Video Multicast over Wireless Mesh Networks with Scalable Video Coding (SVC)

Xiaoqing Zhu[a], Thomas Schierl[b], Thomas Wiegand[b] and Bernd Girod[a]

[a]Information Systems Lab, Stanford University, 350 Serra Mall, Stanford, CA 94305, USA
[b]Fraunhofer HHI - Heinrich Hertz Institut, Einsteinufer 37, 10587 Berlin, Germany

ABSTRACT

We address the problem of rate allocation for video multicast over wireless mesh networks. An optimization framework is established to incorporate the effects of heterogeneity in wireless link capacities, traffic contention among neighboring links and different video distortion-rate (DR) characteristics. We present a distributed rate allocation scheme with the goal of minimizing total video distortion of all peers without excessive network utilization. The scheme relies on cross-layer information exchange between the MAC and application layers. It adopts the scalable video coding (SVC) extensions of H.264/AVC for video rate adaptation, so that graceful quality reduction can be achieved at intermediate nodes within each multicast tree. The performance of the proposed scheme is compared with a heuristic scheme based on TCP-Friendly Rate Control (TFRC) for individual peers. Network simulation results show that the proposed scheme tends to allocate higher rates for peers experiencing higher link speeds, leading to higher overall video quality than the TFRC-based heuristic scheme.

Keywords: wireless mesh network, video multicast, scalable video coding (SVC), distributed rate allocation

1. INTRODUCTION

In multicast video delivery, the same video content is streamed to a group of interested receivers via multicast trees. This can be achieved either at the network layer via IP multicast, or by application-layer relaying in a peer-to-peer fashion.[1] With the advance of wireless networking technologies, video multicast over wireless mesh networks is also gaining popularity, for potential applications such as broadband multimedia content sharing within a residential community or among cell phone users.

Rate allocation for video multicast is necessary for accommodating heterogeneity in receiver capabilities and avoiding network congestion. It is preferable to have a scalable representation of the video bit streams, so that the presence of a single user with slow network connection will not reduce the received video quality of other users in the same multicast session.[2] The rate allocation problem is further complicated in wireless networks due to traffic contention among neighboring links and heterogeneity in their transmission speeds. For instance, in 802.11 wireless networks with contention-based MAC protocols, packets of the same size sent over a slow link would occupy the shared wireless channel for a longer duration than those sent over a fast link.[3] The utility of the allocated rate is also different for video streams with different contents: the same rate increase may impact a sequence containing detailed textures and fast motion rather differently than a head-and-shoulder news clip. Therefore, it is important for the rate allocation scheme to take into consideration both wireless link heterogeneity and video distortion-rate (DR) characteristic, especially when multiple video multicast sessions baring different video contents are sharing the same wireless network.

In Ref. 4 and 5, we have developed a distributed rate allocation protocol for multi-user video streaming over wireless mesh networks. The protocol attempts to minimize total video distortion of all participating streams, without incurring excessive network congestion. This is achieved by local maintenance and exchange of congestion prices among neighboring wireless nodes at the MAC layer, in conjunction with vide rate adaptation of each stream at the application layer. The focus of this paper is to extend the framework for video multicast delivery. Due to heterogeneity in the wireless link speeds, different peers in the same multicast tree may have different

This work is supported, in part, by NSF Grant CCR-0325639 and a gift from HiSilicon Inc.
Further author information: [a] E-mail: {zhuxq,bgirod}@stanford.edu, Telephone +1 650 723 3476, Fax +1 650 724 3648;
[b] E-mail: {schierl,wiegand}@hhi.de, Telephone +49 30 31002 227, Fax +49 30 392 7200

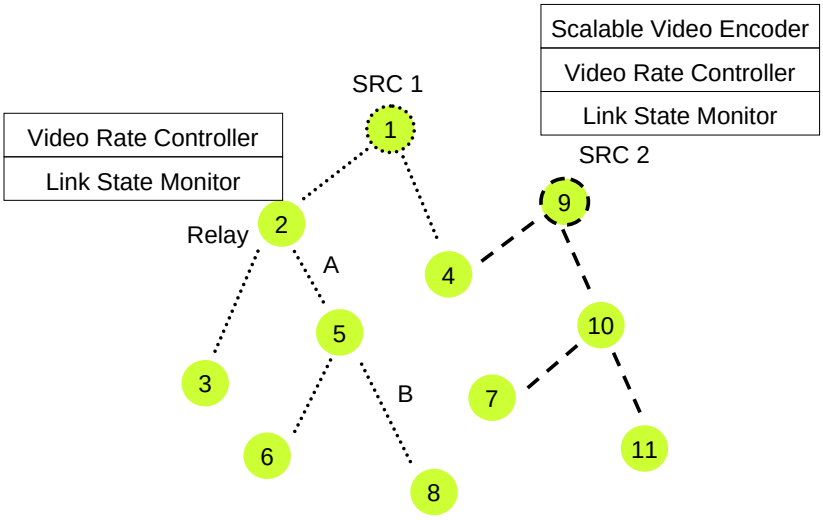


Figure 1. Overview of the wireless video multicast system illustrated with two ALM trees.

allocated rates. We therefore adopt the scalable video coding (SVC) extensions of H.264/AVC,[6] to allow graceful quality reduction at intermediate nodes within a multicast tree.

The remainder of the paper is organized as follows. In the next section, we formulate the rate allocation problem in the convex optimization framework, present a distributed solution, and discuss practical considerations in protocol design. The rate adaptation procedures based on SVC are then described in Section 3. In Section 4, we compare performance of the proposed scheme against a heuristic approach based on TCP-Friendly Rate Control (TFRC) of individual peers. Simulation results are presented for scenarios of single or multiple video multicast sessions sharing an 802.11a network.

2. DISTRIBUTED RATE ALLOCATION

2.1. System Model

An overview of the wireless video multicast system is shown in Fig. 1. The wireless nodes self-organize into a mesh network and form application layer multicast (ALM) trees for video delivery. In each ALM tree, the video content is encoded at the source node into a SVC bit stream, which is subsequently forwarded by intermediate nodes to reach all peers. Each wireless node consists of a link state monitor at the MAC layer and a video rate controller at the application layer.

We index each wireless link with l and denote the set of links as $\mathcal{L}$. The set of video multicast sessions, each served by an ALM tree, is denoted as $\mathcal{S}$. The distortion-rate tradeoff of each SVC stream is described using a parametric model[7]:

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

where the parameters D_0^s, R_0^s and θ^s can be fitted to empirical data from trial encodings using nonlinear regression techniques.

For each video multicast session $s \in \mathcal{S}$, the corresponding ALM tree $\mathcal{T}^s$ consists of all links traversed by that tree. We use $l' \xrightarrow{s} l$ to represent that Link l *succeeds* Link l' in $\mathcal{T}^s$, when the destination node of Link l' is parent of the destination of Link l. In Fig 1, for instance, Link B succeeds Link A in the first tree.

The capacity C_l is defined as the maximum achievable data rate over Link l, when the rest of the network is *not* transmitting. Total traffic rate over that link is $F_l = F_l' + \sum_{s:l\in\mathcal{T}^s} R_l^s$, including both the rate of non-video traffic F_l' and the allocated rates R_l^s's of all video sessions traversing that link.

In wireless networks with contention-based MAC protocols such as IEEE 802.11,[8] neighboring links interfere with each other, so only one link can transmit at any time instant. We define link utilization as the fraction of time during which each link is active: $u_l = F_l/C_l$. The set of links that cannot transmit simultaneously with Link l constitutes its *interference set* $\mathcal{L}_l$. Consequently, total utilization within each interference set $\mathcal{L}_l$ is constrained by

$$\tilde{u}_l = \sum_{l' \in \mathcal{L}_l} u_{l'} < \gamma, \tag{2}$$

where $\gamma < 1$ is an over-provisioning factor. The extra headroom is needed to absorb various effects not included in our model, such as random backoff in a CSMA/CA network to resolve traffic contention over the shared wireless media, or inaccurate estimates of instantaneous link capacity.

2.2. Optimization Objective

The goal of the rate allocation scheme is to minimize total video distortion of all participating peers without incurring excessive network utilization:

$$\min \quad \sum_{s \in \mathcal{S}} \sum_{l \in \mathcal{T}^s} D^s(R_l^s) \tag{3}$$

$$\text{s. t.} \quad \tilde{u}_l < \gamma, \qquad \forall l \in \mathcal{L} \tag{4}$$

$$R_l^s \le R_{l'}^s, \ l' \xrightarrow{s} l \qquad \forall l \in \mathcal{L}, \forall s \in \mathcal{S}. \tag{5}$$

The constraint (5) states that the allocated rate for a peer can never exceed the rate allocated to its parent. Note that the optimization framework can be easily extended to accommodate different weighting factors for each peer, or to maximize total video quality in terms of their PSNR values.

It can be easily verified that the optimization objective (3) is a convex function of the allocated rates R_l^s, therefore can be efficiently solved by centralized algorithms, such as the interior point method.[9] However, the drawback of a centralized scheme is that all the computation is carried out by one node, which also needs to collect global information such as the capacity and the rate of existing traffic on all links in the network. The complexity of the algorithm and the overhead of information collection scale approximately with the square of the total number of nodes in the network. As the network size grows, it is preferable to solve the problem in a distributed manner.

2.3. Distributed Solution

In this section, we present a distributed solution for solving (3) - (5) based on the subgradient method.[10] For each constraint in (4), a corresponding variable λ_l is introduced and is updated according to:

$$\lambda_l = [\lambda_l - \kappa(\gamma - \tilde{u}_l)]^+. \tag{6}$$

In (6), $[.]^+$ denotes that the value of λ_l is constrained to be positive. Its update step size is proportional to the *excess* total utilization within $\mathcal{L}_l$, and is modulated by a scaling factor κ that decreases over time to ensure convergence.[10] Intuitively, the variable λ_l can be interpreted as the *congestion price* within each interference set: it decreases when total utilization within $\mathcal{L}_l$ is below the limit γ to encourage higher allocated rate, and increases vise versa.

At the application layer, the optimal rate for Session s over Link l is calculated as:

$$R_l^s = \underset{R < R_{l'}^s, l' \xrightarrow{s} l}{\operatorname{argmin}} \ [D^s(R) + \Lambda_l R]. \tag{7}$$

where $\Lambda_l = \sum_{l' \in \mathcal{L}_l} \lambda_{l'}/C_l$ is the accumulated price of all links within the interference set of Link l, scaled by its capacity C_l. It reflects the impact of R_l^s on the interference sets of its own and neighboring links. The allocation R_l^s is further constrained by the video rate allocated to its parent. Given the DR parameters and accumulated price, the optimal rate can be calculated analytically:

$$R_l^s = \begin{cases} R_0^s + \sqrt{\frac{\theta^s}{\Lambda_l}}, & \text{if } \ R_{l'}^s < R_0^s + \sqrt{\frac{\theta^s}{\Lambda_l}} \\ R_{l'}^s \quad , & \text{otherwise} \end{cases}. \tag{8}$$

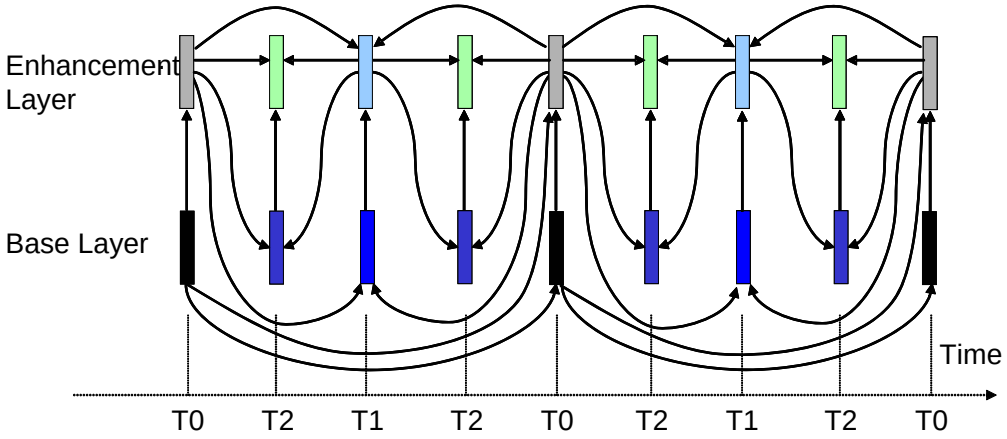


Figure 2. Structure of a group of pictures (GOP) in a SVC encoded stream with SNR and temporal scalability. The GOP length is 4 in this example.

2.4. Protocol Design

Implementation of such a scheme in a practical protocol relies on cross-layer information exchange between the MAC and application layers. As shown in Fig. 1, each wireless node consists of a link state monitor at the MAC layer and a video rate controller at the application layer.

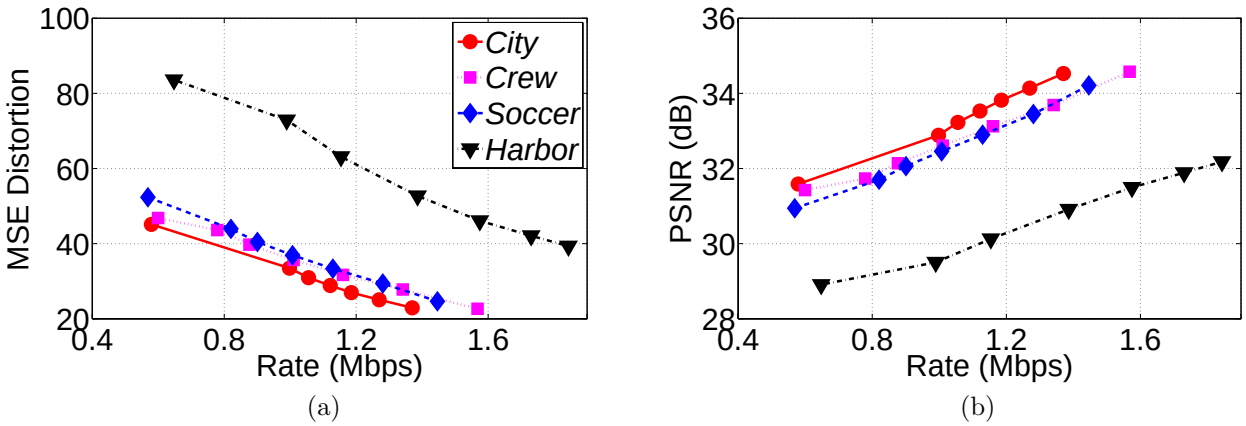
The source node of each wireless link dynamically monitors its outgoing link capacities and non-video traffic flows by logging average packet size and delivery time at the MAC layer. The link state monitor for Link l also records the intended rate allocation R_l^s advertised by all multicast sessions traversing that link, for deriving the corresponding link utilization u_l. Periodic broadcast of link state messages are used to collect utilization from neighboring links for calculating total utilization $\tilde{u}_l$, as well as to collect congestion prices $\lambda_{l'}$ in neighboring interference sets. Upon relaying of a video packet, the congestion price λ_l is updated according to (6). The accumulated price Λ_l is then calculated and reported in the video packet header.

At the application layer, a video rate controller extracts the value of Λ_l, and calculates the optimal video rate R_l^s based on (7). After the allocation has converged, the video rate controller truncates the incoming SVC bitstream accordingly. In practice, the video DR characteristic may change abruptly, for instance when there is a scene cut. Such changes are tracked by the video rate controller at the source node, which periodically fits the video DR model parameters to data points obtained from the current GOP and reports such information in the video packet headers.

3. RATE ADAPTATION WITH SCALABLE VIDEO CODING (SVC)

In the proposed scheme, allocated rates to peers receiving the same video stream may be different due to heterogeneity in their link speeds. We adopt the scalable video coding (SVC) extensions[6, 11] of H.264/AVC standard for video rate adaptation, so as to avoid the computational complexity of transcoding and to achieve more efficient distortion-rate tradeoff than packet pruning from a non-scalable bitstream.

In SVC, the motion-compensated transform coding architecture is extended to achieve a wide range of spatio-temporal and quality scalability. The structure of a group of picture (GOP) in the SVC stream is illustrated in Fig. 2. Pictures labeled T0 form the first temporal layer; pictures labeled T1 and T2 form the first and second temporal layers respectively. Note that the base layer pictures labeled with either T1 or T2 are predicted from the highest enhancement layer pictures. This approach, also known as Medium Granularity Scalability (MGS), provides a balance between high coding efficiency of the base layer and mismatch error control when the reference pictures are missing. Multiple bit rate points can be obtained via temporal scalability within the MGS enhancement layer, by sequentially dropping enhancement pictures according to their temporal priorities. Figure 3 shows the rate-distortion performance of SVC for four 4CIF video sequences later used in the experiments.



(a) (b)

Figure 3. (a) Rate-distortion and (b) Rate-PSNR tradeoff achieved by SVC for four 4CIF sequences: *City*, *Crew*, *Soccer* and *Harbor* with frame rate of 60 fps. The sequences have frame rate of 60 fps, and are encoded with GOP length of 32.

4. EXPERIMENTAL RESULTS

We evaluate performance of the proposed distributed rate allocation scheme with simulations in ns-2[12] over a small mesh network shown in Fig. 1. The MAC parameters of the wireless nodes are set according to the specifications in IEEE 802.11a protocol.[8] The outgoing link speed of all wireless nodes is 54 Mbps unless otherwise stated. Four ITU-T test sequences *City*, *Crew*, *Soccer* and *Harbor* with 4CIF resolution (704x576 pixels) and frame rate of 60 fps are considered. They are encoded using the SVC JSVM 8.8 reference encoder,[6] with one MGS quality layer and 6 temporal layers, at GOP length of 32 frames.

As a basis of comparison, we also implement a heuristic rate allocation scheme, where the rate of each peer is calculated according to the TCP-Friendly Rate Control (TFRC) equation:

$$R_l^s = \frac{B}{RTT_l\sqrt{p_l}}, \tag{9}$$

where B denotes average size of the video packets; p_l and RTT_l are the estimated packet loss ratio and round-trip-time over Link l.[13]

In both schemes, rate adaptation at the intermediate node is achieved by truncating incoming SVC bitstreams according to the allocated rate. The encoded base layer and enhancement layer pictures of each frame are further segmented into packets with maximum size of 1500 bytes, and are reassembled at the receivers. Upon complete receipt of each encoded frame at each peer, an acknowledgment packet of 40 bytes is sent to its parent, containing feedback information (e.g., accumulated price Λ_l, estimated packet loss ratio p_l) required by the rate allocation scheme.

We first consider the simple case of a single video stream, *Harbor*, streaming over the first ALM tree in Fig. 1. Figure 4 shows the traces of total utilization with the interference set of Link $1 \rightarrow 2$, of accumulated link price Λ_l, as well as allocated rate to each peer, when the outgoing link speed of Peer 5 is 24 Mbps. In this simple example, all links are within the same interference set, therefore their link prices λ_l's are the same. However, due to the lower capacity C_l observed over Links $5 \rightarrow 6$ and $5 \rightarrow 8$, the accumulated price Λ_l for the two receiving peers 6 and 8 are higher, leading to lower allocated rate. It can also be noticed that the prices and allocated rates reach convergence within about 2 seconds. The fluctuations in the allocation afterwards reflect variations in the video RD characteristics across GOPs.

Figure 5 compares the two schemes in terms of PSNR values corresponding to average video distortion of all peers, averaged over 60 seconds after convergence. The *Harbor* sequence is streamed over the first tree in Fig. 1. As the outgoing link speed of Peer 5 reduces from 54 Mbps to 6 Mbps, the TFRC scheme reduces allocated rate to all peers due to increased round-trip-time over all links in this small network. The proposed scheme, in comparison, maintains the video quality of most peers at the cost of reducing allocation to the two peers with slower link speed. This leads to improved overall video quality, measured as PSNR of the average video

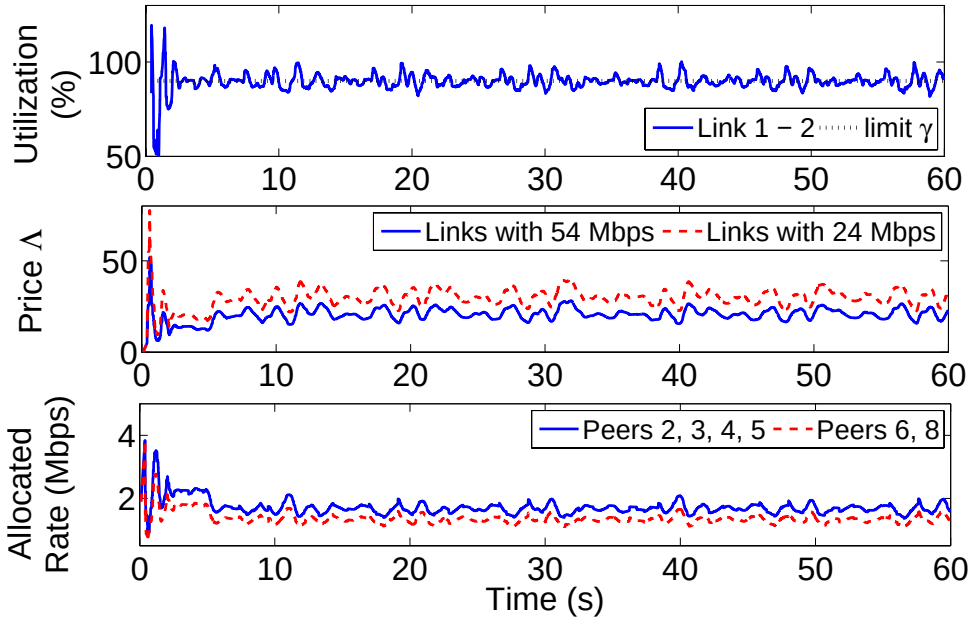


Figure 4. Trace of total utilization with the interference set of Link $1 \rightarrow 2$, accumulated price Λ reported by each link and corresponding allocated rates, when *Harbor* is being streamed over the first ALM tree in Fig. 1. The outgoing link speed of Peer 5 is 24 Mbps whereas all other nodes has outgoing link speeds of 54 Mbps. The value of utilization limit γ is chosen to be 0.90, as plotted in the top figure in black dotted lines.

distortion of all peers, as shown in Fig. 6 (a). Similar behavior is observed when the *City* sequence is streamed over the first tree in Fig. 6 (b). The performance gain ranges from 0.2 dB to 1.4 dB, depending on the video sequence content and wireless link speeds.

In the next experiment, two video streams are delivered over two ALM trees respectively, as in Fig 1. The outgoing link speed of Peer 5 varies from 6 Mbps and 54 Mbps. Figure 7 (a) shows the comparison between the proposed scheme and the TFRC-based heuristics, when *Soccer* is streamed over the first tree and *City* is streamed over the second tree. The proposed scheme benefits from explicit knowledge of the video DR characteristics and wireless link speeds, therefore outperforms the TFRC-based heuristic by about 0.5 dB in terms of PSNR of average video distortion of all peers in both trees. Similar performance gains can be observed in Fig. 7 (b), with *Crew* streaming over the first tree and *City* over the second tree.

5. CONCLUSIONS

When multiple video multicast sessions are present in the same wireless mesh network, careful rate allocation is needed to balance the goals of preventing network congestion and improving received video quality at all peers. In this work, we present a distributed rate allocation scheme, which attempts to minimize the total video distortion of all peers, without excessive network utilization. The allocation results are adapted to heterogeneity in wireless link speeds, as well as the time-varying video distortion-rate (DR) characteristics in each multicast session. In addition, the scalable video coding (SVC) extension of the H.264/AVC standard is adopted for video rate adaptation, so that more graceful quality reduction can be achieved at intermediate nodes within each multicast tree. Performance of the proposed scheme is compared to a TFRC-based heuristic scheme in network simulations involving single or multiple video multicast sessions. Results show that the proposed scheme tends to maintain good video quality for peers with higher wireless link speeds, by reducing allocation to peers with slower links. This leads to performance gains of up to 1.4 dB over the TFRC-based heuristic scheme, in terms of PSNR of average video distortion of all peers.

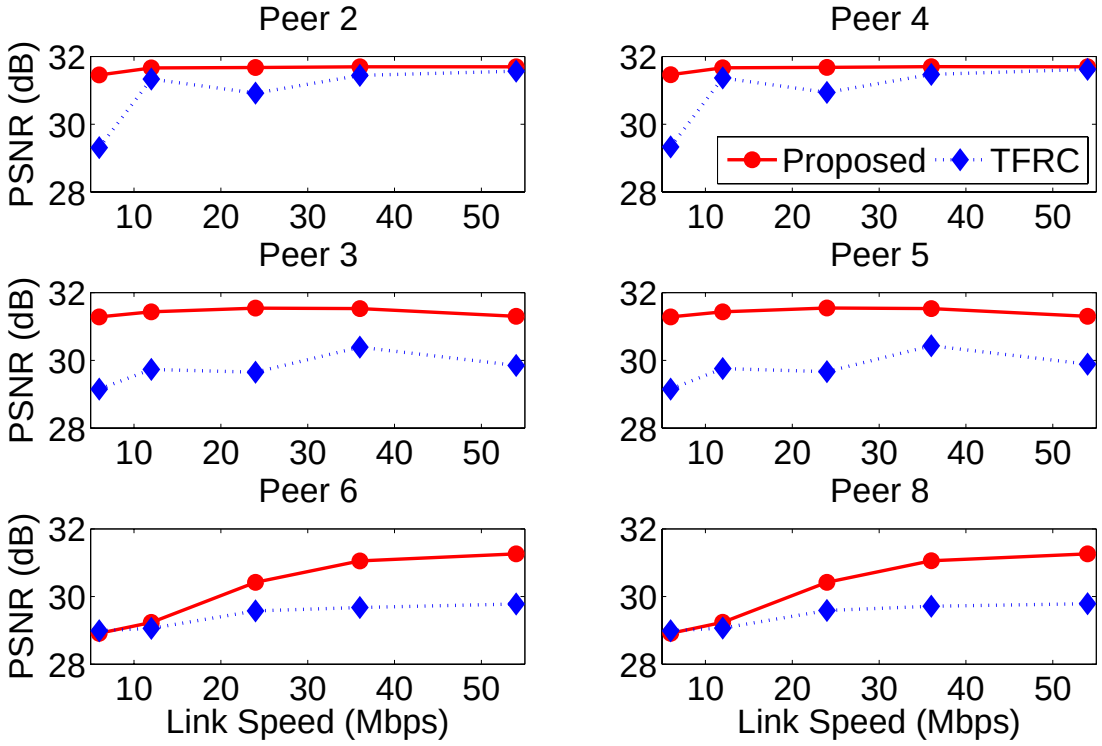


Figure 5. Video quality of each peer when the *Harbor* sequence is streamed over the first tree in Fig. 1. As the outgoing nominal link speed of Peer 5 varies from 54 Mbps to 6 Mbps, the video quality of all peers are reduced in the TFRC scheme. The proposed scheme, in comparison, maintains high video quality for the faster peers (2, 3, 4, 5) by reducing allocation only for the two peers (6 and 8) with lower link speeds.

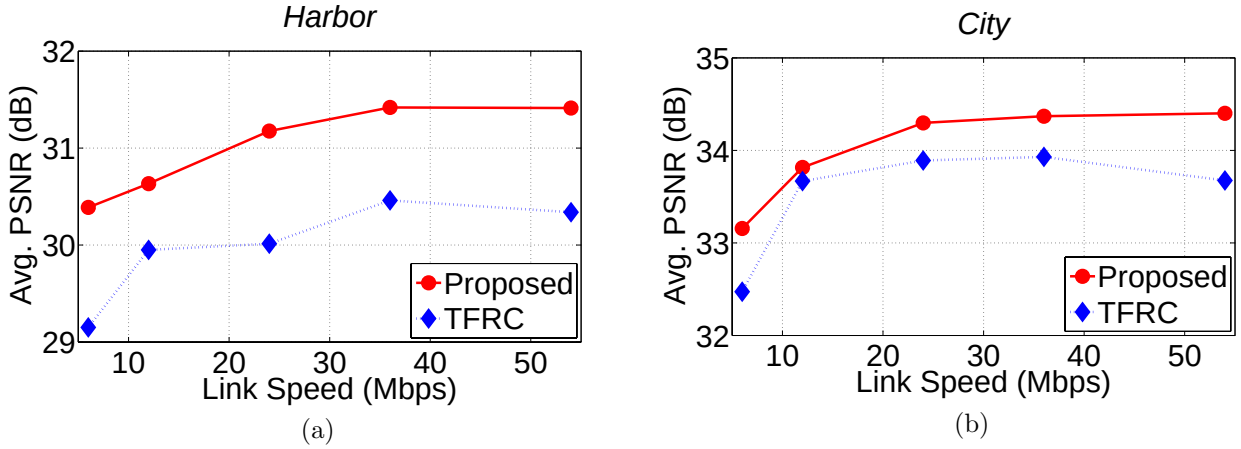


Figure 6. Average video quality measured as PSNR of average video distortion of all peers. The nominal link speed of Peer 5 varies from 6 Mbps to 54 Mbps. (a) *Harbor* streaming over the first ALM tree in Fig. 1; (b) Same experiment with *City*.

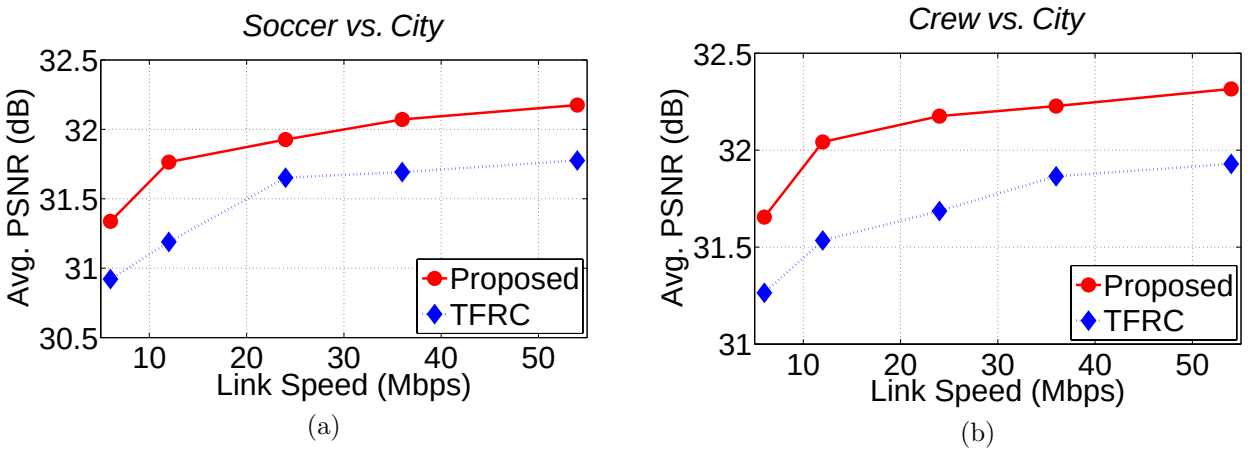


Figure 7. Average video quality measured as PSNR of average video distortion of all peers. The outgoing link speed of Peer 5 varies from 6 Mbps to 54 Mbps. (a) Experiment with *Soccer* over the first tree and *City* over the second tree; (b) Experiment with *Crew* over the first tree and *City* over the second tree.

REFERENCES

1. A. Ganjam and H. Zhang, "Internet multicast video delivery," *Proceedings of IEEE* **93**, pp. 159–170, Jan. 2005.
2. S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," *Proc. SIGCOMM'96* **26**, pp. 117–130, Aug. 1996.
3. M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. INFOCOM'03, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, **2**, pp. 836–843, (San Francisco, CA, U.S.A.), Mar. 2003.
4. X. Zhu and B. Girod, "Distributed rate allocation for multi-stream video transmission over ad hoc networks," in *Proc. IEEE International Conference on Image Processing (ICIP'05)*, **2**, pp. 157–160, (Genoa, Italy), 2005.
5. X. Zhu, P. van Beek, and B. Girod, "Distributed channel time allocation and rate adaptation for multi-user video streaming over wireless home networks," in *IEEE International Conference on Image Processing (ICIP'07)*, pp. 69–72, (San Antonio, TX, USA), 2007.
6. ITU-T and ISO/IEC JTC 1, *ITU-T Recommendation H.264 - ISO/IEC 14496-10(AVC), Advanced Video Coding for Generic Audiovisual services, Amendment 3: Scalable Video Coding*, 2005.
7. K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications* **18**, pp. 1012–32, June 2000.
8. *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, P802.11*, Nov. 1997.
9. Y. Nesterov and A. Nemirovsky, "Interior-point polynomial methods in convex programming," *Studies in Applied Mathematics, SIAM, Philadelphia, PA* **13**, 1994.
10. N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer Series in Computational Mathematics. Springer, New York, USA.
11. H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC," *IEEE Trans. Circuits and Systems for Video Technology* **17**(9), pp. 1103–1120, 2007.
12. "The Network Simulator NS-2." `http://www.isi.edu/nsnam/ns/`.
13. S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the Internet," *IEEE/ACM Trans. on Networking* **7**, pp. 458–472, Aug. 1999.