OPTIMIZED VIDEO STREAMING FOR NETWORKS WITH VARYING DELAY

Susie Wee, Wai-tian Tan, John Apostolopoulos

Streaming Media Systems Group
Hewlett-Packard Laboratories, USA

Minoru Etoh

Multimedia Signal Processing Laboratory
NTT DoCoMo, Yokosuka, Japan

ABSTRACT

This paper presents a method for distortion-optimized streaming of predictively coded video over packet networks with varying delay. In networks with significant delay variations, coded video frames can arrive late at the decoder and miss their respective display deadlines. Furthermore, due to predictive coding, a late frame can also prevent a number of subsequent frames from being displayed properly, where the number of affected frames or degree of distortion depends on the particular coding dependencies of the late frame. In this paper, we present an optimized video streaming strategy based on frame reordering for networks with significant delay variations. This streaming strategy minimizes distortion by exploiting the fact that different late frames result in different degrees of distortion. We model the router-induced delay in a wired network with an analytical PDF and we model the link-layer retransmission delay of a wireless network with the 3GPP specification for WCDMA radio link control. We compute the distortion for different frame reorderings using the network delay models and a source model that accounts for the prediction dependencies of predictively coded video. Our optimized streaming strategies are shown to reduce the number of late frames by 14 to 23% for the situations examined.

1. INTRODUCTION

Streaming media systems transmit media such as video and audio over wired and wireless networks afflicted with packet loss and delay. This has motivated advancements in a number of research areas, including rate-distortion (RD) techniques for video compression [1], error-resilient video delivery for lossy packet networks [2, 3, 4], and media networking system design coupling media delivery and networking [5, 6]. However, relatively little prior work has explicitly addressed the problem of optimized streaming for network channels with varying delay. A framework for RD-optimal streaming for network channels with varying loss and delay was presented in [7]; this provided an analytical treatment and presented experimental results for audio streaming.

A number of considerations arise when developing streaming strategies for video because of the use of predictive coding in most video coders. Predictive coding introduces temporal dependencies in the coded data that provide improved compression, but also result in error propagation in the event of packet loss or late arrival. The degree of distortion from a lost or late packet depends on the specific coding dependencies of the affected frame. In addition, predictive coding leads to different delay requirements for different types of coded video data – while some coded frames need to be received in time for their own display, others are needed earlier to decode other frames. Furthermore, the delay variation over a network can be quite large, causing packets to arrive late (after their required playback time). However, even if a coded frame arrives after its playback time, it can still be useful for decoding subsequent frames.

The fact that coded video frames have different delay requirements and different impact on overall video quality suggests that an improved streaming strategy could involve reordering the coded video frames so that more important and delay-sensitive data is transmitted earlier than others. In this paper, we develop a streaming strategy based on frame reordering that allows us to improve the transmission of predictively coded video over network channels with significant delay variations. We optimize delivery by improving the probability of on-time delivery for more important media packets. We show that in this manner, the expected distortion can be reduced under the same network conditions.

This paper continues by discussing network delay models that capture the delay variations of wired and wireless networks. A media source model for predictively coded video is briefly reviewed. The streaming optimization formulation is then developed using the network delay and source models. Optimized streaming methods are determined by minimizing the expected distortion. Finally, experimental results are provided which show that streaming strategies based on transmission frame reordering reduce distortion for networks with varying delay.

2. NETWORK DELAY MODEL

Wired Network Paths: In wired networks, network paths consist of a number of links connected with network nodes. Packets sent over a single network path experience varying end-to-end delays [8]. Following [9, 10], we model this delay in three parts, the propagation delay of the packet travelling over the links, the queue delay incurred by the routers, and the retransmission delay due to packets lost to network congestion. The propagation delay is modelled by a constant κ. Each node is a network router that incurs delay due to the queue length at that node. The total delay due to n routers can be modeled as an M/M/1 queue. Thus, the PDF of the total queue delay τ is given by:

$$p(\tau) = \frac{\alpha}{\Gamma(n)}(\alpha\tau)^{n-1}e^{-\alpha\tau} \qquad (1)$$

where

$$\Gamma(n+1) = \int_0^\infty y^n e^{-y} dy \qquad n > -1 \qquad (2)$$

This has a mean $\mu = n/\alpha$ and variance $\sigma^2 = n/\alpha^2$. Note that for integer values of n, $\Gamma(n+1) = n!$ and the Gamma PDF becomes an Erlang PDF. Figure 1 plots the Gamma PDFs for $n = 5$ nodes with variances $\sigma^2 = .1$ and $\sigma^2 = .05$.

Wireless Network Links: Wireless channels are afflicted with significant delay variations due to factors such as link-layer retransmissions in radio access networks. We derived a network delay model for a WCDMA channel based on the 3GPP Radio Link

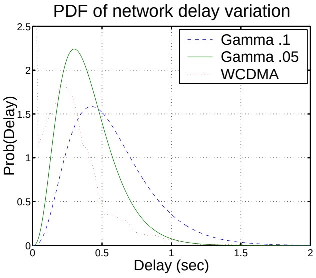

Figure 1: Model PDFs for delay variation (without offset for propagation delay): (1) Gamma PDF with $\sigma^2 = .1$, (2) Gamma PDF with $\sigma^2 = .05$ sec, and (3) WCDMA channel model (3GPP RLC).

Control (RLC) Specification [11] in acknowledged mode. According to the above model, the normalized histogram of the packet delay is shown in Figure 1, assuming a channel with 10% packet loss, 160 msec round trip time, and 10 maximum retransmissions.

3. MEDIA SOURCE MODEL

This section describes the characteristics of the media we transmit over the network path. Media can be video or audio stored on a server. However, because of the additional challenges associated with the prediction dependencies and temporal reordering of coded video, we focus on video in the remainder of this discussion.

3.1. Streaming Media and Video Coding Properties

Streaming media performance can be characterized by a number of parameters, including the start time for the session, and the transmit time, receive time, and playback time of each coded video frame. While our analysis can be applied to many transmission scenarios, in our experiments we assume that the sender streams the coded video at a constant framerate beginning at the session start time. Since each coded video frame has a different amount of data, this requires variable bitrate (VBR) transmission with an average bitrate corresponding to that of the coded media stream. We assume that each coded video frame may be afflicted by additional delay incurred by the channel. The playback time of the first frame begins at delay τ_{preroll} from the session start time, and each subsequent frame is spaced apart by $\Delta T = 1/F$, where F is the video frame rate. Note that longer preroll times (with larger buffer sizes) can reduce the timing or delay sensitivity of the system.

Video coders such as MPEG code frames as I, P, and B frames which have different compression efficiencies, prediction dependencies, and delay constraints. In general, I, P, and B frames have increasing coding efficiency and thus decreasing coded frame sizes. The sequence of frame types completely determines the prediction dependencies among frames and the number of subsequent frames that depend on each specific frame. We use this number as a measure of the relative importance of each coded frame.

Figure 2 shows the transmission requirements of a video sequence coded with I, P, and B frames. If this sequence is transmitted in coding order at regular frame intervals, then different frames have different lengths of time to their required playback. Furthermore, because of the prediction dependencies of the coded frames, while some I and P frames do not have to be displayed until their playback time, they may need to arrive at the decoder earlier so that other frames can be decoded/displayed at their appropriate time.

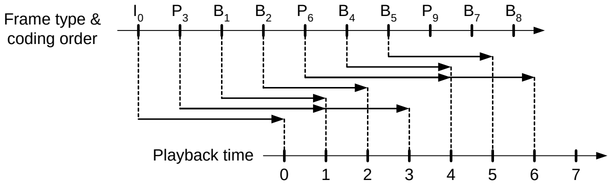

Figure 2: Playback delay: Different frames have different transmission time requirements. All coded frames must be received before their playback time, and some must be received earlier to decode other frames. Even if a frame is received late, it may still be useful for correctly decoding subsequent frames.

3.2. Effects of delay and distortion model

When coded video is streamed to a client over a channel with varying delay characteristics, the quality of the reconstructed video will depend on the time of the received coded video frames in relation to their playback time. If a coded frame is received late, obviously it can not be decoded in time to be displayed. However, a late coded frame is still useful if subsequent frames depend on it because of temporal prediction. For this reason, the quality of reconstructed video depends on (1) whether each coded frame is received before its playback time, and (2) the actual delay as it may affect other frames through error propagation. Note that coded video frames have different delivery requirements and error propagation effects depending on the frame type and GOP structure.

We measure distortion by counting the number of frames that can not be displayed at their appropriate playback time. We assume that frames are decoded as soon as they are received. We also assume that a frame can be correctly decoded only if all of the frames in its prediction tree have also been received.

4. STREAMING OPTIMIZATION FORMULATION

Due to dependence in video frames, the *distortion* associated with the late arrival of different frames is different. By optimizing the transmission schedule of the coded frames so that more important frames are transmitted earlier at the expense of less important frames, the overall distortion at the receiver can be minimized.

In this paper, the distortion we attempt to minimize is the expected number of late video frames. A frame is late when the frame itself, or one of the frames it depends upon, is late. Specifically, let $\mathcal{T}(t)$ denote the cumulative distribution function of the transmission delay. If we denote the transmission time and display time of a frame p by $t_p^{(T)}$ and $t_p^{(D)}$, respectively, the probability that p will be late, λ_p, is given by:

$$\lambda_p = 1 - \mathcal{T}(t_p^{(D)} - t_p^{(T)}) \prod_{q \in S_p} \mathcal{T}(t_p^{(D)} - t_q^{(T)}) \qquad (3)$$

where S_p is the set of frames that frame p depends on. This conveys the fact that for a frame to be decoded and displayed properly, both it and all the frames in its prediction dependency tree must arrive before its required display time. This is shown in Figure 3. We assume the decoder is intelligent in that it is not necessary that the frames in S_p are displayed on time in order for frame p to be displayed on time. Instead, the decoder can make use of late dependent frames as long as they all arrive before the display time of frame p. Thus, for the transmission of a set of frames G, the

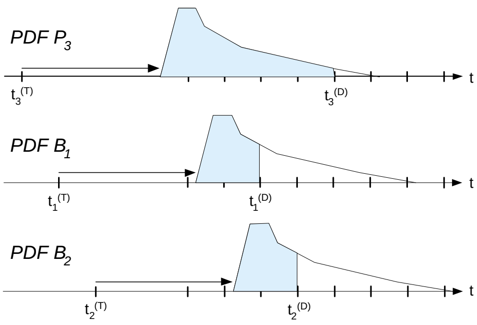

Figure 3: Simplified PDF of the arrival time for each transmitted coded frame: The shaded area represents the probability that the coded frame will arrive before its display time. For a frame to be decoded/displayed properly, all the coded frames that it depends upon must also arrive before its display time.

distortion D_G is given by:

$$D_G = \sum_{p \in G} \lambda_p \qquad (4)$$

In our optimization, the delay characteristics, $\mathcal{T}()$ and the display times $t^{(D)}$ are fixed, and the optimization is performed to identify the optimal transmission schedule corresponding to a set of $t^{(T)}$. Specifically, we consider the case when the video source is made up of a periodic GOP structure, and we construct our transmission schedule using *elementary operations*. An elementary operation is illustrated in Fig. 4, where when the transmission of a frame p in a GOP i is moved ahead by k time slots into GOP $i-1$, (1) k frames are moved back one time slot, and (2) the same operation is repeated in all other GOPs. Using only elementary operations to construct transmission schedules has two advantages. First, an elementary operation ensures that for each GOP, the net movement of all the frames is zero, so that there is always a cost in moving the transmission time of a frame early. Second, the resulting transmission schedule is periodic, so that expected distortion (Equation 4) can be easily computed using results of a single GOP.

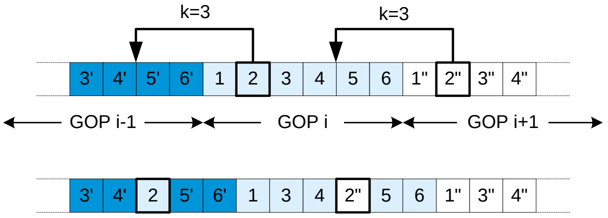

Figure 4: Elementary reordering operation.

In summary, the optimization problem is to find the optimal transmission schedule $\mathbf{s}^*$ in a GOP G that minimizes overall distortion:

$$\mathbf{s}^* = \min_{\mathbf{s}} \sum_{p \in G} \lambda_p(\mathbf{s}) \qquad (5)$$

where **s** is constructed using elementary operations. To limit the search space of potential transmission schedules, we restrict the net movement of any single frame from its original bitstream order by a threshold, refered to as the *advancement range*.

5. EXPERIMENTAL RESULTS

These experiments examine sequences with 13-frame GOPs with a coding order structure of IPBBPBBPBBPBB. We consider advancing the I and P frames, but not the B frames. The maximum advancement range is 10 frames, and the preroll is 1 sec. We assume that the total end-to-end delay is represented by the three different delay models shown in Figure 1: two Gamma PDFs with $\sigma^2 = .1$ sec and $\sigma^2 = .05$ sec, corresponding to long and short delay spreads with $\mathcal{T}(1 \text{ sec}) = .92$ and $\mathcal{T}(1 \text{ sec}) = .99$, respectively; and one WCDMA PDF derived from the 3GPP RLC described in Section 2. In the conventional case of no reordering, this corresponds to expected number of late frames of 2.39, .29, and .44 frames/GOP, respectively. Clearly, the smaller σ^2 corresponds to a significantly higher probability of reception within the preroll interval, which leads to a significantly smaller expected number of late frames. These three situations present very different delay/late arrival characteristics, which lead to different optimized results.

Performance Change when Only I Frames are Advanced Figure 6 examines the reduction in expected number of late frames as we vary the advancement of the I frame in each GOP. Advancing the I frame in the Gamma distribution with $\sigma^2 = .1$ generally reduces the probability of late frames, however the probability of late frames does not monotonically decrease with increasing advancement of the I frame. Specifically, the distortion (which relates to the probability of late frames) increases for advancements of 6 and 9 as these advancements lead to delaying P frames – while these advancements increase the probability of the I frame arriving in time, this is offset by the decreased probability of the P frames arriving in time. For the Gamma distribution with $\sigma^2 = .05$ the tradeoff is even more pronounced, as advancing the I frame by more than 2 frames has little positive effect, and advancing by more than 5 frames has a decidedly negative effect. For the WCDMA channel, a 5 frame advancement minimizes the expected number of late frames.

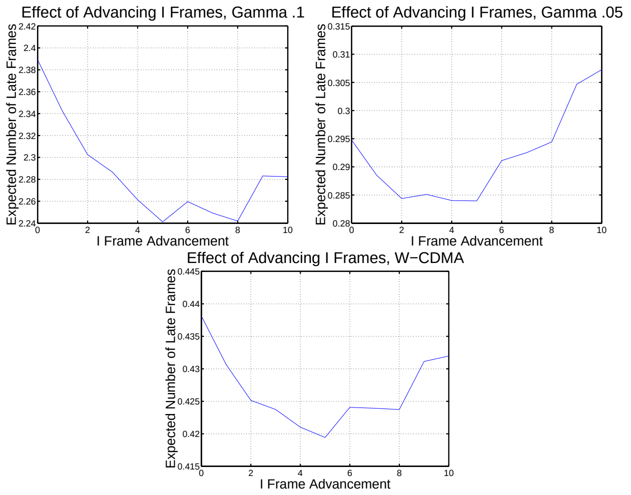

Figure 6: Expected number of late frames when only I frames are advanced: Results for Gamma PDF with $\sigma^2 = .1$ (left) and $\sigma^2 = .05$ (right), and WCDMA PDF derived from 3GPP RLC (bottom).

Performance Change as Both I and P Frames are Advanced We now examine the improvements when both the I and P frames can be advanced. Specifically, we consider five cases where we advance (1) only the I frame, (2) only the I and first P, (3) only the

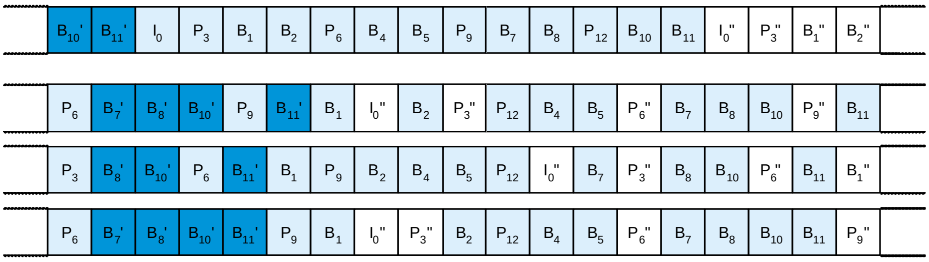
B_{10}' B_{11}' I_0 P_3 B_1 B_2 P_6 B_4 B_5 P_9 B_7 B_8 P_{12} B_{10} B_{11} I_0'' P_3'' B_1'' B_2''
P_6 B_7' B_8' B_{10}' P_9 B_{11}' B_1 I_0'' B_2 P_3'' P_{12} B_4 B_5 P_6'' B_7 B_8 B_{10} P_9'' B_{11}
P_3 B_8' B_{10}' P_6 B_{11}' B_1 P_9 B_2 B_4 B_5 P_{12} I_0'' B_7 P_3'' B_8 B_{10} P_6'' B_{11} B_1''
P_6 B_7' B_8' B_{10}' B_{11}' P_9 B_1 I_0'' P_3'' B_2 P_{12} B_4 B_5 P_6'' B_7 B_8 B_{10} B_{11} P_9''

Figure 5: Optimal streaming strategy by frame reordering for 13-frame GOP. Top to bottom: Original coded GOP structure, and optimized streaming strategies for (1) Gamma distribution with long delay spread, $\sigma^2 = .1$sec, (2) Gamma distribution with short delay spread, $\sigma^2 = .05$sec, and (3) WCDMA channel with 10% loss rate and retransmission with 160 msec RTT.

I, first P, and second P, (4) only the I and first, second, and third P's, and (5) all the I and P's. Figure 7 plots the expected number of late frames for the *optimized reordering* in each case, and the percentage improvement in the expected number of late frames as compared to the conventional case where no frames are advanced. For the Gamma distribution with $\sigma^2 = .1$ the expected number of late frames can be reduced by 6 to 23% as compared to the conventional case with no reordering, and even for the case of very short delay spread ($\sigma^2 = .05$) a reduction of about 3 to 14% can be achieved. For the WCDMA distribution, the expected number of late frames is reduced by 4 to 19% over the conventional approach.

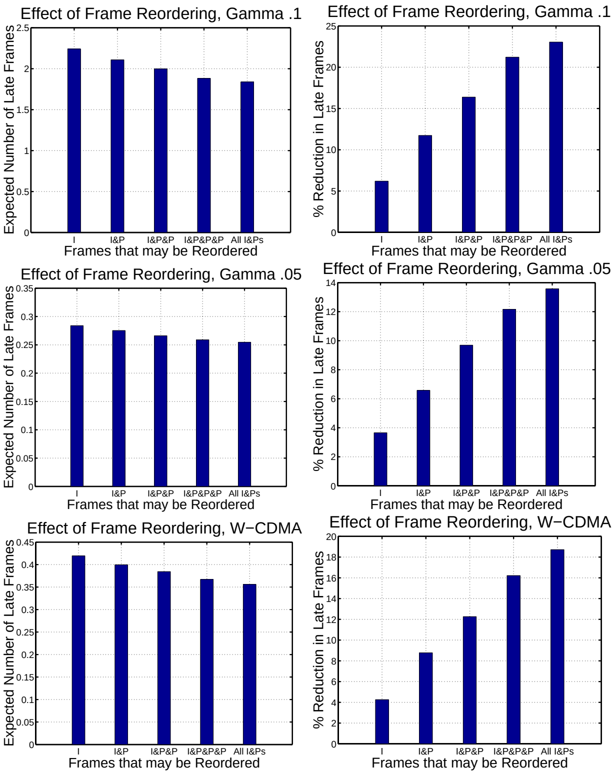

Figure 7: Expected number of late frames (left) and percentage improvement over conventional streaming without reordering (right) as we vary the subset of frames that can be advanced. Results for Gamma PDF with $\sigma^2 = .1$ (top) and $\sigma^2 = .05$ (middle), and WCDMA PDF derived from 3GPP RLC (bottom).

The optimal reordering structures when allowing the advancement of all the I and P frames are shown in Figure 5 for the different delay distributions. The optimal reorderings result in an expected number of late frames/GOP of 1.84, .25, and .36, respectively, as compared to 2.39, .29, and .44 for the conventional cases which do not use reordering. The sum of the total advancements and delays in the reordering is equal to zero in each case.

6. SUMMARY

We have examined the problem of optimal streaming strategies for transmitting predictively coded video over networks with varying delay characteristics. Given knowledge of the delay distribution, optimal frame reordering on a GOP basis can lead to significant reductions in the number of late frames (14 to 23% in these experiments). The optimal reorderings can be precomputed and stored, enabling simple, adaptive, optimized real-time streaming.

7. REFERENCES

[1] A. Ortega and K. Ramchandran, "Rate-distortion techniques in image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, November 1998.

[2] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proceedings of the IEEE*, pp. 1707–1723, October 1999.

[3] J.G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Visual Communications and Image Processing (VCIP)*, January 2001.

[4] Y.J. Liang, E.G. Steinbach, and B. Girod, "Real-time voice communication over the internet using packet path diversity," *Proc. ACM Multimedia*, Sept/Oct 2001.

[5] J.G. Apostolopoulos, T. Wong, W. Tan, and S.J. Wee, "On multiple description streaming for content delivery networks," in *Proceedings of INFOCOM*, Manhatten, NY, July 2002.

[6] T. Yoshimura, Y. Yonemoto, T. Ohya, M. Etoh, and S. Wee, "Mobile streaming media CDN enabled by dynamic SMIL," in *Proceedings of the International World Wide Web Conference*, May 2002.

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

[8] V. Paxson, "End-to-end internet packet dynamics," *Proc. of the ACM SIGCOMM*, pp. 139–152, Sept. 1997.

[9] A. Mukherjee, "On the dynamics and significance of low frequency components of internet load," *University of Pennsylvania Technical Report*, December 1992.

[10] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 2 edition, 1992.

[11] *3GPP Technical Spec. Group Radio Access Network; Radio Link Control protocol specification, TS 25.322 v 3.5.0*, December 2000.