DISTRIBUTED MEDIA-AWARE RATE ALLOCATION FOR WIRELESS VIDEO STREAMING

Xiaoqing Zhu and Bernd Girod

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

Invited Paper

ABSTRACT

This paper addresses the rate allocation problem for wireless video streaming. Results from a simple subjective viewing test indicate that perceptually preferred allocation can be closely approximated by minimizing the total mean-squared-error (MSE) distortion of all participating streams. This finding motivates the design of a distributed media-aware rate allocation protocol, which achieves the optimal solution by allowing cross-layer information exchange between the video end hosts and relaying wireless nodes. Various network simulation results confirm the effectiveness of the proposed scheme, which consistently outperforms conventional media-unaware TCP-friendly Rate Control (TFRC).

Index Terms— wireless networking, video streaming, distributed rate allocation protocols

1. INTRODUCTION

Growing networking capabilities of modern wireless devices and increasingly sophisticated techniques in video coding and streaming have spurred many applications for wireless video streaming. Technical challenges abound in such a system, which demands high-rate and low-latency delivery of loss-sensitive compressed video streams over inherently time-varying and error-prone wireless channels. In particular, careful rate allocation is crucial in preventing multiple simultaneous video streaming sessions from congesting a shared wireless network. As pointed out in [1] and [2], the rate allocation problem is further complicated by heterogeneity in both the video contents and the wireless link capacities.

Two major questions await to be addressed, in designing a rate allocation scheme for wireless video streaming:

- *What to optimize?* The goal of a rate allocation scheme is to optimally trade off the video qualities among multiple competing streams, without overloading the wireless network with excessive traffic. While different quality metrics have been adopted in the past for various rate allocation schemes [3][4], it remains unclear which optimization objective would better reflect human preference in a realistic application scenario. The study in [1] suggests that minimizing MSE distortion is the subjectively preferred criteria when multiple video streams of the same resolution time-share a wireless LAN. Such conclusion has yet to be tested in more general settings, e.g., for video streams with heterogeneous resolutions, or over multi-hop wireless networks.
- *How?* Even if an optimization objective has been determined, one still needs to devise a practical scheme to achieve the optimal solution. Ideally, the rate allocation scheme should adapt fast to changes in both the video contents and the wireless link capacities. Preferably, the scheme is distributed in nature, and incurs minimal traffic overhead.

This paper addresses both questions. We first motivate our optimization objective with results from a simple subjective viewing test in the next section. Section 3 then presents the optimization framework for the wireless video rate allocation problem, which accommodates heterogeneity in both the video rate-distortion (RD) characteristics and the wireless link speeds. We explain in Section 4 how the optimal solution can be achieved by a distributed rate allocation protocol, based on cross-layer information exchange between video end hosts and relaying wireless nodes. Finally, Section 5 showcases performance of the proposed scheme in simulations of high-definition (HD) and standard-definition (SD) video streaming over various wireless 802.11a networks.

2. SUBJECTIVE VIEWING TEST

In this section, we briefly describe procedures and findings of a subjective viewing test, designed to collect perceptual ratings of HD/SD video pairs in various quality combinations. A detailed account of the test can be found in [5].

2.1. Test Setup and Procedures

In the subjective viewing test, different video quality levels are represented in the form of still image snapshots. Pairs of test images are displayed side-by-side on two identical LCD monitors at a distance from the viewer. The test data set includes 4 pairs of HD/SD sequences covering a range of video RD characteristics. Each video is represented in 5 quality levels. In addition to the 25 HD/SD image pairs, each test data set also contains pairs of *identical* HD or SD images on both displays, for all 5 quality levels. Ratings of these extra 10 pairs indicate perceptual qualities of individual HD/SD images.

Each test comprises of one training session and two scoring sessions. The training session contains pairs of HD and SD images at the highest, intermediate and lowest quality levels. Its purpose is to familiarize the viewer with the expected range of perceived image qualities. In the scoring session, the viewer provides an opinion score for each presented image pair, based on the perceived overall quality of *both* images, on a scale from 1 to 5.

2.2. Test Results and Analysis

Analysis of the viewers' perceptual ratings, collected from 20 participants for each of the 4 video pairs, leads to a simple subjective quality model. The mean-opinion-score (MOS) of an image pair can

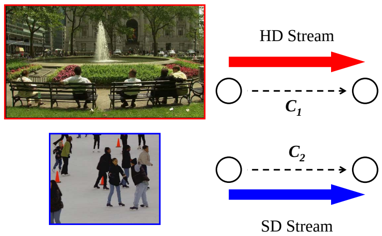
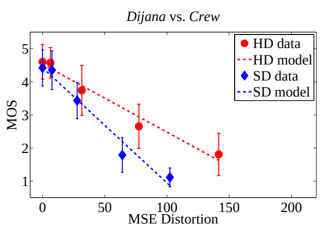
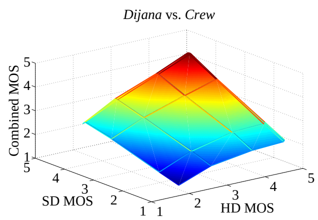


Fig. 1. Network topology for subjective evaluation.



(a)



(b)

Fig. 2. (a) Subjective mean opinion score (MOS) of individual HD/SD images, as a function of their mean-squared-error (MSE) distortions. (b) Subjective mean opinion score (MOS) of image pairs, as function of individual HD/SD image MOS values.

be approximated as a bilinear function of individual image MOS, which, in turn, fits linearly to the MSE distortion of that image. Figure 2 illustrates this model with data collected from one of the HD/SD video pairs.

Now consider one HD and one SD stream over two wireless links, as illustrated in Fig. 1. Various rate allocation schemes would result in different rate pairs (R_{HD}, R_{SD}) along the tradeoff line: $R_{HD}/C_1 + R_{SD}/C_2 = \gamma$, where γ is the target utilization level. Figure 3 (a) shows the MOS derived from the subjective quality model along this tradeoff line. In Fig. 3 (b), the comparison in MOS is presented against varying capacities of the second link. It can be noted that the two schemes minimizing the (weighted) sum of MSE distortion of both video streams closely approximate the subjectively preferred allocation (denoted as ”MAX-MOS” in the figure). In contrast, the media-unaware TCP-friendly allocation leads to significantly lower MOS values, especially when network resources are limited.

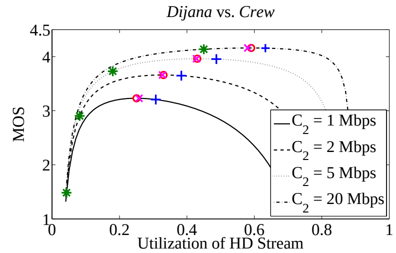
3. OPTIMIZATION FRAMEWORK

3.1. Wireless Network Model

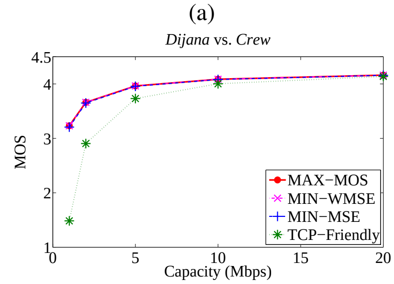
A wireless network is abstracted as a collection of links, each indexed by l. The effective capacity of each link is denoted as C_l[1]; the rate of non-video traffic as F_l. Each video stream is indexed by s, with an allocated rate of R^s. The effect of traffic contention among neighboring links is captured by two key concepts:

- *Channel time utilization*: the fraction of time occupied by active transmission over each link, denoted as u_l. Total link utilization can be calculated as $u_l = (F_l + \sum_s R^s)/C_l$.

[1]Effective capacity of a link is defined as the maximum throughput over that link *without* competition or interference from traffic over any other links.



(a)



(b)

Fig. 3. MOS of rate allocation results, (a) as utilization of the HD stream increases; (b) against varying capacities of the second wireless link. Capacity of the first link is fixed at 20 Mbps. The utilization limit γ is chosen as 90%.

- *Interference set*: the collection of wireless links that cannot transmit simultaneously with a given link l, denoted as $\mathcal{L}_l$. Consequently, total channel time allocation over an interference set is $\tilde{u}_l = \sum_{l' \in \mathcal{L}_l} u_{l'}$.

Note that our model is generic enough to accommodate various types of wireless networks, in which neighboring links time-share their transmission opportunities. In this paper, we focus on 802.11 networks, as a concrete example.

3.2. Video Rate-Distortion Model

For each encoded video stream s, the MSE distortion D^s typically decreases nonlinearly with increasing rate R^s, and can be fitted to a parametric RD model [6]:

$$D^s(R^s) = D_0^s + \frac{\theta^s}{(R^s - R_0^s)}, \tag{1}$$

where the parameters D_0^s, θ^s and R_0^s depend on the coding scheme, the encoder configurations, and the video scene complexity. They need to be updated periodically to track time-varying video contents. A natural choice for the update period is the duration of a group of pictures (GOP) at the encoder, typically in the range of 0.5-2.0s.

3.3. Optimization Objective

Motivated by the subjective viewing test results described earlier, the goal of our rate allocation scheme is to minimize the weighted sum of MSE distortion of all streams:

$$\min_{R^s} \quad \textstyle\sum_s w^s D^s(R^s) \tag{2}$$

$$\text{s. t.} \quad \tilde{u}_l < \gamma. \tag{3}$$

In (2), the scaling factor w^s indicates relative importance of each video stream. The last constraint (3) limits the total channel time utilization within each interference set below a prescribed target $\gamma < 1$.

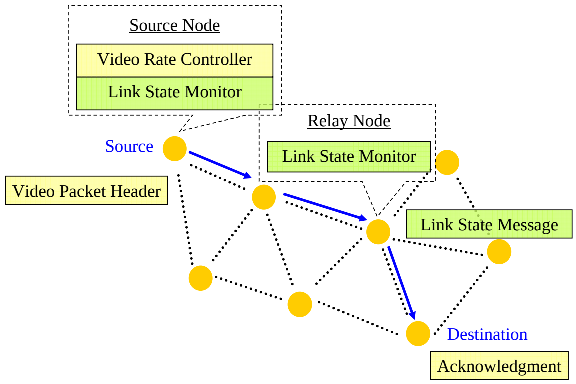


Fig. 4. Cross-layer information exchange among the MAC-layer link state monitors (LSM) and the application-layer video rate controllers (VRC) in the proposed distributed rate allocation protocol.

One can easily verify that the problem in (2) - (3) has a convex objective function with linear constraints. It can be readily solved in a distributed manner using the subgradient method [7]. In the following, we explain how such distributed algorithm can be embodied in the design of a practical rate allocation protocol.

4. DISTRIBUTED RATE ALLOCATION PROTOCOL

4.1. Protocol Overview

Figure 4 illustrates the various components in the proposed protocol. For the purpose of distributed rate allocation, each node consists of a link state monitor (LSM) at the MAC layer and a video rate controller (VRC) at the application layer. Cross-layer information exchange is achieved by granting intermediate wireless nodes access to a set of special video packet header fields. In addition, neighboring nodes exchange knowledge of their local link utilization and congestion level by periodically broadcasting link state messages.

4.2. Link State Monitor

At the MAC layer, the LSM on each wireless node dynamically estimates link state information such as capacity and background traffic rate on its outgoing links. It records *advertised* video rate R^s from the video packet header and calculates the total link utilization u_l of the link traversed by that stream accordingly.

In addition, the LSM needs to track total utilization within the interference sets centered around each of its outgoing links, based on utilization information reported in the link state messages from neighboring nodes. It maintains a congestion price λ_l associated with each interference set, and periodically updates its value as:

$$\lambda_l(t) = \max[\lambda_l(t-\tau) + \kappa(\tilde{u}_l - \gamma)C_l, 0], \tag{4}$$

where τ indicates the interval from last price update and κ is a scaling factor controlling the update step sizes. For $\lambda_l > 0$, the price update is proportional to instantaneous *excess* total utilization over $\mathcal{L}_l$. Intuitively, λ_l increases if total channel time utilization $\tilde{u}_l$ temporarily exceeds the specified limit γ in order to induce rate reduction by all streams affecting $\mathcal{L}_l$. Conversely, as long as $\tilde{u}_l$ is below the target γ, the corresponding congestion price decreases to encourage higher rates from all contributing streams.

Upon relaying each video packet, the LSM updates the field of accumulated congestion price Λ_s in the video packet header as:

$$\Lambda^s := \Lambda^s + \sum_{l' \in \mathcal{L}_l} \frac{\lambda_{l'} C_{l'}}{C_l}. \tag{5}$$

Here, the values of $\lambda_{l'}$ and $C_{l'}$ are collected from link state messages from neighboring nodes. Note that contribution of individual congestion price $\lambda_{l'}$ to the accumulated congestion price Λ^s is weighted by the ratio of the link capacities $C_{l'}/C_l$, reflecting the observation that the same packet transmitted over a fast link would occupy less channel time than over a slow link.

4.3. Video Rate Controller

At the application layer, the VRC at the source node is in charge of tracking video RD characteristics over time. In this work, we assume that the video sequence is pre-encoded at different quality levels, resulting in a discrete set of RD tradeoff points $(R_1, D_1), \ldots, (R_K, D_K)$. The VRC records one set of RD data points and the fitted model parameters (θ, R_0, D_0) for each GOP.

The current allocated video rate R^s_{opt} is advertised in the header field of each outgoing video packet. As the video packet traverses its path, congestion prices of all interference sets encountered by the stream are weighted and accumulated by the LSMs along the way. The VRC at the destination of the stream extracts the end-to-end accumulated congestion price from the video packet header and reports it back to the sender in the same header field of an acknowledgment (ACK) packet. Upon receipt of an ACK, the VRC at the source re-calculates the allocated rate based on current video RD parameters and the accumulated congestion price, as:

$$R^s_{opt} = R^s_0 + \sqrt{\frac{w^s \theta^s}{\Lambda^s}}. \tag{6}$$

The updated video rate is then stamped into subsequent video packet headers, initiating the next round of iteration.

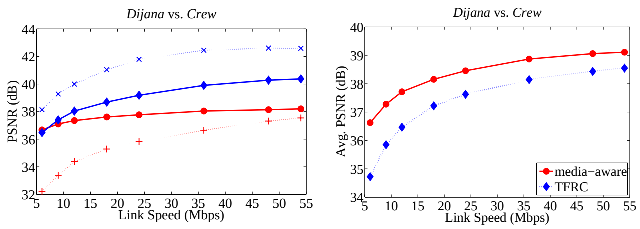
5. PERFORMANCE EVALUATION

5.1. Simulation Setup

Performance of the proposed rate allocation protocol is evaluated in `ns-2` simulations [8]. Parameters of the wireless nodes are chosen according to specifications of the IEEE 802.11a standard [9]. The basic rate for header and control packet transmissions is set at 6 Mbps, whereas the nominal link speed for payload transmissions varies between 6 to 54 Mbps.

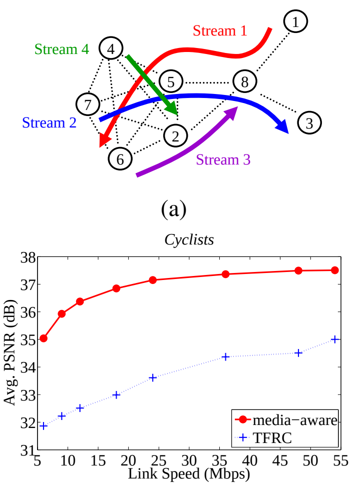
Two HD video sequences (*Dijana* and *Cyclists*) and one SD video sequence (*Crew*) are considered for streaming. They are encoded using `x264` [10], a fast implementation of the H.264/AVC standard [11], at various quantization step sizes. The GOP length is 30 frames; the GOP structure is IBBPBBP..., similar to that used in MPEG-2 streams. Encoded video frames are segmented into packets with maximum size of 1500 bytes for streaming. Packet transmission intervals are evenly dispersed within each GOP to avoid queuing delays due to large intra-coded frames.

Performance of the proposed media-aware rate allocation protocol is compared against the conventional approach of regulating video streaming rate according to the TCP-Friendly Rate Control (TFRC) equation [12], calculated as function of end-to-end observations of packet loss ratios and round trip times.



(a) (b)

Fig. 5. Comparison of media-aware allocation against TFRC in terms of (a) video quality in PSNR of each stream and (b) average video quality of both streams.



(a)

(b)

Fig. 6. Average video quality achieved by the media-aware and TFRC schemes, for four HD streams sharing a multi-hop network.

5.2. Single-hop network

We first consider one HD stream and one SD stream over two parallel links, as shown in Fig. 1. The first link has a nominal speed of 54 Mbps while the speed of the second link varies between 6 Mbps and 54 Mbps. Figure 5 (a) compares the media-aware and TFRC schemes in terms of average video quality of each stream. Since TFRC is agnostic of the different RD characteristics in the two streams, it allocates similar rates for both streams. Such a strategy leads to a large quality gap between the two streams. The media-aware scheme, in contrast, shifts the network resource from the SD stream over the slower link to the HD stream over the faster link, thereby reducing the video quality gap. Consequently, it also improves the overall video quality over TFRC, by 0.6 - 1.7 dB in PSNR of the average distortion of both streams, as shown in Fig. 5 (b).

5.3. Multi-hop network

Next, we consider a more complex multi-hop network, as shown in Fig. 6 (a). In this case, four HD sequences with the same content from *Cyclists* stream over paths of different hop counts. All links operate at 54 Mbps except the one from Node 1 to Node 8, the nominal speed of which varies between 6 Mbps and 54 Mbps. As shown in Fig. 6, improvement of the media-aware scheme over TFRC ranges between between 2.1 - 4.1 dB in PSNR of average video quality.

6. CONCLUSIONS

We present in this paper a distributed rate allocation protocol for wireless video streaming, which aims at minimizing total mean-squared-distortion (MSE) distortion of all streams. The optimization objective is based on subjective viewing test and the observation that MOS-based rate allocation optima are very broad. Hence, in practice, minimizing total MSE performs almost as well. Simulation studies of high-definition (HD) and standard-definition (SD) video streaming over various 802.11 networks demonstrate the benefit of media-aware rate allocation over TCP-Friendly Rate Control (TFRC). Performance gain varies between 0.7 to 4.1 dB in PSNR of average video quality. Perceptual improvements of the competing video streams are manifested in the form of reduced quality gaps.

7. REFERENCES

[1] M. Kalman and B. Girod, "Optimal channel-time allocation for the transmission of multiple video streams over a shared channel," in *Proc. IEEE International Workshop on Multimedia Signal Processing*, Shanghai, China, Oct. 2005.

[2] X. Zhu and B. Girod, "Distributed rate allocation for video streaming over wireless networks with heterogeneous link speeds," in *Proc. International Symposium on Multimedia over Wireless*, Honolulu, HI, USA, Aug. 2007, pp. 296–301.

[3] T. Schierl, S. Johansen, C. Hellge, T. Stockhammer, and T. Wiegand, "Distributed rate-distortion optimization for rateless channel coded scalable video in mobile ad hoc networks," in *Proc. International Conference on Image Processing*, San Antonio, TX, USA, 2007, vol. 6, pp. 497–500.

[4] S. Khan, S. Duhovnikov, E. Steinbach, and W. Kellerer, "MOS-based multiuser multiapplication cross-layer optimization for mobile multimedia communication," *Advances in Multimedia*, 2007.

[5] X. Zhu and B. Girod, "Subjective evaluation of multi-user rate allocation for streaming heterogeneous video contents over wireless networks," in *Proc. IEEE International Conference on Image Processing (ICIP'08)*, San Diego, CA, USA, Oct. 2008, pp. 3092–3095.

[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] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer Series in Computational Mathematics. Springer, NY, USA.

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

[9] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, P802.11*, Nov. 1997.

[10] "x.264," `http://www.videolan.org/x264.html`.

[11] ITU-T and ISO/IEC JTC 1, *Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264 - ISO/IEC 14496-10(AVC)*, 2003.

[12] 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.