DISTRIBUTED CHANNEL TIME ALLOCATION AND RATE ADAPTATION FOR MULTI-USER VIDEO STREAMING OVER WIRELESS HOME NETWORKS

Xiaoqing Zhu and Bernd Girod

Information Systems Laboratory
Stanford University
Stanford, CA 94305 U.S.A.
{zhuxq,bgirod}@stanford.edu

Peter van Beek

Sharp Laboratories of America
5750 NW Pacific Rim Blvd.,
Camas, WA 98607, U.S.A.
pvanbeek@sharplabs.com

ABSTRACT

Simultaneous support of multiple video streaming sessions over a shared wireless network requires careful resource allocation to achieve high utilization while dynamically adapting to network and video fluctuations. We propose a distributed algorithm for *channel time allocation* among multiple video streams, and investigate several heuristic packet pruning schemes for rate adaptation of high-definition (HD) video streams. Simulation results are presented for streaming multiple HD video sequences over an 802.11a network. In comparison with TCP-Friendly Rate Control (TFRC) and a basic scheme without rate adaptation, it is shown that the proposed scheme can sustain higher video quality with lower packet delivery delay.

1. INTRODUCTION

With decreasing equipment cost and increasing data rate achieved by recent wireless networking technologies, the prospect of supporting multiple high-definition (HD) video streaming sessions over a wireless home media network gives rise to many attractive applications, as well as numerous technical challenges. The design of such a system needs to address the unpredictable nature of wireless communication channels, while meeting the high data rate and low latency requirements of media streaming. In addition, careful rate allocation is needed to prevent multiple simultaneous video streaming sessions from congesting the shared wireless network. The utility of the allocated rate is also different for streams with different contents: the same rate increase may impact a sequence containing fast motion rather differently than a head-and-shoulder news clip. Rate allocation should therefore maximize total utility across all streams in the network, preferably in a distributed manner.

Multi-user rate allocation is an important and well-studied problem. Practical solutions such as TCP congestion control [1] and TCP-Friendly Rate Control (TFRC) [2] are widely used over the Internet. A mathematical framework of multi-user rate allocation is presented in [3], where the authors also analyzed two classes of distributed solutions, corresponding to the primal and dual decomposition of the optimization objective. In wireless networks, adaptive transmission techniques are typically used to protect the video stream against the time-varying channel [4]. When multiple streams are involved, centralized channel time allocation among multiple wireless stations has been investigated in [5] and [6]. Distributed algorithms have also been proposed, using rate-distortion (RD) optimized packet scheduling in [7] for rate allocation among streams sharing a bottleneck link, and using the subgradient method in [8] for streams competing over a wireless mesh network.

In this work, we focus on multiple HD video streaming sessions sharing a common wireless home network. Even though the video streams may traverse disjoint single-hop links, their rates need to be jointly optimized, as they compete for transmission opportunities over the shared wireless radio channel. In particular, wireless links may experience different channel conditions, resulting in different link speeds. As a consequence, the same allocated rate over a fast link would result in lower fraction of channel time than over a slow link. The rate allocation problem is therefore more conveniently formulated in terms of *channel time allocation*, with the goal of minimizing total video distortion of all streams sharing the network. We present a distributed protocol for solving the allocation problem, by allowing cross-layer information exchange between the MAC and application layers. To avoid the complexity of transcoding HD video streams, rate adaptation is achieved via pruning encoded packets. Several candidate RD-based and heuristic packet dropping patterns are investigated in terms of their computational complexity and RD efficiency.

The rest of the paper is organized as follows. We present the multi-user channel time allocation problem in Section 2, together with a distributed solution based on subgradient method. Section 3 explains the procedures for adapting the video rate via packet pruning, and compares the performance of several candidate pruning schemes. Simulation results of multiple HD video streams over a wireless 802.11a network are discussed in Section 4.

2. CHANNEL TIME ALLOCATION

Consider N video streams sharing a common network. Each Stream i travels over a single-hop connection with effective channel bandwidth C_i, defined as the maximum data rate achieved over that link, *without* contention from any other link in the network. The distortion-rate trade-off of the stream is characterized by $D_i(R_i)$, which may vary for each stream and needs to be updated over time. The goal of channel time allocation is to assign a fraction of channel time s_i to Stream i, with average rate $R_i = s_iC_i$ and distortion $D_i(R_i)$, so as to minimize total video distortion:

$$\min \quad \sum_{i=1}^{N} D_i(s_iC_i), \tag{1}$$

$$\text{s. t.} \quad \sum_{i=1}^{N} s_i < \gamma \tag{2}$$

$$s_i > 0, i = 1, \ldots, N \tag{3}$$

where $\gamma < 1.0$ is an over-provisioning factor to prevent network congestion.

This work is partially supported by NSF Grant CCR-0325639 and a gift from Sharp Labs of America.

The optimization problem can be readily solved in a distributed manner using the subgradient method [9]. Given current observation of its own video distortion-rate trade-off $D_i(R_i)$ and effective link bandwidth C_i, each user updates its channel time allocation factor s_i such that:

$$s_i = \arg\min_s \; D_i(s_i C_i) + \lambda s_i. \tag{4}$$

In (4), λ corresponds to the variable of the dual function of the objective in (1). It is updated according to:

$$\lambda = \lambda - \mu(\gamma - \sum_{i=1}^{N} s_i). \tag{5}$$

The update step size is modulated by a scaling factor μ that decreases over time, and is proportional to the *excess* of total allocated channel time with respect to the constraint γ. Intuitively, the variable λ can be understood as a shadow price that helps to regulate the channel time allocation: if the total channel time exceeds the limit γ, the price increases accordingly, resulting in reduced s_i for each stream according to (4).

Implementation of such a scheme in a practical protocol relies on cross-layer information exchange between the MAC and application layers. The effective channel bandwidth C_i for Stream i is measured as:

$$C_i = \alpha C_i^{prev} + (1-\alpha)\frac{\overline{B}_i}{\overline{T}_i}, \tag{6}$$

by logging average packet size $\overline{B}_i$ and average delivery time $\overline{T}_i$ at the MAC layer, including the overhead of header and ACK transmissions, as well as retransmissions in case of packet losses. In (6), C_i^{prev} denotes the previous instantaneous estimate, which is smoothed over time with the value of α empirically chosen at 0.95. At the application layer, each sending node advertises its intended allocation $s_i = R_i/C_i$ and the value of λ in the video packet header, so that other nodes can overhear such information, and keep its value of λ in sync with the stream bearing the smallest index. Note that s_i is updated upon the transmission of every packet at each node, therefore convergence is fast, due to frequent packet transmissions in HD video streaming.

3. VIDEO RATE ADAPTATION

In order to avoid the complexity of transcoding high-definition (HD) video sequences, we choose to perform rate adaptation by packet pruning, i.e., dropping encoded video packets according to predetermined *omission patterns* within each group of pictures (GOP). Rate R_k of a GOP under omission pattern o_k is calculated by summing over all transmitted packet sizes:

$$R_k = \frac{1}{T_{GOP}} \sum_{l \notin o_k} B_l, \tag{7}$$

where T_{GOP} denotes time duration of one GOP during encoding. Distortion D_k is the empirical mean-squared-error measured from decoding the video sequence *without* the packets in o_k. By varying over different omission patterns, a collection of RD trade-off points $\{(R_1, D_1), (R_2, D_2), \ldots, (R_K, D_K)\}$ can be obtained.

Determination of the omission patterns in a rate-distortion optimized manner would require trial decodings of each GOP with all possible combinations of packet drops. The computational complexity is therefore prohibitive, especially for HD video streams containing hundreds of packets per GOP. Alternatively, one can derive heuristic pruning algorithms based on the distortion contribution of individual packets, as in [7], reducing the number of trial decodings to the order of hundreds per GOP. Also, the packet pruning order may depend on the type of the frame each packet belongs to. Over home media networks, the GOP structure of encoded broadcast media streams typically contains IBBP... frames. In this case, one may choose to drop packets from B frames before omitting any packet from a P frame. We therefore consider the following three candidate schemes for rate adaptation [1]:

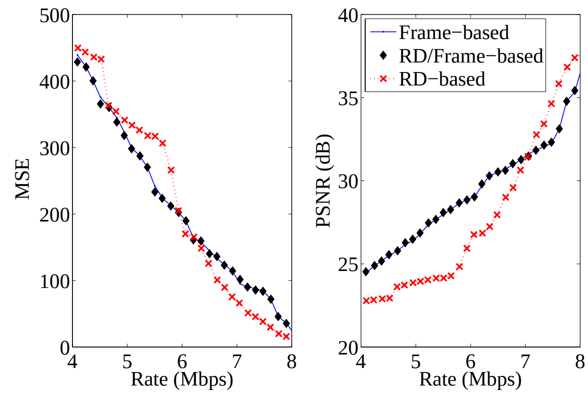

Fig. 1. Rate-Distortion (left) and Rate-PSNR (right) trade-off plots resulting from three pruning schemes for the first GOP of 30 frames in the *City* HD sequence. The H.264/AVC reference codec JM10.2 is used, with maximum slice size of 1400 bytes and RTP packetization of each slice during encoding. Packet-level error concealment is enabled at the JM reference decoder.

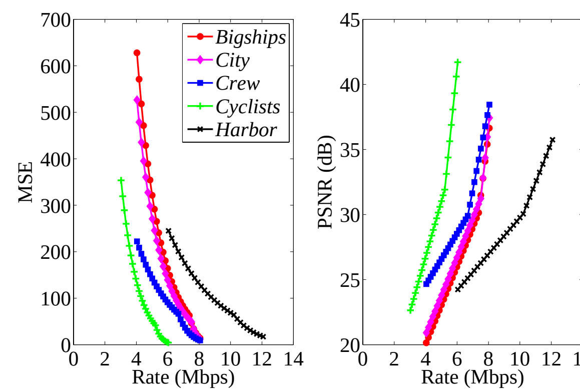

Fig. 2. Fitted Rate-Distortion (left) and Rate-PSNR (right) trade-off curves using the *Frame-based* pruning scheme for the five HD video sequences: *Bigships*, *City*, *Crew*, *Cyclists* and *Harbor*. The fitting is averaged over all GOPs in each sequence.

- *RD-based*: Packets are pruned according to their individual distortion contributions, regardless of frame types.
- *RD/Frame-based*: The omission patterns first include packets from every other B frames, starting from the head of each GOP, followed by packets in the remaining B frames, and finally those from P frames, starting from the end of the GOP.

[1] In all three schemes, packets from I frames and the first packets of each frame are never dropped, to ensure proper operation of the decoder.

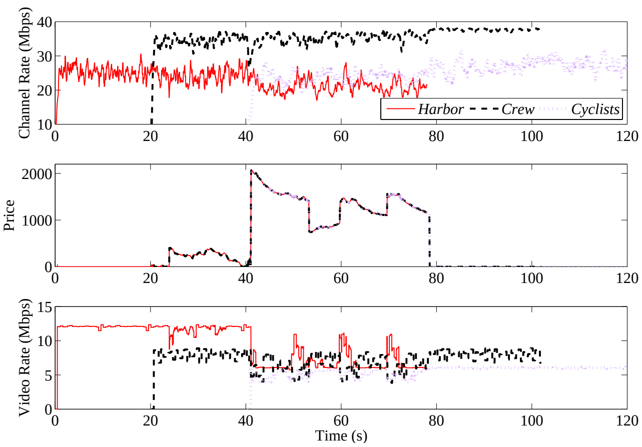

Fig. 3. Traces of estimated channel rate (top), common shadow price (middle) and allocated rate (bottom) for each stream resulting from the proposed rate allocation scheme.

Within each frame, packets of smaller distortion contributions are dropped first.

- *Frame-based*: Packets are dropped according to the frame type they belong to in the same order as in the previous scheme. Within each frame, packets are pruned starting from the end of each encoded frame.

Figure 1 compares the trade-off between rate and encoded video quality obtained by the various pruning schemes, for the first GOP in the *City* HD sequence. It can be observed that the *Frame-based* scheme achieves almost identical rate-distortion trade-off as the *RD/Frame-based* scheme. On the other hand, while the *RD-based* scheme explicitly tries to minimize the distortion contributions from dropped packets and can achieve lower distortions with mild packet drops, as more packets need to be pruned for lower target rates, the underlying additive distortion assumption no longer holds, yielding even higher distortions than the other two schemes guided by frame types. Similar results are observed in other HD sequences. We therefore choose the *Frame-based* scheme for its simplicity.

It can be further noted that the slopes of PSNR values versus rate from the *Frame-based* scheme tend to have two linear segments: the transiting knee point corresponding to where all packets from B frames are dropped. We therefore fit the distortion-rate function $D_i(R_i)$ accordingly, to be used in Eq. (4) for channel time allocation. The fitted trade-off curves of all five HD sequences used in our experiments are plotted in Fig. 2.

4. SIMULATION RESULTS

We simulate in ns-2 [10] a small wireless network with 15 nodes randomly placed in a 100m-by-100m square, all within transmission range of each other. Each node follows the IEEE 802.11a protocol, with a rate of 54 Mbps for payload and 6 Mbps for MAC headers and ACK packets. A 2-state Markov model is used to characterize random packet losses over each link, the good state (G) being successful delivery, and the bad state (B) a packet loss. Effective channel bandwidth is varied by introducing random losses according to the 2-state Markov model with different state transitional probabilities [2].

Five high-definition (HD) video sequences of varying content complexity (*Bigships*, *City*, *Crew*, *Cyclists* and *Harbor*) are considered for streaming over single-hop connections. The sequences have spatial resolution of 1280×720 pixels, and frame rate of 60 fps. The video sequences are encoded using the H.264/AVC reference codec JM10.2 [12], with GOP length of 30 and IBBP... structure. Each slice is constrained to have maximum size of 1400 bytes, and fits into one RTP packet. Rate adaptation is achieved using the *Frame-based* packet pruning scheme in Section 3. In our experiment, we choose 6 Mbps as the starting rate of *Cyclists* sequence with slow motion, 12 Mbps for *Harbor* with the most complex content, and 8 Mbps for the rest. For each pruned version of the encoded bitstream, packet transmission intervals are spread out evenly in the entire GOP duration, so as to avoid unnecessary bursts due to large I frames. Playout deadline is chosen to be 500 ms. Since small ACK packets incur much MAC-layer overheads during transmission, only one ACK is sent per ten received packets.

Figure 3 plots the traces of estimated channel rate, corresponding shadow price λ and rate allocation for three video streams: *Harbor*, *Crew* and *Cyclists*, each over a single-hop connection with link-level packet loss rates of 8.3%, 0.0% and 5.2% respectively. Initially, the network can accommodate one or two simultaneous video streaming sessions at full rate, therefore the common shadow price remains low. At the time of 40 second, the start of the third stream *Cyclists* causes temporary over-utilization of total channel time. Consequently, the shadow price is increased significantly, leading to rate reduction of all three streams. Note that convergence of the shadow price and allocated rates is achieved within a very short period, less than 1.0 second after entrance of the third stream, and that the allocated rate is different for each stream, reflecting the difference in their channel qualities and video contents.

The performance of the proposed allocation scheme is compared against TCP-Friendly Rate Control (TFRC) [2] in Fig. 4. Since TFRC assumes that competing streams share a common bottleneck queue and relies on end-to-end observations such as packet loss and round trip delays to adjust the rates, it fails to converge when the competing video traffic are sent over neighboring wireless links, especially in the case of abrupt changes in the network, e.g., when a new stream joins the network or when an existing stream leaves. As a result, the allocated rates from TFRC tend to experience greater fluctuations and higher packet delivery delay than the proposed scheme.

In the next set of experiments, multiple identical video streams are transported over parallel single-hop connections with the same effective channel bandwidth. We then vary the number of video streams, as well as the channel conditions (by choosing different state transitional probabilities in the 2-state Markov model). Figure 5 compares the decoded PSNR averaged over all video streams achieved by the proposed scheme against the case without rate adaptation. For the experiments with the *Bigships* sequence, the proposed scheme can sustain 3 streams at an acceptable average video quality of 31.2 dB with MAC-level packet loss ratio of 14.2% and effective channel bandwidth of 25 Mbps, whereas the scheme without adaptation fails with milder channel conditions at 27 Mbps, corresponding to 8.3% of MAC-level packet losses. Similar observations can be made for experiments with the other sequences *City*, *Harbor* and *Cyclists*.

[2]The state transitional probabilities p_{GB} and p_{GB} are fitted to a 15-second packet delivery trace collected in [11], with MAC-level packet loss ratio of 8.3%, average duration of 0.8 ms for the bad state, and 8.8 ms for the good state. We then choose similar state transitional probability values, to simulate channels with MAC-level loss ratios in the range of 3 - 14%.

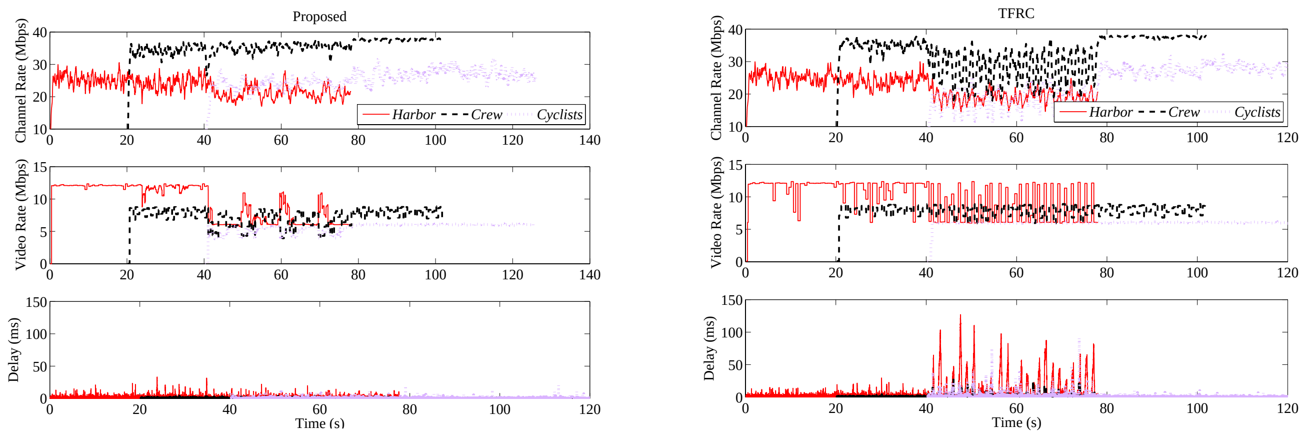

Fig. 4. Estimated channel rate (top), allocated rate (middle) and packet delivery delay (bottom) comparing the proposed scheme and TFRC.

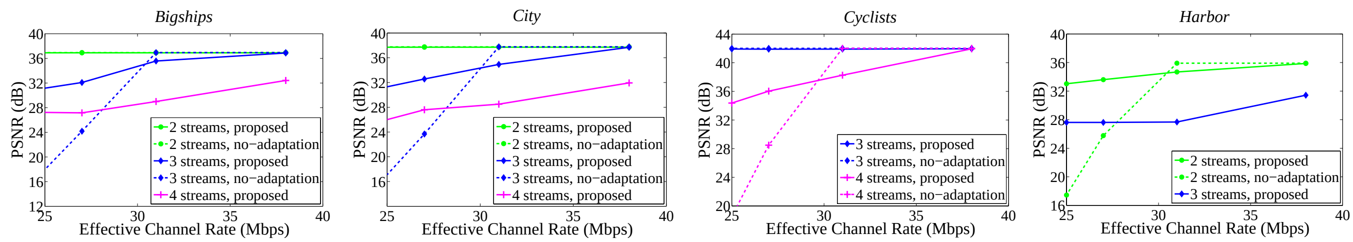

Fig. 5. Decoded PSNR values from the proposed scheme (solid lines) in comparison with the case of non-adaptation (dashed lines).

5. CONCLUSIONS

We investigate a distributed channel time allocation scheme for simultaneous streaming of multiple video sessions over a common wireless network. Given the effective channel bandwidth over the wireless link and the video distortion-rate tradeoff achieved by packet pruning, each video stream can dynamically adjust its outgoing rate according to the common shadow price. Simulation results of multiple HD video streams over a shared 802.11a wireless network show that the proposed rate allocation scheme converges fast in case of abrupt changes in the network (e.g., when a new stream joins or an existing stream leaves), and yields more stable allocation results and lower packet delivery delay than TCP-Friendly Rate Control (TFRC). In comparison with the case without rate adaptation, the proposed scheme is proactive in avoiding congestion, and selectively drops less important packets, therefore it can sustain multiple video streams at acceptable received video quality even with harsher channel conditions.

6. REFERENCES

[1] V. Jacobson, “Congestion avoidence and control,” *Proc. SIGCOMM’88*, vol. 18, no. 4, Aug. 1988.

[2] S. Floyd and K. Fall, “Promoting the use of end-to-end congestion control in the Internet,” *IEEE/ACM Trans. on Networking*, vol. 7, pp. 458–472, Aug. 1999.

[3] F. Kelly, A. Maulloo, and D. Tan, “Rate control for communication networks: Shadow prices, proportional fairness and stability,” *Journal of Operations Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[4] Y. Shen, P. C. Cosman, and L. B. Milstein, “Error resilient video communications over CDMA networks with a bandwidth constraint,” *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3241–3252, Nov. 2006.

[5] M. Kalman and B. Girod, “Optimal channel-time allocation for the transmission of multiple video streams over a shared channel,” *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP’05), Shanghai, China*, Oct. 2005.

[6] M. van der Schaar and N. Sai Shankar, “Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms,” *IEEE Wireless Communications*, vol. 12, no. 4, pp. 50 – 58, Aug. 2005.

[7] J. Chakareski and P. Frossard, “Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources,” *IEEE Trans. on Multimedia*, vol. 8, no. 2, pp. 207–218, Apr. 2006.

[8] X. Zhu and B. Girod, “Distributed rate allocation for multistream video transmission over ad hoc networks,” *Proc. IEEE International Conference on Image Processing (ICIP’05), Genoa, Italy*, vol. 2, pp. 157–160, Dec. 2005.

[9] N. Z. Shor, “Minimization Methods for Non-differentiable Functions,” *Springer Series in Computational Mathematics. Springer*, 1985.

[10] “NS-2,” `http://www.isi.edu/nsnam/ns/`.

[11] P. van Beek and M. Umut Demircin, “Delay-constrained rate adaptation for robust video transmission over home networks,” *Proc. IEEE International Conference on Image Processing (ICIP’05), Genoa, Italy*, vol. 2, pp. 173–176, 2005.

[12] “JM 10.2.,” `http://iphome.hhi.de/suehring/tml`.