OPTIMIZED VIDEO STREAMING OVER LOSSY NETWORKS WITH REAL-TIME ESTIMATION OF END-TO-END DISTORTION

Rui Zhang, Shankar L. Regunathan and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Abstract

This paper is concerned with adaptive and robust video streaming over lossy networks. A general system is described which is compatible with virtually all application settings. The delivery procedure is optimized within a rate-distortion framework with the aid of an efficient real-time end-to-end distortion estimation algorithm. The estimate fully accounts for the effects of (prior) quantizaton, packet loss and error propagation, as well as error concealment. It features high accuracy and low complexity, and requires a small amount of pre-computed side information. It is shown that various error-resilient schemes can be optimized within the proposed framework. In particular, unequal error protection through forward error control is employed in the simulations to demonstrate the achievable performance gains.

1. INTRODUCTION

Due to the explosive growth of the Internet and wireless networks, video streaming has attracted increasing attention. However, the unreliable packet delivery through these lossy networks poses robustness requirements on the design of the streaming system, to ensure that the perceptual quality varies gracefully during periods of fluctuations in network quality of services (QoS).

While standard source-channel coding algorithms can be used to adaptively optimize the encoding/delivery of live video, they are incompatible with applications that stream pre-compressed video. The major difficulty is due to the fact that network conditions are unknown during the compression stage. It is only at the time of delivery that adaptation to the channel conditions can be performed [1].

In this work, we propose an adaptive and robust video streaming system which is optimized within a rate-distortion (RD) framework. Its generality ensures applicability to many application settings with various error-resilience schemes. Recent related work has been reported in [2] [3]. It is well recognized that the ideal delivery strategy should adapt to the actual bandwidth and packet loss statistics, so that the expected end-to-end distortion is minimized. This work is motivated by the absence of an efficient and accurate method to estimate the end-to-end distortion – a challenging but crucial task. The proposed system employs an efficient estimation algorithm, which accounts for the effects of (prior) quantizaton, packet loss and error propagation, as well as error concealment. It features high accuracy and low complexity, and requires only a small amount of side information.

The paper is organized as follows: Section 2 describes the general RD-optimized delivery system for video streaming. The algorithm for end-to-end distortion estimation is briefly described in Section 3. Section 4 integrates this estimate into the RD framework for optimizing various error resilient strategies. An example of a practical system is constructed in Section 5 where unequal error protection is provided by forward error control. Simulation results evaluate the performance in the context of transmission over packet-erasure channels and bit-error channels, and demonstrate substantial performance gains.

2. RD-BASED ADAPTIVE VIDEO STREAMING

Figure 1 illustrates the general framework for adaptive delivery of pre-compressed video through RD optimization. For concreteness, let us consider a "video on demand" application. The raw video content is compressed offline without knowledge of the eventual network status. The compressed bitstream is stored on the disk, together with some side information. The server employs a given set of adaptation tools for combating channel loss. Note that practical restrictions on the server complexity often restrict the choice to simple transport-level tools, such as forward error control (FEC) or automatic retransmission on request (ARQ).

This work was supported in part by the NSF under grants EIA-9986057 and EIA-0080134, the University of California MICRO program, Conexant Systems, Inc., Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, and Qualcomm, Inc.

R. Zhang is currently with Stanford University, Stanford, CA 94305.

S. L. Regunathan is currently with Microsoft Corp., Redmond, WA 98052.

0-7803-7304-9/02/$17.00 C2002 IEEE

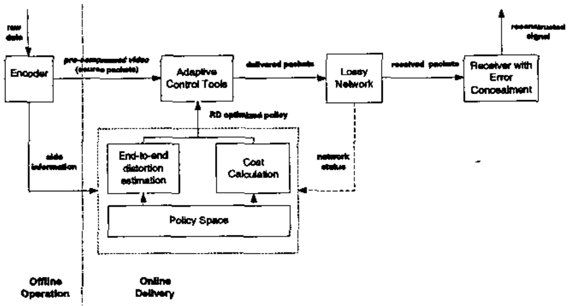


Figure 1: RD-optimized delivery for pre-compressed video

Each adaptation scheme is associated with a policy space consisting of all the possible operational choices. For example, a specific policy may determine how many parity check packets to send in the case of FEC, or whether to retransmit the particular packet in the case of ARQ. Upon a delivery request, the server seeks to identify the policy that minimizes the expected end-to-end distortion under the current network status. the packets are then generated using the selected adaptation policy and are sent through the network. The receiver receives the packets and decodes the information. The decoder usually approximates the information in missing packets by error concealment.

Note the generality of this framework. A large variety of video compression, online adaptation, and error concealment techniques are covered. Moreover, it is suitable for applications other than "video on demand", including the following examples: 1) Media distribution through heterogeneous networks involves a number of intermediate servers where the employment of transcoding can potentially improve the overall quality of service. A particularly relevant example involves a mixed wired/wireless transmission system with channel adaptation to enhance error resilience performed at the gateway prior to a last wireless hop [4]; 2) In multicast transmission of pre-compressed video, there are multiple receivers with differing channel conditions and processing power. Distortion estimation is, further, essential for RD optimized subscription/desubscription in receiver-driven applications [3]. In insisting on its generality we wish to emphasize that the end-to-end RD solution is fundamental in all of these applications, which hence stand to benefit considerably from optimal estimation.

3. END-TO-END DISTORTION ESTIMATION

End-to-end distortion estimation is fundamental to robust video streaming. It has also been recognized as a highly non-trivial problem. On the one hand, the original video is not accessible during delivery. This represents a major difficulty in estimating the end-to-end distortion, which quantifies the difference between the original signal and the decoder reconstructed signal (after compression, loss and error concealment). Further, the computation is complicated by many inter-related factors, including (prior) quantization, effective packet loss statistics which is a function of the network condition and the error resilience strategy, and error concealment. Moreover, the use of inter-frame prediction in video coders results in spatial and temporal error propagation, which introduces additional inter-dependencies between packets.

In this work we adopt an approach we recently developed [5], whose objective is to fully account for all the above mentioned effects. The algorithm features high accuracy and requires a small amount of side information. It is of low complexity, and is applicable to virtually all coding techniques, including standard (predictive) video coders. A brief review of the algorithm is given below.

Assume N source packets per independent group of packets (GOP). The distortion for all packets in one GOP must be jointly calculated due to inter-packet dependencies. Let p_i denote the effective packet loss rate (PLR) of packet i. The PLR vector for the entire GOP is then given by $\mathcal{P} = \{p_0, p_1, ..., p_i, ..., p_{N-1}\}$. Let binary random variable b_i denote the status of packet i. The delivery status of the entire GOP is denoted by the binary random vector $\mathcal{B}$. A particular GOP transmission event is represented by the kth event vector: $\mathcal{B}^{(k)} = \{b_0^{(k)}, b_1^{(k)}, ..., b_i^{(k)}, ..., b_{N-1}^{(k)}\}$, and occurs with probability $p^{(k)} = \prod_{i=0}^{N-1}(1-p_i)^{(1-b_i^{(k)})} p_i^{b_i^{(k)}}$. The corresponding end-to-end distortion, $D^{(k)}$, can be *exactly computed* during compression. The expected end-to-end distortion during delivery is given by:

$$E_{\mathcal{P}}\{D\} = \sum_{k=0}^{2^N-1} p^{(k)} D^{(k)}. \tag{1}$$

Note that this calculation is *exact*. It averages over all possible error events, and takes into account the effects of compression, loss, error propagation and error concealment. However, it is highly demanding in side information storage and introduces high complexity in the delivery optimization procedure. For practical usefulness, we derive a simple approximation which involves a slight loss in accuracy but offers substantial advantages in storage and complexity. The expected distortion of (1) is approximated via a first-order Taylor expansion about a reference PLR vector $\bar{\mathcal{P}}$:

$$E_{\mathcal{P}}\{D\} \approx E_{\bar{\mathcal{P}}}\{D\} + \sum_{i=0}^{N-1} \gamma_i (p_i - \bar{p}_i), \tag{2}$$

where

$$\gamma_i = \frac{\partial E\{D_{\mathcal{P}}\}}{\partial p_i}|_{\mathcal{P}=\bar{\mathcal{P}}} \tag{3}$$

is the partial derivative with respect to the PLR of packet i. Both $E_{\bar{\mathcal{P}}}\{D\}$ and γ_i are easily pre-computed. We refer to this estimate as the first-order distortion estimate (FODE). Note that one may use multiple reference PLR vectors and thereby trade complexity for better estimation.

4. OPTIMIZATION OF ERROR-RESILIENT DELIVERY SCHEMES

In this section, an optimal delivery scheme within an RD framework is proposed as a general solution for various video streaming scenarios. The FODE algorithm is integrated within the RD framework to optimize error-resilient streaming schemes during delivery. This provides the optimal linear solution at the cost of only modest complexity.

Without implied loss of generality, let any adaptive error-resilience scheme be described in terms of a set of policy choices,

$$\pi \in \{\pi^{(0)}, \pi^{(1)}, ..., \pi^{(S)}\}, \tag{4}$$

for each packet. Note that virtually all delivery strategies are naturally covered by this paradigm. The policies could involve whether/when to retransmit the current packet, or the strength of the error correction code. Clearly, the effective loss probability for each packet is a function of the policy choice (and the given channel loss rate p_{ch}). The cost of the policy choice c is usually measured by the total number of bits needed to send the packet, which of course also depends on the delivery policy:

$$p = p(\pi), c = c(\pi). \tag{5}$$

The policy vector for a source GOP specifies the policy decision per packet:

$$\Pi = \{\pi_0, ..., \pi_i, ..., \pi_{N-1}\}. \tag{6}$$

Correspondingly, the effective PLR vector and the cost vector of the GOP are:

$$\begin{aligned} \mathcal{P}(\Pi) &= \{p_0(\pi_0), ..., p_i(\pi_i), ..., p_{N-1}(\pi_{N-1})\}, \\ \mathcal{C}(\Pi) &= \{c_0(\pi_0), ..., c_i(\pi_i), ..., c_{N-1}(\pi_{N-1})\}. \end{aligned} \tag{7}$$

Using FODE, the expected end-to-end distortion for a GOP can be estimated as:

$$E_{\mathcal{P}(\Pi)}\{D\} = E_{\bar{\mathcal{P}}}\{D\} + \sum_{i=0}^{N-1} \gamma_i (p_i(\pi_i) - \bar{p}_i), \tag{8}$$

while the cost is simply the sum of the packet costs:

$$\mathcal{C}(\Pi) = \sum_{i=0}^{N-1} c_i(\pi_i). \tag{9}$$

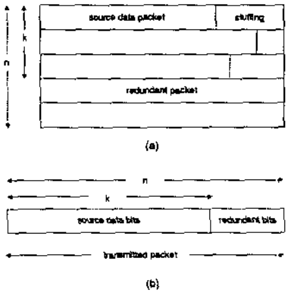


Figure 2: Generation of FEC codes. (a) across packet, (b) within packet.

The optimal adaptive delivery scheme should then select the policy that minimizes the expected distortion $E_{\mathcal{P}(\Pi)}\{D\}$ while satisfying a constraint on the cost $\mathcal{C}(\Pi)$:

$$\min_{\Pi} E_{\mathcal{P}(\Pi)}\{D\}, \quad \text{subject to } \mathcal{C}(\Pi) \leq C_c. \tag{10}$$

This problem is conveniently recast as an unconstrained minimization of the Lagrangian:

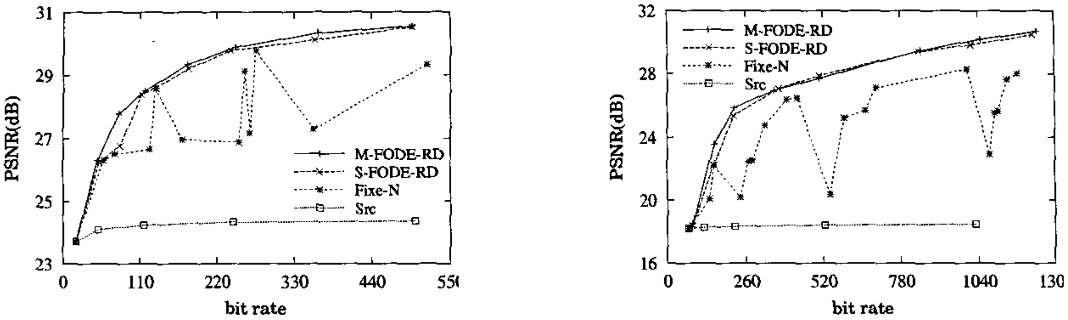
$$\begin{aligned} & E_{\mathcal{P}(\Pi)}\{D\} + \lambda \mathcal{C}(\Pi) \\ = \; & E_{\bar{\mathcal{P}}}\{D\} + \sum_{i=0}^{N-1} (\gamma_i (p_i(\pi_i) - \bar{p}_i) + \lambda c_i(\pi_i)). \end{aligned} \tag{11}$$

5. SIMULATION RESULTS

The proposed framework is applicable to diverse applications with various error-resilient delivery schemes. For illustration of its performance, we set up a delivery system of layered coding with unequal transport prioritization through forward error control. We consider both the packet-erasure channel and bit-error channel, and propose cross-packet FEC codes and within-packet FEC codes for them, respectively.

In packet-erasure channels, such as the best-effort IP networks, the entire packet is either received perfectly intact and on time, or considered as "erased" if the intermediate routers discard it or if it arrives too late. In this case, we apply the FEC code across packets and generate redundant packets for protection, as shown in Figure 2(a). For concreteness, we adopted systematic Reed-Solomon (RS) codes. In the case of a bit-error channel, e.g., a wireless channel, the packets are corrupted mainly due to random bit errors. Thus, the packet loss rate can be reduced by inserting redundant bits into each packet, as shown in Figure 2(b). Here we used rate-compatible punctured convolutional (RCPC) codes.

In the simulations, we generated a five-layer bitstream for the QCIF sequence *carphone*. Three online delivery



(a) (b)

Figure 3: PSNR vs. total bit rate for different delivery schemes. QCIF sequence "carphone", 10 fps, 5-layer bitstream at 16/64/112/240/496kbps. (a) Packet-erasure channel, (b) Bit-error channel.

schemes are compared. The first is the RD-optimized scheme using our multi-FODE model (M-FODE-RD), where multiple reference PLR vectors are used. The second uses only the single-FODE model (S-FODE-RD), where one reference PLR vector (the all-zero vector) is used. Both of them dynamically choose the best error protection code (n, k), from a set of value n for a fixed k, to minimize the RD cost for packets in each layer. The third scheme uses fixed unequal error protection schemes for each layer, with more protection given to lower layers, through RS or RCPC codes (fixed-N). Note that while the first two schemes can adapt to any rate constraint, the fixed-N scheme can only operate at prespecified rates, which may result in poor bit allocation at some channel conditions. The performance of unprotected transmission of the source bitstream (Src) is also presented for reference.

The three bitstreams generated by these schemes go through the same time-varying channels with PLR in the range of $1\% \sim 20\%$ for the packet-erasure channel and BER in the range of $2 \times 10^{-6} \sim 5 \times 10^{-3}$ for the bit-error channel. Figure 3 shows the decoder distortion for each of them under different bit rates. Figure 3 (a) and (b) give the results for cross-packet protection in packet-erasure channel and within-packet protection in the bit-error channel, respectively. The results illustrate that FODE-RD schemes achieve substantial gains and enhanced flexibility relative to the fixed-N scheme. They provide more graceful degradation as the network bandwidth decreases. Moreover, we note that in these simulations there is only a small performance difference between the multi-FODE-RD and the single-FODE-RD schemes. This suggests that in practical application, the simple single-FODE model, which only requires minimal side information, may be sufficient to ensure good performance.

6. CONCLUSION

We proposed an RD-optimized streaming video system. An efficient algorithm for estimating the overall end-to-end distortion is integrated within the RD framework for the optimal streaming decisions. The framework is general and covers most application settings. The method involves low storage and computational complexity, and provides graceful degradation during QoS fluctuations.

7. REFERENCES

[1] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman,and Y. A. Reznik, "Video coding for streaming media delivery on the Internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, p. 269-81. Mar. 2001.

[2] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," submitted to *IEEE Transactions on Multimedia*, Feb. 2001.

[3] W. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, p. 373-86, vol. 11, no. 3, Mar. 2001.

[4] G. Reyes, A. R. Reibman, S. Chang, and J. C. I. Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE Journal of Selected Areas in Communications, special issue on Error-Resilient Image and Video Transmission*, vol. 18, no. 6, p. 1063-74, June 2000.

[5] R. Zhang, S. L. Regunathan and K. Rose, "End-to-end distortion estimation for RD-based robust delivery of pre-compressed video," *35rd Asilomar Conference on Signals, Systems, and Computers*, Oct. 2001, Pacific Grove, CA, USA.