CONGESTION-OPTIMIZED SCHEDULING OF VIDEO OVER WIRELESS AD HOC NETWORKS

Eric Setton, Xiaoqing Zhu and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA94305-9510, USA
{bgirod, esetton, zhuxq}@stanford.edu

Invited Paper

ABSTRACT

We analyze the benefits of information sharing between the application layer and the transport layer, for streaming video encoded at several different qualities, in a mobile wireless network. The application relies on statistics collected at the transport layer and on a video distortion model to select the highest quality that can be supported by the network. At the transport layer, congestion-distortion optimized scheduling is performed to select packets which maximize the received video quality. Experiments performed over a simulated multi-hop wireless network, with H.264 encoded video, show benefits of the proposed approach.

1. INTRODUCTION

Recently, there has been a growing interest in multi-hop ad hoc wireless networks, where nodes communicate with each other without the support of a fixed infrastructure. While this provides appealing features of rapid deployment and flexible configuration for many applications, technical challenges also arise when the network is required to support demanding applications. The traditional OSI layering approach inherent to most network designs does not provide a mechanism for protocol layers to adapt to underlying channel and network conditions, nor to specific application requirements. Meeting the end-to-end performance requirements of demanding applications such as low-latency video streaming is extremely challenging without interaction between protocol layers. Cross-layer design allows information sharing across different layers, and considers adaptive power control, media access control, routing, scheduling and source coding jointly for efficient utilization of network resources. This idea has been explored both in its theoretical aspects, with convex optimization formulations [1, 2], and for more practical system concerns [3, 4].

This work illustrates the advantages of information sharing between the application layer and the transport layer of a video streaming system, operating in a mobile wireless ad hoc network using the 802.11 protocol. In the proposed approach, the transport layer relies on information from the application layer to choose which packets should be transmitted, and when, to maximize decoded video quality. It shapes traffic optimally and reacts to small network impairments by dropping packets, if needed, in order to retransmit more important ones. In addition, the application layer reacts to network capacity fluctuations, estimated at the transport layer, by switching video quality according to a distortion model.

In the next section we describe the video distortion model which extends that presented in [5] to the case where only a subset of video qualities are available. The model is used by the application layer to decide when transmitting all the packets of a lower quality video stream is better than streaming a pruned representation encoded at a higher quality. In Section 3, we explain the congestion-distortion optimized (CoDiO) scheduling algorithm which operates at the transport layer. In Section 4, we analyze experimental results on a simulated ad hoc network. We compare the performance of the proposed scheme with a heuristic scheduler which uses only a limited amount of application layer information to transmit video.

This work was supported, in part, by National Science Foundation Grant CCR-0325639.

2. VIDEO DISTORTION MODEL

For live video streaming applications, video packets are transmitted over the network and need to meet a playout deadline. Decoded video quality at the receiver is therefore affected by two factors: distortion introduced by the encoder compression, denoted by D_{enc}, and distortion due to packet loss or late arrivals, denoted by D_{loss}. Assuming an additive relation of these two independent factors, a video distortion model was derived in [5]. The decoded video distortion, D_{dec}, is given by:

$$D_{dec} = D_{enc} + D_{loss}, \tag{1}$$
$$D_{enc} = D_0 + \theta/(R - R_0), \tag{2}$$
$$D_{loss} = \kappa(P_r + (1 - P_r)e^{-(C-R)T/L}). \tag{3}$$

In (2), R is the rate of the video stream, and the parameters D_0, θ and R_0 are estimated from empirical rate-distortion curves via regression techniques [6]. This is illustrated in Fig. 1.

The second distortion term, D_{loss}, depends linearly on the packet loss rate. The scaling factor κ indicates the sensitivity of the stream to losses which depend on the encoding structure. The other factor reflects the combined rate of random losses and late arrivals. P_r is the random packet loss rate and T is the time within which each packet should reach the receiver (typically a few hundred milliseconds). The parameters C and L depend on the maximum allowable rate and on the average packet size.

This model reflects the impact of the rate on video distortion. At lower rates, reconstructed video quality is limited by coarse quantization, whereas at high rates, more packets are delayed beyond their playout deadline due to network congestion. For live video steaming in a bandwidth-limited environment, we therefore expect to achieve maximum decoded quality for some intermediate rate. This is illustrated in Fig. 1, by the bell shape of the curve representing decoded video quality.

When only a discrete number of video qualities are available the application should decide which video representation to transmit. In the simplest case, the highest quality representation which can be accommodated by the network is chosen. However, this is not always optimal, as the scheduler may transmit a higher quality representation by pruning less important packets. The optimal performance achievable by pruning is shown in Fig. 1 for the *Foreman* sequence encoded at 4 different qualities. The inefficiency of pruning compared to re-encoding is illustrated by the steepness of the curves. In this example, pruning can help reduce the rate of the streams by up to 15%. Beyond this point a lower quality representation should be chosen by the application layer. This is illustrated by the horizontal lines on the figure which characterize quality switches. This criterion will be used for the experiments, presented in Sec. 4, where we consider 4 streams encoded at 915 kbps, 495 kbps, 275 kbps and 120 kbps.

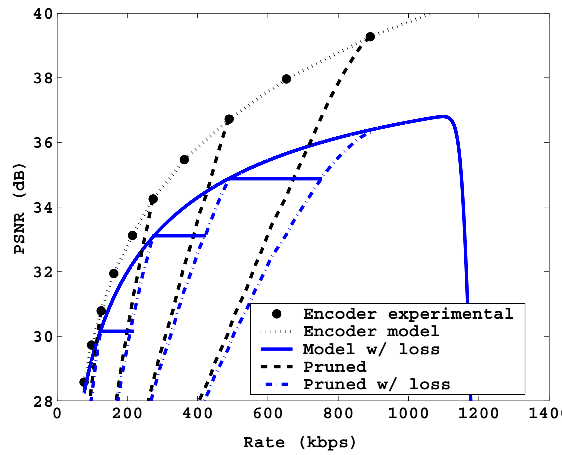

Fig. 1. Rate-distortion characteristic of the CIF *Foreman* sequence encoded at different qualities with H.264. The dotted curve is the fit of the encoder performance, the dashed lines represent the rate-distortion performance of optimally pruned representations. Expected decoded video quality, for a network bandwidth of 1.2Mbps, a loss rate of 1% and a playout deadline of 0.5 s is also represented.

3. CONGESTION-DISTORTION OPTIMIZED SCHEDULING

In this section, we describe how to determine an optimal transmission schedule for the packets of a video stream. This schedule indicates when the packets of the stream will be sent to maximize the decoded video quality at the receiver. To limit the exponential number of possible schedules, the time horizon covered by the schedule is limited. Furthermore, rather than optimizing jointly the schedule for all the packets of the stream, only a small number of packets are selected and the optimization is performed iteratively for each packet, following the approach first described in the seminal work [7].

The aim of CoDiO is to determine a schedule minimizing the expected Lagrangian cost $D + \lambda\Delta$, where D is the distortion of the received video stream and Δ is the end-to-end delay which serves as the congestion metric. In [8], we analyze the benefits of using this metric rather than the traditional objective $D+\lambda R$ used in rate-distortion optimized scheduling [7]. In particular, end-to-end delay is inherently adaptive to time-varying network conditions. In addition, it reflects better the impact of a user operating on a bandwidth-limited network. To minimize the Lagrangian cost, CoDiO selects the most important packets in terms of video distortion reduction, and transmits them in an order which minimizes the congestion created on the network. For example, I frames are transmitted in priority whereas B frames might be dropped. In addition, CoDiO avoids transmitting packets in large bursts as this has the worse effect on the queuing delay.

In the following, we briefly describe how to estimate the expected end-to-end delay and distortion corresponding to a given transmission schedule. This elementary step is repeated several times, at the network layer, to evaluate the performance of different schedules and choose the schedule which performs best.

3.1. Determining the end-to-end delay

In a mobile ad hoc network scenario, the available transmission rate depends on several factors, such as the level of interference, the distance between different hosts or the number of hops of the path. In the scenario we consider, this rate is estimated at the transport layer, by a TCP-friendly rate control (TFRC) agent [9], based on packet loss and end-to-end delay statistics collected from received acknowledgements. Given this capacity estimate, denoted by C, the average end-to-end delay over the time horizon can be estimated for a given schedule. For this purpose the route chosen by the network layer is modelled as a succession of high bandwidth links followed by a bottleneck link of capacity C. The size of the queue at this virtual bottleneck may easily be computed given the transmitted rate at each time instant. A typical illustration of the size of this queue as a function of time is shown in Fig. 2.

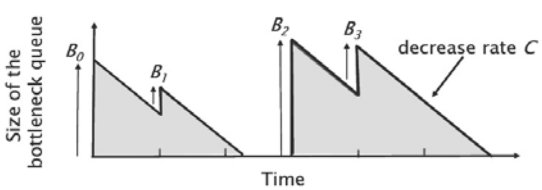

Fig. 2. Backlog at the bottleneck queue

3.2. Determining the video distortion

The expected value of the distortion for the video stream decoded by the client is computed as in [10]. Namely, if copy error concealment is used, an undecodable frame is replaced with the nearest correctly decoded frame for display. Hence, to capture the effect of packet loss on the video quality, only a limited number of display outcomes need to be identified and associated with different distortions. Let $D(s, f)$ denote the distortion resulting from substituting frame s to frame f, the expected distortion when displaying frame f is:

$$D(f) = \sum_{s} D(s, f) Pr\{s\} \tag{4}$$

In Eq. (4), $Pr\{s\}$ represents the probability that frame s is displayed instead of f. This probability may be computed, as described in [10], by combining the probabilities that different packets do not reach the client by their playout deadline. The difficulty resides in estimating the delay distribution function needed to derive these quantities. We model it by a shifted exponential distribution, where the shift is time varying and reflects the backlog at the virtual bottleneck queue. We choose a fixed standard deviation of 30 ms to account for the uncertainty in the capacity estimate.

4. SIMULATION RESULTS

We evaluate the performance of CoDiO in a simulated ad hoc network using the 802.11 protocol in NS-2 [11]. The network consists of 15 mobile nodes randomly placed within a 500m-by-500m square. Each node follows the random walk mobility model with average speed 2.5 m/s. At the network layer, Dynamic Source Routing protocol (DSR) [12] is used to discover and maintain routes. As a basis of comparison, we also use a simple scheduler which does not rely on application layer information other than the packet playout deadline. Packets are sent sequentially and unacknowledged packets are retransmitted 200 ms after their last transmission, as long as their playout deadline has not expired. For this scheduler, the application layer handles quality switches in the same way as for CoDiO.

The *Foreman* CIF video sequence is encoded by the H.264 codec at 30 frames per second at different qualities, using a typical IBBP... coding structure with GOP length 15. In the experiments, packet losses are caused by link failures or overflow of transmission queues due to congestion. Packets arriving at the receiver after their deadlines are discarded. Error concealment is performed by replacing an undecodable frame with the nearest correctly decoded frame. A video streaming session is setup between two nodes lying at opposite extremities of the network. Results are collected for a duration of 40 s which corresponds to 1200 video frames.

Tables 1 and 2 indicate the loss rate and the peak signal to noise ratio (PSNR) for both schedulers for different playout deadlines. For most playout deadlines CoDiO maintains lower loss rate and higher video quality than the sequential scheduler. For a playout deadline of 450 ms, the gain reaches 1.7 dB. This gain is also illustrated in Fig. 3 which shows the PSNR as a function of time. As the latency decreases, both schedulers suffer higher losses as more frames do not reach the decoder by their playout deadline. For the sequential scheduler this is because the number of possible retransmissions is lower. For CoDiO, pruning is used more often to limit network congestion. This degrades the performance at very low latencies. The sequential scheduler has a very low loss rate for a playout deadline between 600 and 650 ms. This is because of the tradeoff between having enough time to reschedule packets and network congestion caused by unnecessary retransmissions.

Playout (ms)	350	400	450	500	600	700
CoDiO	7.9	5.7	4.2	2.5	2	2.2
Sequential	14.2	13.1	11.8	2.9	0.4	1.8

Table 1. Losses (in %) for different playout deadlines.

Playout (ms)	350	400	450	500	600	700
CoDiO	34.9	35.5	36.9	37.2	37.3	37.2
Sequential	35.0	35.2	35.2	36.6	37.0	36.7

Table 2. PSNR in dB for different playout deadlines.

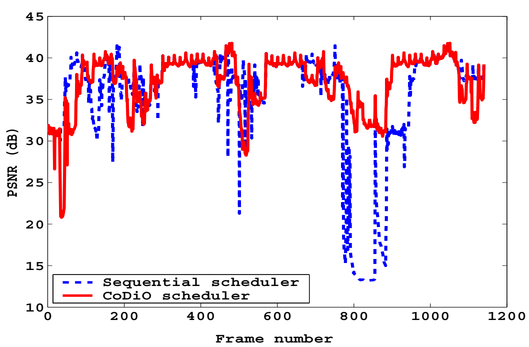

Fig. 3. PSNR trace for both schedulers.

5. CONCLUSION

In this paper, we analyze the benefits of sharing information between the transport layer and the application layer for video streaming in a wireless ad hoc network. Based on a video distortion model and on information collected from received acknowledgements, the application selects the optimal operating quality. At the transport layer, the scheduler prioritizes transmissions to maximize decoded video quality. Experiments performed on a simulated network show the advantages of the scheduler.

6. ACKNOWLEDGMENTS

The authors would like to thank Mark Kalman at Stanford University from whom they obtained optimally pruned video representations.

7. REFERENCES

[1] Y. Wu, P. Chou, Q. Zhang, W. Zhu K. Jain, and S-Y. Kung, "Network planning in wireless ad hoc networks: a cross-layer approach," *IEEE Journal on Selected Areas on Communications, to appear*, Jan. 2005.

[2] R. L. Cruz and A. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," *Proc. INFOCOM, San Francisco, USA*, pp. 702–711, Mar. 2003.

[3] M. van der Schaar and N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles and new paradigms," *Wireless Communications Magazine, to appear.*

[4] A. Goldsmith, E. Setton, T. Yoo, X. Zhu, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," in *Wireless Communications Magazine, to appear.*

[5] E. Setton, X. Zhu, and B. Girod, "Minimizing distortion for multipath video streaming over ad hoc networks," *International Conference on Image Processing, Singapore*, pp. 1751–1754, Oct. 2004.

[6] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–32, June 2000.

[7] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *Microsoft Research Technical Report MSR-TR-2001-35*, Feb. 2001.

[8] E. Setton and B. Girod, "Congestion-Distortion Optimized Scheduling of Video," *Multimedia Signal Processing Workshop (MMSP), Sienna, Italy*, pp. 99–102, Oct. 2004.

[9] M. Handley, S. Floyd, J. Pahdye, and J. Widmer, "Tcp friendly rate control (tfrc): Protocol specification," *RFC 3448*, Jan. 2003.

[10] M. Kalman, P. Ramanathan, and B. Girod, "Rate-distortion optimized streaming with multiple deadlines," *Proc. International Conference on Image Processing, Barcelona, Spain*, Sept. 2003.

[11] "The Network Simulator - ns-2," *http://www.isi.edu/nsnam/ns/*.

[12] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Kluwer Academic Publishers, Ed. 1996.