network that jointly considers video source coding, packet classification and error concealment within the scope of cost-distortion optimization. Our formulation incorporates the random network delay for each packet into the calculation of the probability of packet loss and manages the end-to-end packet delay by selecting the encoding parameters and packet priority. We formulate two approaches to evaluate the performance of the proposed framework: a minimum distortion approach and a minimum cost approach, in which the encoding mode and priority class for each packet are optimally selected so as to minimize the total distortion subject to cost constraints, or to minimize the total cost subject to end-to-end distortion constraints. Simulation results demonstrate the advantage of jointly adapting the source coding and packet classification.

1. INTRODUCTION

Continuous Media (CM) Internet applications, such as streaming video and videoconferencing, are gaining increased popularity. CM applications have more stringent quality of service (QoS) demands than traditional TCP-based applications such as web browsing, data file transfer, and electronic mail. Today's Internet, with a best-effort design, works well for "elastic" applications since they are not sensitive to delay, and reliable transmission can be realized through retransmission. In contrast, CM applications, such as real-time streaming, which have strict delay constraints, can suffer significantly.

When transmitting over unreliable networks, source coding should be adapted to dynamic network conditions. The approaches toward this task range from rate control and mode selection to forward error correction (e.g., [1-3]).

The incompatibility between the nature of the current Internet and the QoS requirements for CM applications has also led to proposed modifications of the Internet itself. One example is differentiated services (DiffServ or DS) approach standardized by the Internet engineering task force (IETF) [4]. DiffServ supports QoS by discriminately allocating resources to aggregated traffic flows according to service classes. In using DiffServ, the sender needs to choose the appropriate DS priority level for each packet. Different DS levels result in different QoS levels, such as different packet loss rate and network delay.

In this paper, we propose a novel framework for the joint adaptation of source coding and packet priority to maximize the system performance. Our formulation incorporates the random network delay for each packet into the calculation of the probability of packet loss and manages the end-to-end packet delay by selecting the encoding parameters and packet priority. The aim is to minimize the end-to-end distortion subject to cost constraints, or, alternatively, to minimize the overall cost given end-to-end distortion constraints. This work is an extension of the work in [5] which considered constant network delay.

In related studies, a distortion-based classification scheme was presented in [6] for video transmission over DiffServ networks with perceptually important macroblocks (MBs) grouped into premium packets. Error concealment was used at the decoder. However, this approach does not consider the selection of source coding parameters. Cost-distortion optimized multimedia streaming over DiffServ networks was also studied in [7] and [8]. Although the framework in [7,8] is very general, the selection of encoding parameters and error concealment are not considered.

The rest of this paper is organized as follows. The major components of the proposed framework are described in Section 2. In Section 3, we present the problem formulations and the solution algorithm. Simulation results and discussion are reported in Section 4. We draw conclusions in Section 5.

2. VIDEO STREAMING SYSTEM

Fig. 1 depicts the block diagram of the proposed video streaming system. On the sender's side, a raw bit-stream of live video is continuously fed into the video encoder, which generates a stream of encoded video packets. The controller assigns a coding parameter and a service class to each video packet, based on the QoS information associated with each service class, the error concealment strategy used at the decoder, and the fullness of the encoder buffer. The video packets are fed into a first-in-first-out (FIFO) encoder buffer before being transmitted over a DiffServ network. Some packets may be dropped in the network (due to congestion) or at the receiver (due to excessive delay). Packets that reach the decoder in time are buffered. The video decoder reads video packets from the decoder buffer and displays the resulting video frames in real-time. Lost packets are concealed at the decoder.

0-7803-7750-8/03/$17.00 ©2003 IEEE.

In the DiffServ network, each class may have different cost specified in the service level agreement (SLA) and different stochastic characteristics, such as packet loss probability and network delay [4]. By adjusting the prices for each service class, the network can influence the class a user selects. Typically, a higher priority service class has a higher cost but better QoS.

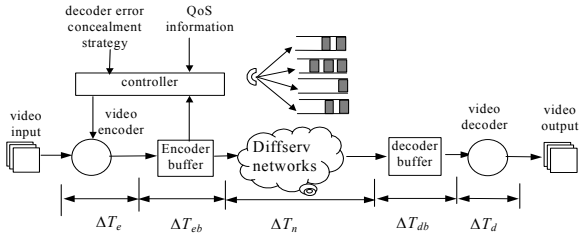

Fig. 1: Video streaming system block diagram

We consider the situation where the individual user's traffic is a small part of the overall traffic in the network, and thus has a negligible effect on the network's characteristics. Based on this assumption and estimates of the network state, transmission rate bounds for each class are obtained from a congestion controller, e.g., a TCP-friendly congestion controller [1], [7].

2.1. Channel Model

In this paper, the network is modeled as an independent time-invariant packet erasure channel with random delays, as in [7], [8]. Thus the packet loss probability is made up of two components: the packet loss probability in the network and the probability that the packet experiences excessive delay. For each service class $\pi \in \mathbf{\Pi}$, combining these two factors, the overall probability of packet loss is

$$p(\pi) = \varepsilon_\pi + (1-\varepsilon_\pi) P\{\Delta T_n(k) > \tau\}, \quad (1)$$

where ε_π is the packet loss probability for service class π, ΔT_n is the network delay, and τ is the maximum allowable network delay for a packet.

The proposed framework is general and not limited to any specific packet loss or packet delay model. All that is needed is a stochastic model of the packet losses and delays. Packet losses in the network can be modeled in various ways, e.g., a Bernoulli process, a 2-state or *k*-th order Markov chain [9]. Since the time distribution of packet arrivals often follows a self-similar law where the underlying distributions are heavily-tailed [10], the packet delay could be modeled, for example, by a shifted Gamma distribution with heavy tail [7], [8], [10].

2.2. Delay Components

In a video streaming system, the end-to-end delay, i.e., the time between when a frame is captured at the encoder and displayed at the decoder, should be constant, if the encoder and decoder operate at the same frame rate of *F* frames per second [4], [11]. As shown in Fig. 1, the end-to-end delay T of each packet can be decomposed into

$$T = \Delta T_e + \Delta T_{eb} + \Delta T_n + \Delta T_{db} + \Delta T_d, \quad (2)$$

where ΔT_e, ΔT_{eb}, ΔT_n, ΔT_{db}, and ΔT_d are the encoder delay, encoder buffer delay, network delay, decoder buffer delay and decoder delay for each packet, respectively. Let *M* be the number of packets in a video frame and *k* the packet index. Without loss of generality, we assume that the processing times for both encoding and decoding a packet are constants and equal to T_p=1/(*MP*). Based on this assumption, the maximum encoder buffer and network delay is $T_{\max} = T - (M+1)T_p$ [11].

Let *w*(*k*) be the waiting time in the encoder buffer for the *k*-th packet. From [12], *w*(*k*) can be recursively calculated by

$$w(k) = w(k-1) + \frac{B(\mu^{k-1})}{R(\pi^{k-1})} - T_p, \quad (3)$$

where $B(\mu^k)$ and $R(\pi^k)$ are the packet length in bits and the transmission rate in bits/sec for packet k using class π, respectively. The *k*-th packet has a particular class $\pi^k \in \mathbf{\Pi}$ and coding parameter $\mu^k \in \mathbf{Q}$, where $\mathbf{\Pi}$ and $\mathbf{Q}$ are the set of available service classes and source coding parameters, respectively. In this case, $\Delta T_{eb}(k)$ is equal to

$$\Delta T_{eb}(k) = w(k) + \frac{B(\mu^k)}{R(\pi^k)}. \quad (4)$$

From (2), to avoid decoder buffer underflow, i.e. $\Delta T_{db}(k) \geq 0$, the maximum allowable network delay for packet *k* is

$$\tau(k) = T_{\max} - \Delta T_{eb}(k) = T_{\max} - w(k) - \frac{B(\mu^k)}{R(\pi^k)}. \quad (5)$$

3. PROBLEM FORMULATION AND SOLUTION

3.1. Packetization and Expected Distortion

First, we assume that each row of MBs is coded as one packet, and every packet is independently decodable. The framework presented here can easily be extended to include other packetization schemes.

We measure the received video quality by the expected distortion at the receiver. The expectation is taken with respect to the probability of packet loss, which depends on the selected service class and the current encoder buffer state. The expected distortion for packet *k* can be written as

$$\mathrm{E}[D(k)] = (1-p(k))\,\mathrm{E}[D_r(k)] + p(k)\,\mathrm{E}[D_l(k)], \quad (6)$$

where $\mathrm{E}[D_r(k)]$ and $\mathrm{E}[D_l(k)]$ are the expected distortion if the packet is received and lost respectively. Due to channel losses, the reference frames at the decoder and the encoder may not be the same. Therefore, both $D_l(k)$ and $D_r(k)$ are random variables. The value of $\mathrm{E}[D_l(k)]$ depends on the specific error concealment scheme used at the decoder. In our simulations, the distortion measurement is based on exact per-pixel distortion calculations, which ensure accurate estimation of the overall end-to-end distortion [2-4].

3.2. Minimum Distortion and Minimum Cost Problems

The first problem we consider is to provide the best quality (minimum end-to-end distortion) for given cost and delay constraints. We refer to this as the "minimum distortion problem". This problem can be formulated as

$$\min_{\{\pi^k, \mu^k\}} \sum_{k=1}^{M} \mathrm{E}[D(\mu^k, \pi^k)] \quad (7.a)$$

$$\text{s.t. } \sum_{k=1}^{M} c(\pi^k) B(\mu^k) \leq C_0(i), \quad (7.b)$$

$$\sum_{k=1}^{M} B(\mu^k)/R(\pi^k) \leq T_0(i), \quad (7.c)$$

where $c(\pi^k)$ is the cost per bit for the *k*-th packet, and $C_0(i)$ is the total cost constraint for the *i*-th frame. The transmission delay constraint for the *i*-th frame, $T_0(i)$, can be obtained from a higher-level rate controller, and may vary from frame to frame. Assuming $T_{rc}(i)$ is the deadline assigned by the higher-level rate controller by which the *i*-th frame must leave the encoder buffer, $T_0(i)$ can be recursively obtained as

$$T_0(0) = T_{rc}(0)$$

$$T_0(i) = T_{rc}(i) + T_0(i-1) - \sum_{k=1}^{M} \frac{B_{i-1}(\mu^k)}{R_{i-1}(\pi^k)} \qquad i \geq 1. \qquad (8)$$

As an alternative to (7), we also consider a dual problem of minimizing the total cost subject to constraints on end-to-end distortion and transmission delay. This "minimum cost problem" reflects a case where a desired level of video quality must be maintained.

3.3. Proposed Solution

Next, we present an approach to solve the minimum distortion problem (7) based on Lagrangian relaxation and deterministic dynamic programming (DP). The minimum cost problem can be solved in the same fashion. With the use of Lagrange multipliers, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, the constrained problem (7) is converted into the unconstrained problem

$$\min_{\{\pi^k, \mu^k\}} \sum_{k=1}^{M} J(k) = \sum_{k=1}^{M} \{\mathbf{E}[D(k)] + \lambda_1 c(\pi^k) B(\mu^k) + \lambda_2 B(\mu^k)/R(\pi^k)\}. \qquad (9)$$

The solution of (7) can be obtained by choosing the correct Lagrange multipliers $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, within a convex hull approximation by solving (9). This can be accomplished by sub-gradient techniques [13]. In our simulations, we use a low complexity algorithm that exploits the structure of this problem as given in [14].

From (1), (6), and (9), the cost of each packet *J*(*k*) is a function of π^k, μ^k, $\tau(k)$ and $\mathrm{E}[D_l(k)]$. As shown in (3) and (5), $\tau(k)$ is a function of $w(k)$, and $w(k)$ is recursively calculated. That is, the cost of each packet depends not only on its own coding and priority decision but also on the decisions of all previous packets. In general, evaluating (9) requires an exhaustive search through all coding parameter and priority choices for each packet [15]. Instead as in [12], we approximate the solution to (9) by uniformly quantizing $w(k) \in [0, T_{\max}]$ into a set of N_W values.

As shown in (6), the error concealment scheme used at the decoder introduces dependencies between packets. For example, if temporal concealment based on the motion vectors of neighboring packet(s) is used, $\mathrm{E}[D_l(k)]$ depends on the coding parameter and priority of the neighboring packet(s). In this paper, in order to illustrate the concept of the proposed formulation and without loss of generality, we consider a simple error concealment scheme, i.e., the corrupted packet is replaced with the MBs from the previous frame at the same position. Based on this error concealment strategy, the cost of the *k*-th packet can be described as $J(k) = J(\mu^k, \pi^k, w(k))$. An example with a more complicated error concealment strategy can be found in [14]. This DP problem can be solved using a shortest path algorithm.

4. EXPERIMENTAL RESULTS

We consider transmission of the Foreman sequence in QCIF format encoded using the ITU H.263+ codec. The frame rate is 30 fps. Packet loss in the network is modeled by a Bernoulli process, i.e., each packet is independently lost with probability ε_π. For simplicity, packet delay in our simulations is modeled as a *shifted Gamma distribution* with parameters (n_π, α_π) and right shift γ_π [8]. The service class set is $\mathbf{\Pi} = \{1,2,3,4\}$, whose parameters are defined in Table 1. Each class has a different transmission rate and loss probability. We choose these parameters using a model-based TCP-friendly congestion controller as in [1]. The transmission rates from this model represent long-term average packet transmission rates for each service class. However, in the following, we use these as an approximation for the effective transmission rate for each service class. The quantizer set $\mathbf{Q}$ is given by {8,12,18,24} for INTRA mode, {4,6,8,10} for INTER mode, and SKIP mode. We set $T_{\max}$ = 333ms and N_W =300. This is a relatively small end-to-end delay for streaming applications, but may be appropriate for "real-time" streaming. Our main reason for considering this case is because larger $T_{\max}$ leads to a larger N_W and significantly increases the computational complexity.

probability of packet loss	{0.2, 0.1, 0.05, 0.001}
transmission rate (Kbps)	{210, 280, 350, 420}
cost (microcents per kilobits)	{25, 50, 75, 100}
γ (milliseconds)	{40, 30, 20, 10}
n	{2, 2, 2, 2}
α(1/milliseconds)	{1/40, 1/30, 1/20, 1/10}
Mean delay {milliseconds}	{120, 90, 60, 30}

Table 1: Parameters of four service classes

To evaluate the advantage of the proposed framework, we define a reference system where only one QoS class is available. In this reference system, source coding decisions are made to minimize the expected end-to-end distortion subject to the transmission delay constraint (7.c). The transmission delay constraint for the *i*-th frame, $T_0(i)$, is recursively obtained as in (8). Since our focus here is not on rate control, we set $T_{rc}(i) = 1/F$ seconds for all frames.

First, we compare the solution to the proposed minimum distortion problem with multiple QoS classes (7) to a reference system with only one QoS class. The cost constraint $C_0(i)$ is obtained from the reference system. We consider two reference systems, which only use service class π_2 and π_3, respectively. For each reference system, a corresponding proposed system is simulated using the same cost and delay constraints. As shown in Fig. 2, the proposed minimum distortion approach outperforms the corresponding reference systems by 0.75dB and 0.93dB average PSNR, respectively.

Next, we compare the proposed minimum cost approach with multiple QoS classes to the single-QoS-class reference system. Here, the distortion constraint $D_0(i)$ is the distortion achieved by the reference system. As shown in Fig. 3, the proposed system outperforms the corresponding reference system with a 16% and 23% average cost saving, respectively.

Note that different choices of the parameter set will result in different gains for the proposed system.

The gains illustrated in the above experiments are mainly due to the use of multiple QoS classes in the proposed system. Specifically, for packets that are hard to conceal but easily encoded, the encoder tends to use coarser quantizers and higher QoS. These packets usually correspond to high motion areas that can be predicted from the previous frame. If such a packet is lost, it may be hard to conceal since the associated motion vector is lost. Similarly, for packets that are hard to encode, the encoder may want to use a finer quantizer to reduce the quantization error, and send them with higher QoS as well. These packets are hard to predict from the previous frame and therefore difficult to conceal. In this case, the associated cost will be relatively high. On the other hand, packets that are easier to conceal can be sent using a lower QoS class with relatively low cost. In this way, the encoder can select different quantizers and QoS classes in order to optimally balance the received video quality and the total cost.

5. CONCLUSIONS

We have presented a novel cost-distortion optimized video streaming framework for DiffServ networks. The optimization is achieved by jointly considering error resilience, packet classification and error concealment. End-to-end packet delay is managed through provisioning based on the fullness of the encoder buffer. By jointly adapting the source coding and packet classification, the system optimally allocates resources among packets, giving more protection to the most important parts of the bitstream, thus achieving the maximum performance.

6. REFERENCES

[1] Q. Zhang, W. Zhu, and Y-Q Zhang, "Resource allocation for multimedia streaming over the Internet", *IEEE Trans on Multimedia*, vol. 3, pp. 339-355, Sept. 2001.

[2] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. On Circuits system Video Technology,* vol. 12, no. 6, pp. 411-424, June 2002.

[3] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol.18, pp.966-976, June 2000.

[4] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the differentiated services field (DS field) in the Ipv4 and Ipv6 headers," RFC 2474, IETF, Dec. 1998.

[5] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and packet marking for video transmission over Diffserv networks," *Proc. IWDC*, Sept. 2002.

[6] E. Masala, D. Quaglia, and J. C. De Martin, "Adaptive picture slicing for distortion-based classification of packets," *IEEE Workshop on Multimedia Signal Processing*, 2001.

[7] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media", *IEEE Trans. on Multimedia*, 2001. Submitted.

[8] A. Sehgal and P. A. Chou, "Cost-distortion optimized streaming media over Diffserv networks," *IEEE ICME*, 2002.

[9] M. Yajnik, and *et al*., "Measurement and modeling of the temporal dependence in packet loss," UMASS CMPSCI, 98-78, 1998.

[10] G. Hooghiemstra and P. Van Mieghem, "Delay distributions on fixed Internet paths," Delft University of Technology, report 20011020, 2001.

[11] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 756-773, May 1999.

[12] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE J. Select. Areas Commn.*, 2003*,* To appear.

[13] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[14] F. Zhai, C.E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A.K. Katsaggelos, "Joint source coding and packet classification for video streaming over differentiated services networks," *IEEE Trans. Multimedia*, special issue on streaming media, 2003, submitted.

[15] G. M. Schuster and A. K. Katsaggelos, "Rate-distortion based video compression: optimal video frame compression and object boundary encoding," *Kluwer Academic Publishers*, 1997.

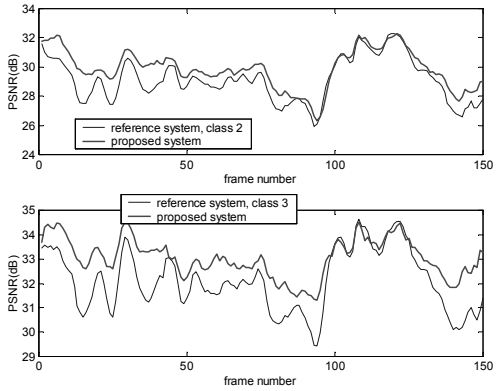

Fig. 2: Comparison of minimum distortion approach with reference system

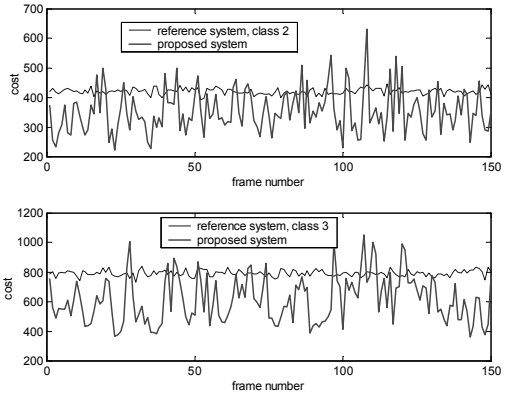

Fig. 3: Comparison of minimum cost approach with reference system