CONGESTION-AWARE RATE ALLOCATION FOR MULTIPATH VIDEO STREAMING OVER AD HOC WIRELESS NETWORKS

Xiaoqing Zhu, Sangeun Han and Bernd Girod

Information Systems Laboratory, Stanford University, CA, U.S.A.
{zhuxq,sehan,bgirod}@Stanford.EDU

ABSTRACT

For multipath video streaming over ad hoc wireless networks, received video quality is influenced by both the encoder performance and the delayed packet arrivals due to limited bandwidth. We propose a rate allocation scheme to optimize the expected received video quality based on simple models of encoder rate-distortion performance and network rate-congestion tradeoffs. Given the embedded bitstreams of a group of video frames and the cross traffic condition on a given set of paths, a gradient descent method is used to determine the amount of bits to be transmitted over the multiple paths for each frame. Two rate allocation strategies, *layered representation* and *partial repetition* are also investigated.

Simulation results of video streaming over a 15-node wireless ad hoc network demonstrates the benefit of performing rate allocation optimization over a simple heuristic scheme. It is also shown that while layered representation is more efficient for mild delay constraints, partial repetition is more robust against packet losses.

1. INTRODUCTION

Video streaming over ad hoc wireless networks offers numerous possibilities for new applications, such as broadband Internet access, highway automation and search-and-rescue missions. There are, however, challenging technical issues. Video data generally have high bandwidth requirements and stringent delay constraints, whereas the achievable transmission rate on each link in an ad hoc network is usually limited due to power constraints and adverse wireless channel conditions. Joint design of source coding and streaming is needed to achieve satisfactory performance.

Consider choosing an appropriate bit rate for video transmission over a bandwidth-constrained path. At the encoder, a higher bit rate would yield higher reconstruction quality. Over the ad hoc network, however, too high a source data rate may lead to network congestion, hence more packets are dropped when they arrive after the playout deadline. The optimal reconstruction is achieved by some intermediate rate.

The above problem also applies to the case of multipath video streaming, a technique gaining much research attention recently. It is shown in [1] that path diversity is an effective tool for combating correlated packet losses over the Internet. For ad hoc wireless networks, multipath routing is supported by practical protocols such as DSR and TORA [2]. For video applications, the authors in [3] propose to combine multiple description coding with mutlipath transport over the ad hoc network, whereas the benefit of a congestion-optimized multipath routing algorithm is studied in [4]. In the context of multipath streaming, the problem of choosing the best video transmission rate generalizes to rate allocation among multiple routes.

We further propose to optimize rate allocation across a group of video frames and over multiple paths for minimum expected video distortion at the receiver. In the next section, we introduce our notation and formally state the rate-allocation problem. We then derive the models for video distortion and network congestion in Section 3 and 4, respectively. The optimization method is discussed in Section 5. Experimental results and discussions from video streaming over a simulated network are shown in Section 6.

2. PROBLEM FORMULATION

Consider a network of N nodes. Using the couplet (i, j) to denote the directional link from node i to node j, a given route from the source node to the destination node can be viewed as a set of concatenated links:

$$\mathcal{P}_m = \{\text{(i,j)} \mid \text{link from node } i \text{ to node } j \text{ in path } m \} .$$

The achievable transmission rate on link (i, j) is C_{ij}, and the existing cross traffic, F_{ij}. Accordingly, the maximum tranmission rate supported by each path is determined by the residual channel capacity on the bottleneck link:

$$C^{(m)} = \min_{(i,j)\in\mathcal{P}_m} C_{ij} - F_{ij}. \tag{1}$$

The number of bits allocated for the lth frame on the mth path is denoted as R_{ml}, and the entire rate allocation scheme can be conveniently expressed as a single matrix R, for M paths and L frames. With video frame rate V_{fps}, the average rate on each path is:

$$\bar{R_m} = \frac{V_{fps}}{L} \sum_{l=1}^{L} R_{ml}. \tag{2}$$

Assuming an embedded representation of each frame at the encoder, the operational distortion-rate function for the lth frame is $d^{(l)}(r)$ when the bitstream is truncated at r bits. Over the mth path in the ad hoc network, the packet drop

This work is partially supported by NSF Grant CCR-0325639.

rate P_m is a function of the average data rate $\bar{R_m}$, the playout deadline T, the packet size B, and the bottleneck capacity $C^{(m)}$.

Furthermore, the received video distortion D_{rec} depends on whether the bitstream is repeated across different paths. In the *partial repetition* mode, the intial portion of the bitstream is duplicated on all M paths:

$$D_{\text{enc}}^{(l)}(R) = d^{(l)}(\max\{R_{1l}, \ldots, R_{ml}\}) \tag{3}$$

whereas in the *layered representation* mode, the bitstreams on each path correspond to consecutive non-overlapping portions of the orginal bitstream:

$$D_{\text{enc}}^{(l)}(R) = d^{(l)}(\sum_{i=1}^{m} R_{il}). \tag{4}$$

Intuitively, the first mode provides better protection for packet losses, whereas the second mode is more efficient in terms of bandwidth utilization. The performance comparison of these two modes will be discussed in greater details in Section 6.

Rate allocation for a GOP of length L over M paths can now be formulated as the following optimization problem:

$$\begin{aligned} \min_{R} \quad & D_{\text{rec}}(R) = f(D_{\text{enc}}^{(l)}(R), P_m(R)) \\ \text{subject to} \quad & R_{ml} \geq R_{\min}^{(l)} \\ & R^{(l)} \leq R_{\max}^{(l)} \\ & \bar{R_m} \leq C^{(m)} \\ & m = 1, \cdots, M, l = 1, \cdots, L. \end{aligned} \tag{5}$$

In (5), the linear constraints correspond to the lower and upper limits of the rates R_{ml}'s, imposed by the source coder and the available transmission rates. From Eq. (3) and (4), we have $R^{(l)} = \sum_{m=1}^{M} R_{ml}$ for layered representation and $R^{(l)} = \max\{R_{1l}, \cdots R_{Ml}\}$ for partial repetition.

3. VIDEO DISTORTION MODEL

We consider a wavelet coder with 3-D motion-compensated subband decomposition across the frames, followed by the SPIHT image coder to compress each subband image into an embedded bitstream [5]. The structure of the coder is depicted in Fig. 1. For simplicity, we use 1-level of Haar wavelet for temporal decomposition and 3-levels of the popular bi-orthogonal 9/7 wavelet for spatial decomposition. The implementation of the SPIHT algorithm is obtained from the publicly available QccPack library [6], which can record all the operational distortion-rate pairs during the encoding process. Note that the recorded distortion is in the 3-D wavelet coefficient domain.

High-rate analysis of transform vector quantization (TVQ) system states that the distortion-rate performance of an entropy-constrained transform vector quantizer can be expressed as summation of exponentials, where each exponential term denotes the distortion contribution from a transform subband [7]. Accordingly, we propose to approximate the operational distortion-rate function of the video coder

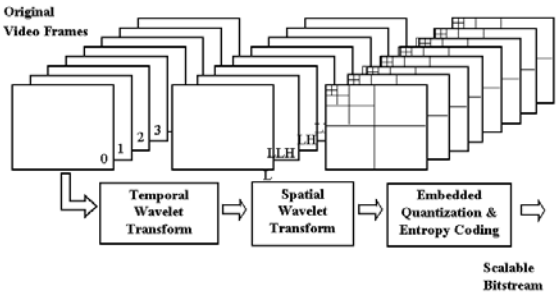

Fig. 1. Structure of the scalable wavelet coder: temporal wavelet decomposition across a GOP followed by spatial wavelet decomposition of each subband image and SPIHT coding of the 3-D wavelet coefficients

as a the sum of exponentials. For each subband image indexed at l, the recorded distortion-rate data is fit with the following expression:

$$d^{(l)}(r) = \sum_{i=1}^{N} \alpha_i e^{-\gamma_i r}. \tag{6}$$

The degree of freedom N is set at 2. The coefficients α_i's and the exponents γ_i's are constrained to be positive. Sample fitting results are given in Figure 2. Relative derivation between the fitted curves and the experimental data $\|d_{\text{model}} - d_{\text{data}}\|^2 / \|d_{\text{data}}\|^2$ is around 0.03. In addition, we assume that distortion in the wavelet coefficient domain is approximately equal to the distortion in the pixel domain, and that contribution from each subband image is additive toward the total distortion. We ignore the possible correlations between the subband images after motion-compensated wavelet decomposition.

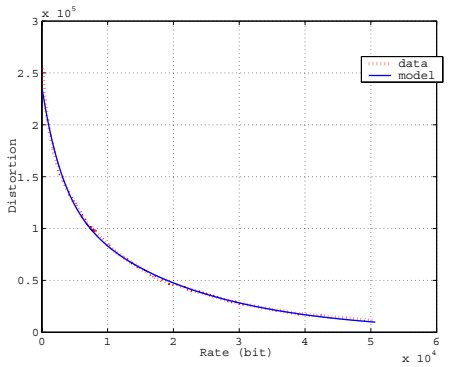

Fig. 2. Sample curve fitting result for the high-band coefficient images. Similar fitting results are observed for the low-band coefficient images

4. NETWORK CONGESTION MODEL

The $M/M/1$ queuing model is used to approximate the packet delay statistics on each path. For further simplification, links in different paths are considered to be disjoint, so that packet drop rates from different paths are independent. We also assume that the end-to-end packet delay is dominated by the bottleneck link along each path, i.e., the link with the minimum residual capacity (see Eq. (1)). This approximation is more accurate at higher rates when a path approaches saturation, which happens to be the region of interest for our optimization. At lower rates, since the packet loss is rare, the simplification in the network model barely affects the optimization.

Packet delay distribution on the mth path is therefore:

$$p_{T^{(m)}}(t) = \lambda_m e^{-\lambda_m t}, t > 0, m = 1, \ldots, M, \tag{7}$$

where λ_m is calculated from:

$$\mathbf{E}\{T^{(m)}\} = 1/\lambda_m = \frac{B}{C^{(m)} - \bar{R_m}}, m = 1, \ldots, M. \tag{8}$$

The packet drop rate due to belated arrivals is:

$$P_m = e^{-\lambda_m T} = e^{-(C^{(m)} - \bar{R_m})T/B} \tag{9}$$

The $M/M/1$ model assumes that the traffic arrival patterns and the packet transmission time follows the Poisson process, which tends to overestimate the packet delays at low rates. To compensate for this effect, a correction factor $\beta_m > 1$ is introduced for each path in the above formula:

$$P_m(R) = e^{-\beta_m (C^{(m)} - \bar{R_m})T/B} \tag{10}$$

The values of β_m's need to be tuned empirically to match simulation results with model prediction, and is set at $3 \sim 6$ for various settings in the experiments. It is noted, however, that although the network model does not match the experimental results closely, it captures the trend between bitrate and packet loss rate properly and is indeed helpful in the optimization.

5. OPTIMIZATION METHOD

Combining the analysis in Section 3 and 4, the objective function of the optimization can now be expressed explicitly. In the case of $M = 2$ paths, for instance,

$$\begin{aligned} D_{\text{rec}}(R) \quad = \quad & \sum_{l=1}^{L} [\, P_1(R)P_2(R)d_0^{(l)} \\ & + P_1(R)(1 - P_2(R))d^{(l)}(R_{2l}) \\ & + (1 - P_1(R))P_2(R)d^{(l)}(R_{1l}) \\ & + (1 - P_1(R))(1 - P_2(R))d^{(l)}(\max(R_{1l}, R_{2l}))] \end{aligned}$$

with partial repetition and

$$\begin{aligned} D_{\text{rec}}(R) \quad = \quad & \sum_{l=1}^{L} [\, P_1(R)d_0^{(l)} \\ & + (1 - P_1(R))P_2(R)d^{(l)}(R_{1l}) \\ & + (1 - P_1(R))(1 - P_2(R)d^{(l)}(R_{1l} + R_{2l})] \end{aligned}$$

with layered representation mode. The expressions for a general M can also be derived in a similar fashion. We focus on the two-path case in the subsequent analysis and simulations.

To achieve the best expected received video quality at different rates, we apply the Lagrangian multiplier technique and solve for

$$\min_R \; D_{\text{rec}}(R) + \lambda \sum_{m=1}^{m=M} \sum_{l=1}^{l=L} R_{ml}, \tag{11}$$

with varying values of λ. Note that the constraints from (5) still apply here. They are incorporated in the gradient descent method in solving the optimization.

To provide an initial solution for the optimization, we propose a heuristic scheme, which performs optimal rate allocation at the encoder by choosing the bitstream length $R^{(l)}$'s to minimize the Lagrangian cost function $d^{(l)}(R^{(l)}) + \lambda R^{(l)}$ at each frame, and then divide each bitstream proportionally with respect to the bottleneck link on each path. This scheme yields optimal performance when no packet losses are involved. It is also noted in the experiments that the heurstic scheme generally provides a fairly good initial solution for the rate allocation matrix R, and that the gradient descent method converges quickly to a local minimum.

6. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed rate allocation optimization, we simulate video streaming over a 15-node static ad hoc wireless network in the ns-2 environment [8]. The nodes are randomly placed within a 100m × 100m square. The transmission rates on the links are calculated from Shannon's channel capacity formula for the AWGN channel, assuming fixed power and simultaneous transmission at all nodes. Link propagation delays and physical layer transmission errors are ignored. Cross traffic on each link is randomly specified up to 70% of the link capacity, with exponentially distributed packet sizes and arrival intervals. Two independent paths are used, each with 4 hops.

The luminance component of the video sequence *Foreman* in QCIF format is encoded with the 3-D wavelet coder in Section 3, for 300 frames, at 30 frames per second. Optimization of the rate allocation is performed over GOP length $L = 16$. Maximum packet size is $B = 1000$ bytes. Bitstreams of length R_{ml}'s are divided into $\lceil R_{ml}/B \rceil$ packets for transmission. Traffic trace files in ns-2 are generated to specify the exact timing and size of each video packet. Queuing trace files on each link are collected to deduce the packet loss information. Actual decoding results are averaged over more than 27 realizations for each bit rate.

Figure 3 and 4 show the video streaming results with playout deadline $T = 350$ ms and $T = 150$ ms, respectively. Received video quality is expressed in terms of *peak-signal-to-noise-ratio (PSNR)* in dB. Rates are denoted in *kilobits per second (kbps)*. Both the layered representation mode and the partial repetition mode are evaluated. The dashed lines on each plot represent the rate-PSNR performance of the heuristic scheme; the solid lines stand for the optimized results. As an upper bound of performance, the encoder rate-PSNR curve is also plotted for comparison.

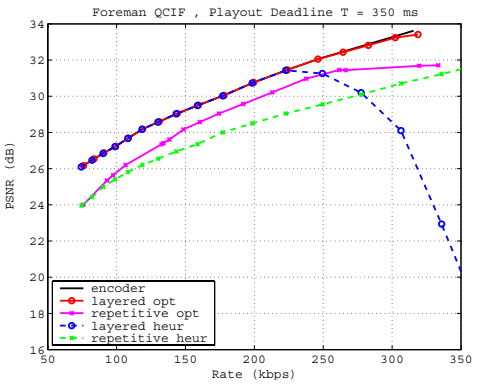

Fig. 3. Rate-PSNR performance for *Forman* QCIF sequence with playout deadline 350ms

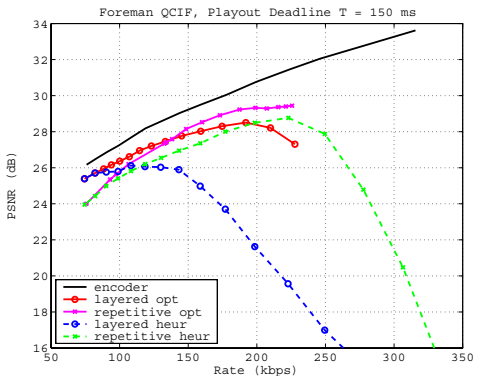

Fig. 4. Rate-PSNR performance for *Forman* QCIF sequence with playout deadline 150ms

As shown in Fig. 3, with layered representation, both the heuristic and the optimal scheme achieve encoder rate-PSNR performance at lower rates, when few packets are dropped due to congestion. As the rate increases, however, the improvement of the optimal scheme over the heuristic scheme becomes more obvious. The best achievable video quality by the optimal scheme is about 2 dB higher in PSNR than that of the heuristics. It is also noted in the experiments that due to losses of initial part of the bistream for low-band coefficient images, the heuristic scheme incurs undesired fluctuation in the received video quality. With partial repetition, on the other hand, video quality of the optimal scheme is $0.5 \sim 1.5$ dB better than the heuristics over a wider range of bit rates. Due to the bitstream repetition on both paths, the efficiency of partial repetition is inferior to that of layered representation.

With a more stringent playout deadline of 150ms, as in Fig. 4, the relative performance of the two modes changes with bitrate. For the optimal scheme, the layered representation mode is more efficient at lower rates and partial repetition is more robust against packet drops at higher rates. The heuristic scheme with layered representation can barely maintain a video quality above 26dB in PSNR, whereas with partial repretition, it can achieve around 29dB in PSNR of decoded video quality. As in the previous experiment, the optimal scheme consistently outperforms the heuristics. The gain in PSNR is $0.5 \sim 1$ dB in the partial repetition mode.

7. CONCLUSIONS

We propose to optimize rate-allocation across a group of frames for multipath video streaming. The proposed objective function, expected distortion at the decoder, incorporates both the influence of encoder performance and network congestion. Based on simple models of encoder video distortion and network congestion, we apply a gradient descent method to solve the optimization. A heuristic scheme is also devised to intialize the optimization and to serve as a basis for comparison. Network simulation results show the benefit of rate allocation optimization, especially at higher rates where the effect of network congestion is more pronounced. The improvement in terms of reconstructed video PSNR is $0.5 \sim 1.5$dB over the heuristic scheme, depending on the delay constraint and the rate allocation strategy in use.

8. ACKNOWLEDGMENTS

The authors would like to thank Professors Andrea Goldsmith and Robert M. Gray at Stanford University for many helpful suggestions and comments.

9. REFERENCES

[1] John G. Apostolopoulos, "Unbalanced Multiple Description Video Communication Using Path Diversity," *Proc. Visual Communicaitons and Image Processing (VCIP)*, pp. 329–409, January 2001.

[2] Elizabeth M. Royer and Chai-Keong Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications*, pp. 46–55, April 1999.

[3] Shiwen Mao, Shunan Lin, Shivendra S. Panar, Yao Wang, and Emre Celebi, "Video transport over ad hoc networks: Multistream coding with multipath transport," vol. 21, pp. 1721 – 1737, December 2003.

[4] Eric Setton, Xiaoqing Zhu, and Bernd Girod, "Congestion-optimized multipath streaming of video over ad hoc wireless networks," *International Conference on Multimedia and Expo (ICME-04), to appear*, July 2004.

[5] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on Set Partitioning in Hierarchical Trees," vol. 6, pp. 243–250, June 1996.

[6] "QccPack-Quantization, Compression, and Coding Library," *http://qccpack.sourceforge.net/*.

[7] M. Goldburg, "Application of wavelets to quantization and random process representations," *Ph.D. Thesis, Stanford University*.

[8] "NS-2," *http://www.isi.edu/nsnam/ns/*.