

Distributed Rate Allocation Policies for Multihomed Video Streaming Over Heterogeneous Access Networks

Xiaoqing Zhu, *Student Member, IEEE*, Piyush Agrawal, Jatinder Pal Singh, *Member, IEEE*, Tansu Alpcan, *Member, IEEE*, and Bernd Girod, *Fellow, IEEE*

Abstract—We consider the problem of rate allocation among multiple simultaneous video streams sharing multiple heterogeneous access networks. We develop and evaluate an analytical framework for optimal rate allocation based on observed available bit rate (ABR) and round-trip time (RTT) over each access network and video distortion-rate (DR) characteristics. The rate allocation is formulated as a convex optimization problem that minimizes the total expected distortion of all video streams. We present a distributed approximation of its solution and compare its performance against H^∞ -optimal control and two heuristic schemes based on TCP-style additive-increase-multiplicative-decrease (AIMD) principles. The various rate allocation schemes are evaluated in simulations of multiple high-definition (HD) video streams sharing multiple access networks. Our results demonstrate that, in comparison with heuristic AIMD-based schemes, both media-aware allocation and H^∞ -optimal control benefit from proactive congestion avoidance and reduce the average packet loss rate from 45% to below 2%. Improvement in average received video quality ranges between 1.5 to 10.7 dB in PSNR for various background traffic loads and video playout deadlines. Media-aware allocation further exploits its knowledge of the video DR characteristics to achieve a more balanced video quality among all streams.

Index Terms—Distributed rate allocation, heterogeneous access networks, multihomed video streaming.

I. INTRODUCTION

WITH the proliferation of broadband access technologies such as Ethernet, DSL, WiMax and IEEE 802.11a/b/g, portable devices tend to possess multiple modes of connecting to the Internet. Most PDAs provide both cellular and WLAN connectivity; laptops are typically equipped with a built-in Ethernet port, an 802.11a/b/g card and a phone jack for dial-up connections. Since a multitude of access technologies will continue

to coexist, increasing efforts are devoted to the standardization of architectures for network convergence. Integration of heterogeneous access networks has been a major consideration in the design of 4G networks [2], IEEE 802.21 [3], and the IP Multimedia Subsystem (IMS) platform [4]. In addition, multihomed Internet access presents an attractive option from an end-host's perspective. By pooling resources of multiple simultaneously available access networks, it is possible to support applications with higher aggregate throughput, lower latency, and better error resiliency [5].

In many applications, each end-host or device needs to simultaneously support multiple application flows with heterogeneous bit rate and latency requirements. One can easily imagine a corporate user participating in a video conference call, while uploading some relevant files to a remote server and browsing web pages for reference. In the presence of many such users, each access network can easily become congested with multiple competing application flows from multiple devices. The problem of resource allocation arises naturally, for determining the source rate of each application flow, and for distributing the traffic among multiple simultaneously available access networks. In this work, we focus on video streaming applications as they impose the most demanding rate and latency requirements. Flows from other applications, such as web browsing and file transfer, are treated as background traffic.

Challenges in the design of a rate allocation policy for such a system are multifold. Firstly, access networks differ in their attributes such as available bit rates (ABRs) and round-trip times (RTTs), which are time-varying in nature. Secondly, video streaming applications differ in their latency requirements and distortion-rate (DR) characteristics. For instance, a high-definition (HD) video sequence containing dynamic scenes from an action movie requires much higher data rate to achieve the same quality as a static head-and-shoulder news clip for a mobile device. Thirdly, unlike file transfer or web browsing, video streaming applications require timely delivery of each packet to ensure continuous media playout. Late packets are typically discarded at the receiver, causing drastic quality degradation of the received video due to error propagation at the decoder. In addition, the rate allocation policy should also operate in a distributed manner to avoid the traffic overhead and additional delay in collecting global media and network information for centralized computation.

Manuscript received August 28, 2008; revised February 19, 2009. First published April 28, 2009; current version published May 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Z. Jane Wang.

X. Zhu, P. Agrawal, and B. Girod are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: zhuxq@stanford.edu; piyushag@stanford.edu; bgirod@stanford.edu).

J. P. Singh is with the Deutsche Telekom Inc., R&D Lab, Los Altos, CA 94022 USA (e-mail: jatinder.singh@telekom.com).

T. Alpcan is with the Deutsche Telekom Laboratories, Berlin 10587, Germany (e-mail: tansu.alpcan@telekom.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2009.2017641

This paper addresses the above considerations, and investigates a suite of distributed rate allocation policies for multihomed video streaming over heterogeneous access networks.¹

- *Media-Aware Allocation*: When devices have information of both video DR characteristics and network ABR/RTT attributes, we formulate the rate allocation problem in a convex optimization framework and minimize the sum of expected distortions of all participating streams. A distributed approximation to the optimization is presented, to enable autonomous rate allocation at each device in a media- and network-aware fashion.
- *H[∞]-Optimal Control*: In the case where media-specific information is not available to the devices, we propose a scheme based on H[∞]-optimal control [6]. The scheme achieves optimal bandwidth utilization on all access networks by guaranteeing a worst-case performance bound characterizing deviation from full network utilization and excessive fluctuations in allocated video rates.
- *AIMD-Based Heuristic*: For comparison, we present two heuristic rate allocation schemes that react to congestion in the network by adjusting the total rate of each stream following TCP-style additive-increase-multiplicative-decrease (AIMD) principle [7]. They differ in their manners of how rates are split among multiple access networks in accordance with observed ABRs.

Performance of all four rate allocation policies are evaluated in ns-2 [8], using ABR and RTT traces collected from Ethernet, IEEE 802.11b and IEEE 802.11g networks in a corporate environment. Simulation results are presented for the scenario of simultaneous streaming of multiple high-definition (HD) video sequences over multiple access networks. We verify that the proposed distributed media-aware allocation scheme approximates the results from centralized computation closely. The allocation results react quickly to abrupt changes in the network, such as arrival or departure of other video streams. Both media-aware allocation and H[∞]-optimal control schemes achieve significantly lower packet delivery delays and loss ratios (less than 0.1% for media-aware allocation and below 2.0% for H[∞]-optimal control), whereas AIMD-based schemes incur up to 45% losses, far exceeding the tolerance level of video streaming applications. As a result, media-aware allocation improves the average received video quality by 1.5 to 10.7 dB in PSNR over the heuristic schemes in various simulation settings. It further ensures equal utilization across all access networks and more balanced video quality among all streams.

The rest of the paper is organized as follows. Section II briefly reviews related work in multiframe, multinet resource allocation. We present our system model of the access networks and expected video distortion in Section III, followed by descriptions of the rate allocation schemes in Section IV. Performances of the four schemes are evaluated in Section V via simulations

¹While our system model is general enough to accommodate best-effort networks such as the Internet, data service over cellular networks and 802.11 wireless home or corporate networks, it may not apply to scenarios where the service provider performs admission control or resource provisioning according to traffic load, e.g., in carrier grade WLAN networks or properly dimensioned UMTS networks for voice services. The extension of the current work to accommodate more general network types and behaviors is an interesting area of future research, and goes beyond the scope of this paper.

of three HD video streaming sessions sharing three access networks under various traffic conditions and latency requirements.

II. RELATED WORK

Rate allocation among multiple flows that share a network is an important and well-studied problem. Internet applications typically use the TCP congestion control mechanism for regulating their outgoing rate [7], [9]. For media streaming applications over UDP, TCP-friendly rate control (TFRC) is a popular choice [10], [11]. Several modifications have been proposed to improve its media-friendliness [12]. In [13], the problem of rate allocation among flows with different utilities is studied within a mathematical framework, where two classes of pricing-based distributed rate allocation algorithms are analyzed. In our work, the notion of utility of each flow corresponds to its expected received video quality, measured in terms of mean-squared-error (MSE) distortion relative to the original uncompressed video signals. We also extend the mathematical framework in [13] to consider rate allocation over multiple networks.

The problem of efficient utilization of multiple networks via suitable allocation of traffic has been explored from different perspectives. A game-theoretic framework to allocate bandwidth for elastic services in networks with fixed capacities is described in [14]–[16]. Our work, in contrast, acknowledges the time-varying nature of the network attributes and dynamically updates the allocation results according to observed available bit rates and round-trip delays. A solution for addressing the handoff, network selection, and autonomic computation for integration of heterogeneous wireless networks is presented in [2]. The work, however, does not address simultaneous use of heterogeneous networks and does not consider wireline settings. A cost-price mechanism is proposed for splitting traffic among multiple IEEE 802.11 access points to achieve end-host multihoming [17], [18]. The work does not take into account existence of other types of access networks or the characteristics of traffic, nor does it specify an operational method to split the traffic. In [5], a flow scheduling framework is presented for collaborative Internet access, based on modeling and analysis of individual end-hosts' traffic behavior. The framework mainly accounts for TCP flows and uses metrics useful for web traffic including RTT and throughput for making scheduling decisions.

Rate adaptation of multimedia streams has been studied in the context of heterogeneous networks in [19], where the authors propose an architecture to allow online measurement of network characteristics and video rate adaptation via transcoding. Their rate control algorithm is based on TFRC and is oblivious of the media content. In [20], media-aware rate allocation is achieved, by taking into account the impact of both packet loss ratios and available bandwidth over each link, on the end-to-end video quality of a single stream, whereas in [21], the rate allocation problem has been formulated for multiple streams sharing one wireless network. Unlike our recent work where the multistream multinet rate allocation problem is addressed from the perspective of stochastic control of Markov decision processes [22] and robust H[∞]-optimal control of linear dynamic systems [23], [24], in this paper we stay within the convex optimization framework for media-aware optimal rate allocation, and compare the

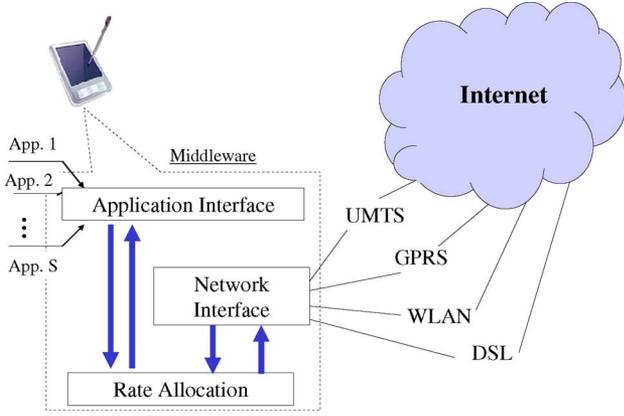


Fig. 1. Middleware functionality in a device. The rate allocation module collects the observed media statistics and network characteristics (e.g., ABR and RTT), and dictates the rate allocation among application streams, over each network interface.

performance of the scheme with prior approaches. Preliminary results from this work have been reported in [1] and [25].

III. SYSTEM MODEL

In this section, we introduce the mathematical notations used for modeling the access networks, and for estimating expected received video distortion of each stream. We envision a middleware functionality as depicted in Fig. 1, which collects characteristic parameters of both the access networks and video streams, and performs the optimal rate allocation according to one of the schemes described in Section IV. A more detailed discussion of the middleware functionality can be found in [25].

A. Network Model

Consider a set of access networks $\mathcal{N} = \{1, 2, \dots, N\}$, simultaneously available to multiple devices. Each access network n is characterized by its available bit rate c_n and round-trip time τ_n , which are measured and updated periodically. For each device, the set of video streams is denoted as $\mathcal{S} = \{1, 2, \dots, S\}$. Traffic allocation can be expressed in matrix form: $\mathbf{r} = \{r_n^s\}_{S \times N}$, where each element r_n^s corresponds to the allocated rate of Stream s over Network n . Consequently, the total allocated rate over Network n is $r_n = \sum_s r_n^s$, and the total allocated rate for Stream s is $r^s = \sum_n r_n^s$. We denote the *residual bandwidth* over Network n as

$$e_n = c_n - \sum_{s \in \mathcal{S}} r_n^s = c_n - r_n. \quad (1)$$

From the perspective of Stream s , the observed available bandwidth is

$$c_n^s = c_n - \sum_{s' \neq s} r_n^{s'}. \quad (2)$$

Note that $e_n = c_n - r_n = c_n^s - r_n^s$.

As the allocated rate on each network approaches the maximum achievable rate, average packet delay typically increases due to network congestion. We use a simple rational function to approximate the nonlinear increase of packet delay with traffic rate over each network as follows:

$$t_n = \frac{\alpha_n}{e_n} = \frac{\alpha_n}{c_n - r_n} = \frac{\alpha_n}{c_n^s - r_n^s}. \quad (3)$$

The value of α_n is estimated from past observations of τ_n and e_n , assuming equal delay on both directions as follows:²

$$\alpha_n = \frac{e_n \tau_n}{2}. \quad (4)$$

We note that despite oversimplification in this delay model, it is still effective in driving a rate allocation scheme with proactive congestion avoidance, as can be verified later by simulation results in Section V.

B. Video Distortion Model

Expected video distortion at the decoder comprises of the following two terms:

$$d_{dec} = d_{enc} + d_{loss} \quad (5)$$

where d_{enc} denotes the distortion introduced by lossy compression performed by the encoder, and d_{loss} represents the additional distortion caused by packet loss [26].

The distortion-rate (DR) characteristic of the encoded video stream can be fit with a parametric model [26] as follows:

$$d^s(r^s) = d_0^s + \frac{\theta^s}{(r^s - r_0^s)} \quad (6)$$

where the parameters d_0^s , θ^s , and r_0^s depend on the coding scheme and the content of the video. They can be estimated from three or more trial encodings using nonlinear regression techniques. To allow fast adaptation of the rate allocation to abrupt changes in the video content, these parameters are updated for each group of pictures (GOP) in the encoded video sequence, typically once every 0.5 s.

The distortion introduced by packet loss due to transmission errors and network congestion, on the other hand, can be derived from [27] as

$$d_{loss}^s = \kappa^s p_{loss}^s \quad (7)$$

where the sensitivity factor κ^s reflects the impact of packet losses p_{loss}^s , and depends on both the video content and its encoding structure. In general, packet losses are caused by both random transmission errors and overdue delivery due to network congestion. Since losses over the former type cannot be remedied by means of mindful rate allocation, we choose to omit their contribution in modeling decoded video distortion.

²For multihomed end hosts, acknowledgement packets for traffic sent over each network interface are returned over the same network. Therefore RTT is a good indication of network congestion, occurring either on the forward or backward path.

For simplicity, p_{loss}^s comprises solely of late losses due to network congestion in the rest of this paper.

IV. DISTRIBUTED RATE ALLOCATION

In this section, we address the problem of rate allocation among multiple streams over multiple access networks from several alternative perspectives. We first present a convex optimization formulation of the problem in Section IV-A, and explain how to approximate the media- and network-aware optimal solution with decentralized calculations. In the case that video DR characteristics are unavailable, we resort to a formulation of H^∞ -optimal control in Section IV-B, which dynamically adjusts the allocated rate of each stream according to fluctuations in observed network available bandwidth. For comparison, we include in Section IV-C two heuristic allocation schemes following TCP-style AIMD principle. All four schemes are distributed in nature, in that the rate allocation procedures performed by each stream does not need coordination or synchronization with other streams. Rather, interactions between the streams are *implicit*, as the ABRs and RTTs observed by one stream are affected by the allocated rates of other competing streams sharing the same interface networks.

A. Media-Aware Allocation

We seek to minimize the total expected distortion of all video streams sharing multiple access networks as follows:

$$\min_{\mathbf{r}} \sum_s d_{dec}^s(r^s, p_{loss}^s) \quad (8)$$

$$s.t. \quad r^s = \sum_n r_n^s, \quad \forall s \in \mathcal{S} \quad (9)$$

$$r_n = \sum_s r_n^s < c_n, \quad \forall n \in \mathcal{N} \quad (10)$$

$$r_n^s = \rho_n r^s, \quad \forall n \in \mathcal{N}. \quad (11)$$

In (8), the expected distortion d_{dec}^s is a function of the allocated rate d^s and average packet loss p_{loss}^s according to (5). The constraint (11) is introduced to impose uniqueness of the optimal solution. We choose $\rho_n = c_n / \sum_{n'} c_{n'}$ to ensure balanced utilization over each interface as follows:

$$\frac{r_n}{c_n} = \frac{\sum_s r_n^s}{c_n} = \frac{\rho_n \sum_s r^s}{c_n} = \frac{\sum_{n'} r_{n'}}{\sum_{n'} c_{n'}}, \quad \forall n \in \mathcal{N}. \quad (12)$$

It can also be shown that $\rho_n = c_n^s / \sum_n c_n^s, \forall s \in \mathcal{S}$. Each stream can therefore calculate the value of ρ_n independently, based on its own ABR observation c_n^s for Network n .

The average packet loss p_{loss}^s for each stream is the weighted sum of packet losses over all networks as follows:

$$p_{loss}^s = \sum_n \rho_n e^{-t_0^s/t_n}. \quad (13)$$

Following the derivations in [27], the percentage of late packets is estimated as $e^{-t_0^s/t_n}$, assuming exponential delay distributions with average t_n for Network n and playout deadline t_0^s for Stream s . Given (3), p_{loss}^s is expressed as

$$p_{loss}^s = \sum_n \rho_n e^{-t_0^s(c_n^s - r_n^s)/\alpha_n}. \quad (14)$$

Combining (5)–(14), it can be easily confirmed that the optimization objective is a convex function of the variable matrix \mathbf{r} . If all the observations and parameters were available in one place, the solution could be found by a suitable convex optimization method [28].

We desire to minimize the objective (8) in a distributed manner, with as little exchange of information among the devices as possible. One approach is to consider the impact of network congestion on one stream at a time, and alternate between the streams until convergence. From the perspective of Stream s , its contribution to (8) can be rewritten as

$$\begin{aligned} \min_{r^s} \quad & d^s(r^s) + \kappa^s \sum_n \rho_n e^{-t_0^s(c_n^s - r_n^s)/\alpha_n} \\ & + \sum_n \sum_{s' \neq s} \rho_n \kappa^{s'} e^{-t_0^{s'}(c_n^s - r_n^s)/\alpha_n} \\ s.t. \quad & r_n^s = \rho_n r^s, \quad \forall n \in \mathcal{N} \\ & r_n^s < c_n^s, \quad \forall n \in \mathcal{N}. \end{aligned} \quad (15)$$

In (15), optimization of rate allocation for Stream s requires knowledge of not only its own distortion-rate function $d^s(r^s)$ and packet loss sensitivity κ^s , but also its impact on the late loss of other streams via the parameters $\kappa^{s'}$ and $t_0^{s'}$. While each stream can obtain information regarding its own packet loss sensitivity and playout deadline, exchange of such information among different streams is undesirable for a distributed scheme.

We therefore further simplify the optimization to

$$\begin{aligned} \min_{r^s} \quad & d^s(r^s) + \sum_n \kappa' \rho_n e^{-t_0^s(c_n^s - r_n^s)/\alpha_n} \\ s.t. \quad & r_n^s = \rho_n r^s, \quad \forall n \in \mathcal{N} \\ & r_n^s < c_n^s, \quad \forall n \in \mathcal{N} \end{aligned} \quad (16)$$

where κ' is empirically tuned to control the scheme's aggressiveness. Even though (16) does not necessarily lead to an optimal solution for (8), it nevertheless incorporates considerations of both network congestion and encoder video distortion in choosing the optimal rates. The impact on other streams is captured implicitly by the second term in (16), reflecting congestion experienced by all streams traversing that network. Effectiveness of this distributed approximation will be verified later in Section V-B.

In essence, optimization of (16) involves a one-dimensional search of r^s and thus can be solved efficiently using numerical methods. Computational complexity of the scheme increases linearly with the number of competing streams S and the number of available access networks N , on the order of $\mathcal{O}(NS)$. In practice, each stream needs to track its observations of c_n^s 's and τ_n 's over all available access networks, and to observe its video DR parameters θ^s and r_0^s . At each time instance, the scheme would update its estimate of α_n according to (4). It then determines the allocated rate r^s by minimizing (16), and divides up the rate in proportion to ρ_n over respective networks. Fig. 2 summarizes these procedures.

B. H^∞ -Optimal Control

In the case when media-specific knowledge is unavailable to the wireless devices, the rate allocation problem can be ad-

Input: ABR and RTT measurements of available access networks c_n^s, τ_n ;
 video DR characteristic θ^s, r_0^s for the current GOP;

Parameters: Level of aggressiveness κ' ;

Output: Allocated rate r_n^s for each access network n ;

foreach Network n available to Stream s **do**
 Update estimate of α_n according to (4);
end

Update ρ_n as $c_n^s / \sum_n c_n^s$;

Update r^s to minimize (16);

foreach Network n available to Stream s **do**
 Update r_n^s as $\rho_n r^s$;

end

Fig. 2. Procedures of the media-aware allocation scheme run by Stream s .

dressed using H^∞ -optimal control [23]. In this approach, we track current and past observations of available bit rate (ABR) of each network, and model variations in ABR as unknown disturbances to a continuous-time linear system. The design goal is to achieve full network utilization while preventing excessive fluctuations in allocated video rates. An optimal rate controller is derived based on H^∞ -optimal analysis [6] to bound the *worst-case* system performance. The scheme is distributed by nature, in that it treats the dynamics of each stream as unknown disturbance for others, thereby decoupling interactions between different streams. Each stream estimates via various online measurement tools [29] the *measured residual bandwidth* as

$$w_n = \begin{cases} e_n, & \text{if } e_n \geq 0 \\ \mu(t_f - t_i), & \text{if } e_n < 0 \end{cases} \quad (17)$$

in which $e_n = c_n - r_n$ is defined by (1); t_i and t_f denote the initial and final time instance when e_n is negative and μ is a negative scaling constant.

We next define a continuous-time linear system from the perspective of a single stream keeping track of a single network. For notational simplicity we subsequently drop the subscript n and omit the time index t . The extension to multiple access networks is discussed in the Appendix. Since each stream is independent of others in the H^∞ -optimal control formulation, the scheme also generalizes immediately to the case with multiple streams [23], [24].

From the perspective of Stream s , its rate update system can be expressed as

$$\dot{x}^s = ax^s + bu^s + w \quad (18)$$

$$\dot{r}^s = -\phi r^s + u^s \quad (19)$$

where the system state variable x^s reflects roughly residual network bandwidth for Stream s and u^s represents the rate control action. In (18), the parameters $a < 0$ and $b < 0$ adjust the memory horizon and the expected effectiveness of control actions, respectively, on the system state x^s . A smaller value of a corresponds to a longer horizon, i.e., smoother values of x^s over time. A higher value of b means a more responsive system, where the rate control action of an individual stream has greater impact on total network utilization. In (19), the rate update is

Input: ABR measurements of available access networks;

Parameters: Stream-specific weighting parameters (a, b) and (h, g) ;

Output: Feedback control u^s and allocated rate r^s ;

foreach Access network available to Stream s **do**
 Measure current ABR (w) and delay;
 Update x^s according to (18);
 Compute u^s according to (21);
 Update rate r^s according to (19);
end

Fig. 3. Procedures of the H^∞ -optimal rate control scheme run by Stream s .

approximately in proportion to the control action, with $\phi > 0$ sufficiently small to guarantee stability [23], [24]. Recall that w is function of residual bandwidth e , which, in turn, is a function of aggregate rates from all video streams. Therefore the evolutions (18) and (19) are connected via a feedback loop.

Ideally, if the network is fully utilized at equilibrium, w is zero while u^s and x^s approach zero for ϕ sufficiently small. To prevent excessive fluctuations in the allocated rate of each video stream, however, fluctuations in the measured available bandwidth cannot be tracked perfectly. Design of the rate controller u^s therefore needs to balance the incentive for full network utilization against the risk of excessive fluctuation in allocated video rates. Such design objective can be expressed in mathematical terms, in the form of a cost function

$$L^s(x^s, u^s, w) = \frac{\|z^s\|}{\|w\|} \quad (20)$$

where $z^s := [hx^s \quad gu^s]^T$ denotes system output with user-specified weights $h > 0$ and $g > 0$ on relative importance of full network utilization and video rate smoothness. In (20), $\|z^s\|^2 := \int_0^\infty |z^s(t)|^2 dt$ and $\|w\|^2 := \int_0^\infty |w(t)|^2 dt$. The cost function captures the proportional change of the system output z^s with respect to system input w . Intuitively, when variations in the observed residual bandwidth w is large, larger variations are allowed in the allocated video rates.

From H^∞ -optimal control theory [6], one can choose the optimal rate controller as

$$u_\gamma^s(x) = -\left(\frac{b}{g^2} \sigma_\gamma\right) x^s \quad (21)$$

with $\sigma_\gamma = (-a \pm \sqrt{a^2 - \lambda h^2})/\lambda$ and $\lambda = 1/\gamma^2 - b^2/g^2$ to ensure a *worst-case* performance factor $\gamma := \sup_w L^s(u^s, w)$. The lowest possible performance factor is calculated as: $\gamma^* = [\sqrt{a^2/h^2 + b^2/g^2}]^{-1}$. In other words, for any given value of $\gamma > \gamma^*$, one can find an optimal rate controller according to (21) to ensure that in the worst case, the cost function (20) will not exceed γ .

Although analysis and controller design are conducted around the equilibrium point, the streams do not have to compute the actual equilibrium values. In practice, the H^∞ -optimal rate control scheme is implemented through the procedures summarized in Fig. 3. Similar as for media-aware allocation, computational complexity of the H^∞ -optimal control scheme

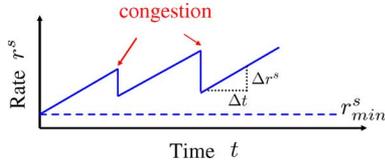


Fig. 4. Illustration of AIMD-based heuristic schemes. Allocated rate for each stream r^s keeps increasing at a rate of $\Delta r^s / \Delta t$ until congestion is detected from packet losses or excessive round-trip delay. In that case, the rate r_n^s is cut down by $(r_n^s - r_{min}^s) / 2$ over the congested network n .

scales linearly with number of competing streams and number of available access networks, on the order of $\mathcal{O}(NS)$.

C. AIMD-Based Heuristics

For comparison, we introduce in this section two heuristic rate allocation schemes based on the AIMD principle used by TCP congestion control [7]. Instead of performing proactive rate allocation by optimizing a chosen objective according to observed network attributes and video characteristics, the AIMD-based schemes are reactive in nature, in that they probe the network for available bandwidth and reduce the allocated rates only *after* congestion is detected.

As illustrated in Fig. 4, each stream initiates at a specified rate r_{min}^s corresponding to the minimum acceptable video quality, and increases its allocation by Δr^s every Δt seconds unless network congestion is perceived, in which case the allocated rate is dropped by $(r_n^s - r_{min}^s) / 2$ over the congested network n .

We consider two variations of the AIMD-based schemes. They differ in how the total allocated stream rate r^s is distributed across multiple access networks during the additive-increase phase.

- *Greedy AIMD*: The increase in rate allocation Δr^s is allocated to the network interface offering the maximum instantaneous available bit rate: $r_n^s = r^s$, if $c_n^s \geq c_{n'}^s, \forall n' \neq n \in \mathcal{N}$.
- *Rate Proportional AIMD*: The increase in rate allocation Δr^s is allocated to all available networks in proportion to their instantaneous available bit rates $r_n^s = (c_n^s / \sum_n c_n^s) r^s$.

In both schemes, congestion over Network n is indicated upon detection of a lost packet or when the observed RTT exceeds a prescribed threshold τ_{th}^s . The value of τ_{th}^s , in turn, is adjusted according to the video playout deadline.

V. PERFORMANCE EVALUATION

A. Simulation Methodology

Performance of all four rate allocation policies are evaluated in ns-2 [8], for an example network topology shown in Fig. 5. Each sender streams one HD video sequence via all three access networks to its receiver. Rate allocation over each network is determined by the middleware functionality depicted in Fig. 1. We collect available bit rate (ABR) and round-trip time (RTT) measurement from three real-world access networks (Ethernet, 802.11b and 802.11g) in a corporate environment

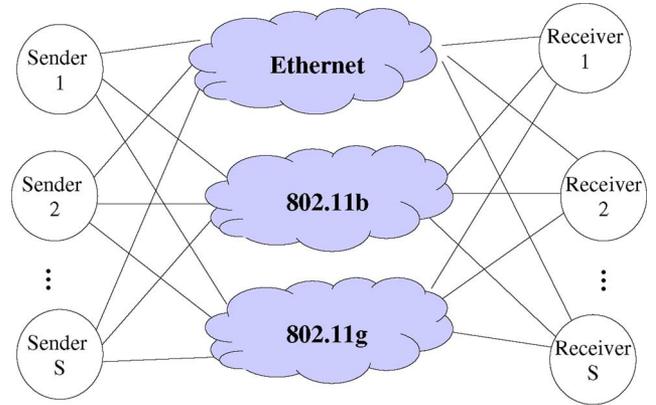


Fig. 5. Topology for network simulations.

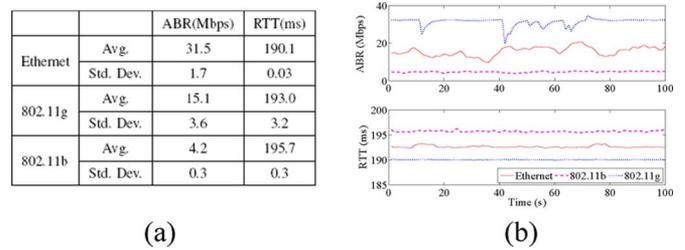


Fig. 6. (a) Statistics and (b) sample trace segments of measured available bit rate (ABR) and round-trip-time (RTT) from Deutsche Telekom Laboratories to Stanford University. The traces are collected over a two-hour duration in a weekday afternoon.

using Abing [29], [30].³ The ABR and RTT values are measured once every 2 s. The traces are then used to drive the capacity and delay over each simulated access network in ns-2.⁴ Statistics of the network measurement, together with a sample segment of the measured traces, are presented in Fig. 6. Fig. 7 shows how average packet delivery delay varies with utilization percentage over each access network, as well as sample packet delay distributions at a given utilization level. In all three interface networks, the average packet delay increases drastically as the utilization level approached 100%, as described in (3). In accordance with our assumptions, the example packet delay distributions also exhibit exponential shapes. We refer to [22] for further details of the trace collection procedures and bandwidth and delay measurements using Abing.

Three high-definition (HD) video sequences: *Bigships*, *Cyclists*, and *Harbor* are streamed by three senders, respectively. The sequences have spatial resolution of 1280×720 pixels, and temporal resolution of 60 frames per second (fps). Each stream is encoded using a fast implementation of the H.264/AVC standard [31], [32] at various quantization step sizes, with GOP length of 30 and IBBPBBP structure similar to that often used in MPEG-2 bitstreams. Fig. 8 shows the tradeoff of encoded video quality measured in MSE distortion and PSNR versus average

³In both 802.11b and 802.11g networks, the transmission rate over each interface is automatically adjusted according to wireless channel conditions. The effect of link rate adaptation is reflected in fluctuations in the ABR traces observed by Abing. Note that the rate allocation schemes under discussion only passively *react* to, instead of *interact* with, such fluctuations.

⁴Both forward and backward trip delays are simulated as half of measured RTTs.

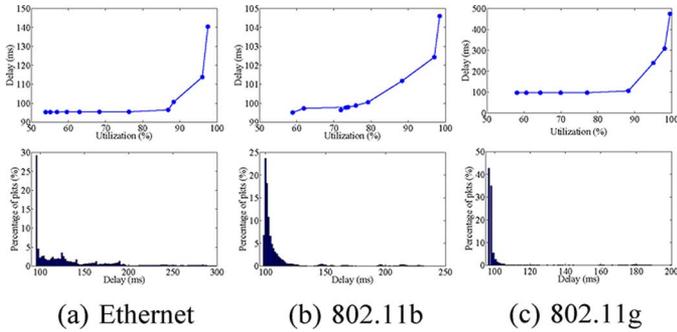


Fig. 7. (Top) Average packet delivery delay as function of network utilization, as well as (bottom) example packet delay distributions at a given utilization level, from the three interfaces used in ns-2 simulations. The packet delay distributions are plotted with a utilization level of 87%, 88%, and 88% over the three interfaces, respectively.

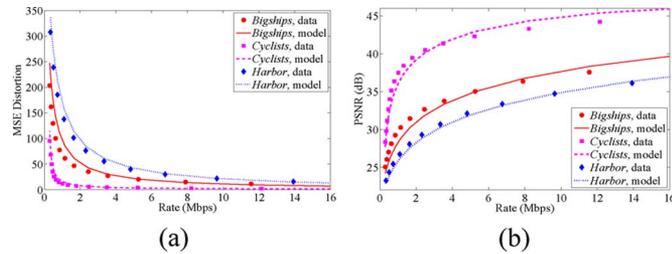


Fig. 8. (a) Rate-distortion and (b) rate-PSNR curves of three HD video sequences used in the experiments: *Bigships*, *Cyclists*, and *Harbor*, all encoded using the fast implementation of the H.264/AVC standard [32] at 60 fps, GOP length of 30. The measured data points obtained from encoding are plotted against model curves fitted according to (6).

bit rate over the entire sequence durations. The measured data points are plotted against fitted model curves according to (6). Encoded video frames are segmented into packets with maximum size of 1500 bytes. Transmission intervals of each packet in the entire GOP are spread out evenly to avoid unnecessary queuing delay due to the large sizes of intra-coded frames.

In addition to the video streaming sessions, additional background traffic is introduced over each network interface by the exponential traffic generator in ns-2. The background traffic rate varies between 10% and 50% of the total ABR of each access network. We also employ an implementation of the Abing agent in ns-2 to perform online ABR and RTT measurement over each access network for each stream. This allows the simulation system to capture the interaction among the three competing HD streams as they share the three access networks simultaneously. For consistency, measurement frequency of the Abing agents in ns-2 is also once every 2 s. Update of video rate allocation is in sync with the time instances when new network measurements are obtained for each stream. Note that no coordination or synchronization is required across rate updates in different streams, due to the distributed nature of the rate allocation schemes.

In the following, we first focus on the media-aware allocation scheme. Its allocation results are compared against optimal solutions for (15) in Section V-B and its convergence behavior is compared against H^∞ -optimal control in Section V-C. The performances of all four allocation schemes are evaluated

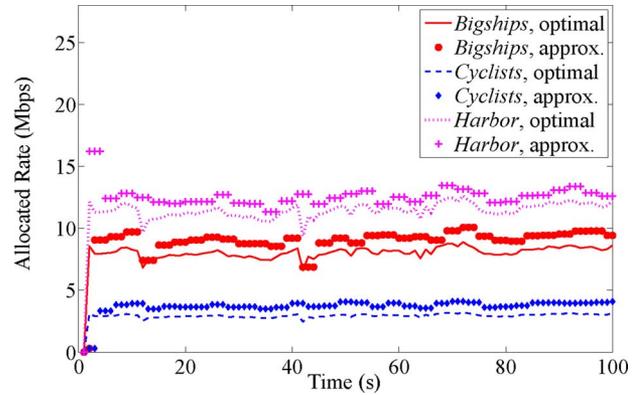


Fig. 9. Comparison of allocated rate to each video stream, from the optimal solution for (15) and its distributed approximation (16). Background traffic load is 20% and the playout deadline is 300 ms.

with 20% of background traffic load over each network and a playout deadline of 300 ms in Section V-D. Section V-E compares allocation results from networks with or without random packet losses. The impact of background traffic load on the allocation results obtained from different schemes is studied in Section V-F. The effect of different video streaming playout deadlines is investigated in Section V-G.

B. Comparison With Optimal Allocation

We first verify how well the distributed solution from (16) can approximate optimal solution for (15). Fig. 9 compares the traces of the allocated rates to each video stream calculated from both solutions. The value of κ' used in the distributed approximation corresponds to the sum of κ^s for all three streams: $\kappa' = \sum_s \kappa^s$. It can be observed that allocation from the distributed approximation tracks the optimal solution closely. Since the congestion term in (16) ignores the impact of a stream on the expected distortion of other streams, the distributed approximation achieves slightly higher rates.

C. Comparison of Convergence Behavior

Fig. 10 shows traces of allocated rate, when the number of competing streams over the three access networks increases from one to three. In this experiment, all three streams are the *Harbor* HD video sequence, hence the allocated rate to each stream is expected to be the same after convergence. The second and third streams start at 50 and 100 s, and complete at 200 and 250 s, respectively. Correspondingly, abrupt drops and rises in allocated rate can be observed in Fig. 10(a) for media-aware allocation. It is also interesting to note the fluctuations in the allocated rates after convergence, reflecting slight variations in the video contents and network attributes. The H^∞ -optimal control scheme, on the other hand, requires longer time for the allocation to converge, as shown in Fig. 10(b).

Next, we measure the allocation convergence times when one, two, or three competing streams join the network simultaneously. Convergence time is defined as the duration between the start of the streams and the time at which allocated video rates settle between adjacent quality levels. Fig. 11 compares results from media-aware allocation against H^∞ -optimal control. While both schemes yield similar allocated rates and video

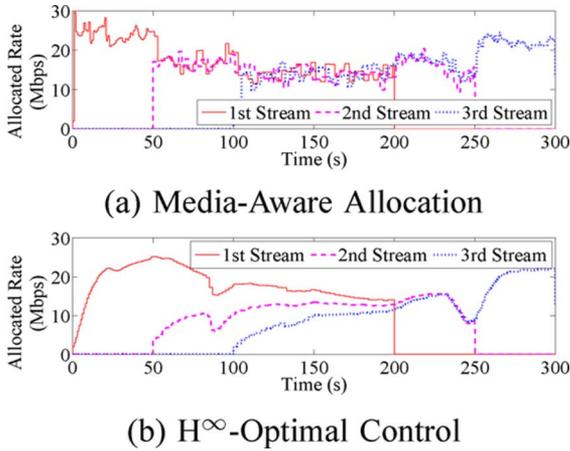


Fig. 10. Trace of total allocated rate to each stream, as the number of competing video streams, all *Harbor*, increases from one to three. Background traffic load is 20% and the playout deadline is 300 ms.

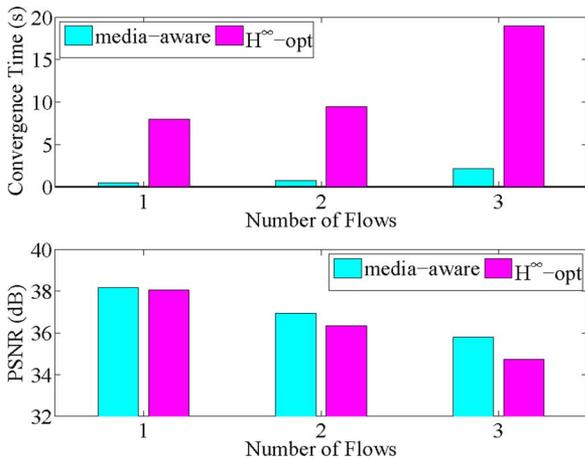


Fig. 11. Convergence time and corresponding average video quality, achieved by media-aware and H^∞ -optimal control schemes, with one, two, or three competing streams of the HD video sequence *Harbor*. Background traffic load is 20%; the playout deadline is 300 ms.

qualities, convergence time from media-aware allocation is shorter than H^∞ -optimal control.

D. Comparison of Allocation Traces

Fig. 12 plots the traces of aggregate rate allocated over the Ethernet interface for all four allocation schemes, together with the available bit rate over that network. It can be observed in Fig. 12(a) that media-aware allocation avoids much of the fluctuations in the two AIMD-based heuristics. Fig. 12(b) shows that it achieves higher network utilization than H^∞ -optimal control, as the latter is designed to optimize for the worst-case scenario. Similar observations also hold for traces of aggregate allocated rate over the other two interfaces.

In Fig. 13, we compare the traces of total allocated rate for each video stream, resulting from the various allocation schemes. In greedy AIMD allocation, the total rate of each stream increases until multiplicative decrease is triggered by either packet losses or increase in the observed RTTs from one of the interfaces. Therefore traces of the allocated rates bear a saw-tooth pattern. Behavior of the rate proportional AIMD

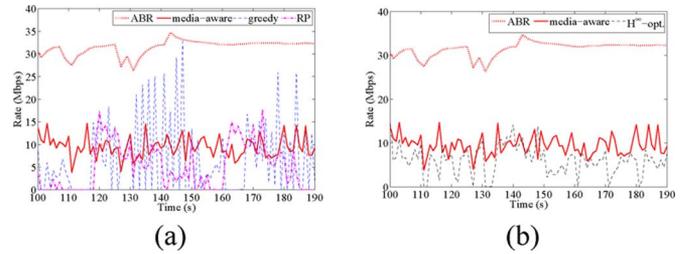


Fig. 12. Trace of aggregated rate over the Ethernet interface. (a) Media-aware allocation versus AIMD-based heuristics. (b) Media-aware allocation versus H^∞ -optimal control. In this experiment, background traffic load is 20% and the playout deadline is 300 ms. The network available bit rate is also plotted as a reference.

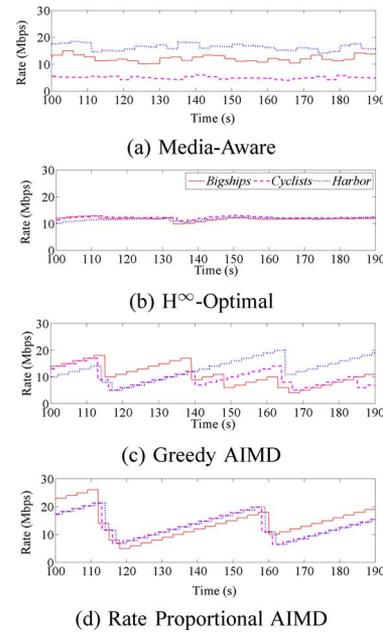


Fig. 13. Trace of allocated rate to each video stream, aggregated over three interfaces. Background traffic load is 20% and the playout deadline is 300 ms.

scheme is similar, except that rate drops tend to occur at around the same time. The H^∞ -optimal control scheme yields less fluctuations in the allocated rates. In both the rate proportional AIMD allocation and the H^∞ -optimal control schemes, allocated rates are almost identical to each video stream, since all flows are treated with equal importance. The media-aware convex optimization scheme, in contrast, consistently allocates higher rate for the more demanding *Harbor* stream, with reduced allocation for *Cyclists* with less complex contents.

E. Impact of Random Packet Loss

Fig. 14 compares the average utilization over each interface, allocated rate to each stream, and corresponding received video quality achieved by the four allocation schemes, for background traffic load of 30%. The media-aware scheme allocates lower rate for *Cyclists* and higher rate for *Harbor*, compared to the other schemes. This improves the video quality of *Harbor*, the stream with the lowest PSNR amongst the three, at the expense of reducing the quality of the less demanding *Cyclists*. Consequently, the video quality is more balanced among all three streams.

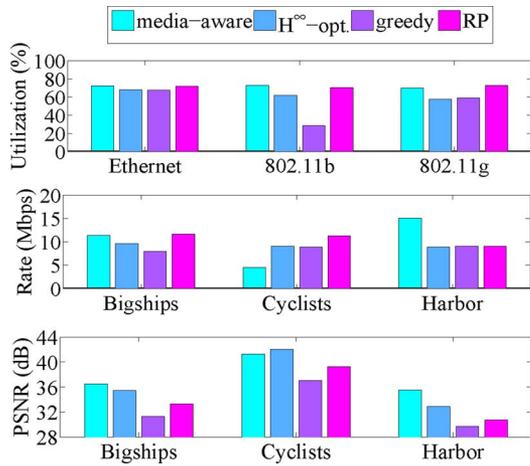


Fig. 14. Comparison of allocation results from different schemes with background traffic load of 30% and playout deadline of 300 ms. (Top) Aggregated network utilization over each interface. (Middle) allocated video rate for each stream. (Bottom) received video quality in PSNR.

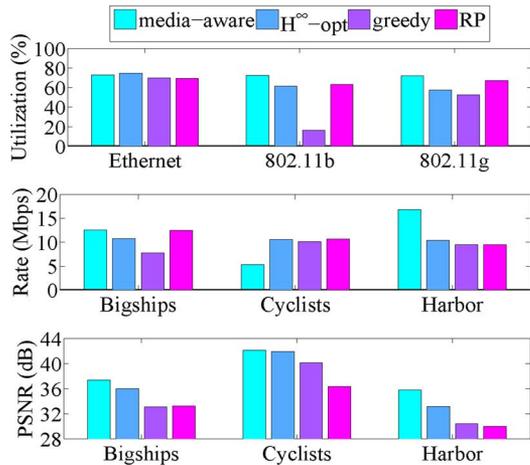


Fig. 15. Comparison of allocation results from different schemes with background traffic load of 30% and playout deadline of 300 ms. There is 1% of random packet loss over each interface. (Top) Aggregated network utilization over each interface. (Middle) allocated video rate for each stream. (Bottom) received video quality in PSNR.

A similar graph is shown in Fig. 15, for the same simulation with 1% random packet loss over each network interface. While the presence of random packet losses tends to reduce received video quality, its impact cannot be mitigated by means of careful rate allocation. Consequently, relative performance of the four rate allocation schemes remain the same in both scenarios. This justifies the absence of a term representing random packet losses when formulating the media-aware rate allocation problem. For the rest of the simulations, we therefore focus on comparisons without random packet losses.

F. Varying Background Traffic Load

Next, we vary the percentage of background traffic over each network from 10% to 50%, with playout deadline of 300 ms. The impact of the background traffic load on the allocation results is shown in Fig. 16. It can be observed that total utilization over each interface increases with the background traffic load.

For the media-aware, H[∞]-optimal and rate proportional AIMD schemes, utilization varies between 60% to 90%, whereas for the greedy AIMD scheme, the 802.11b interface is underutilized.⁵ Note that media-aware allocation ensures balanced utilization over all three access networks, as dictated by (12).

It can be observed from Fig. 16 that increasing background traffic load leads to decreasing allocated rate in each stream. While the other three schemes treat the three flows with equal importance, the media-aware allocation consistently favors the more demanding *Harbor*, thereby reducing the quality gap between the three sequences. The two AIMD-based heuristics achieve lower received video quality than media-aware and H[∞]-optimal allocations, especially in the presence of heavier background traffic load.

Fig. 17 compares the average packet delivery delay and packet loss ratios due to late arrivals. In the two AIMD-based schemes, allocated rates are reduced only *after* congestion has been detected. The media-aware allocation and the H[∞]-optimal control schemes, on the other hand, attempt to avoid network congestion in a proactive manner in their problem formulations. They therefore yield significantly lower packet loss ratios and delays. This leads to improved received video quality, as shown in Fig. 18. The performance gain ranges between 1.5 to 8.8 dB in PSNR of the decoded video, depending on the sequence content and background traffic load. Note also that the packet delivery delays and packet loss ratios also indicate the impact of each scheme on background traffic sharing the same access networks. Lower delays and losses achieved by the media-aware and H[∞]-optimal schemes means that they introduce less disruption to ongoing flows, as a result of proactive congestion avoidance.

G. Varying Playout Deadline

In the next set of experiments, we vary the playout deadline for each video stream from 200 ms to 5.0 s, while fixing the background traffic load at 20%. Fig. 19 compares the allocation results from the four schemes. As the playout deadline increases, higher network congestion level can be tolerated by each video stream. The media-aware allocation scheme therefore yields higher allocated rate and improved video quality, saturating as the playout deadline exceeds 1.0 s. Allocation from the other three media-unaware schemes, in comparison, are not so responsive to changes in the playout deadlines of the video streams.

Figs. 20 and 21 compare the average packet delivery delay and packet loss ratios due to late arrivals. Similar to results in the previous section, the media-aware and H[∞]-optimal allocations achieve much lower packet delivery delays and loss ratios than the two AIMD-based heuristics. The performance gap increases as the playout deadline becomes more relaxed. The packet loss ratios are almost negligible (less than 0.1%) from media-aware allocation, and very small (less than 2.0%) from H[∞]-optimal control. In comparison, the packet loss ratios range between 16%–45% for greedy AIMD, and between 12%–37% for rate proportional AIMD allocation, far exceeding the tolerance level of video streaming applications. Consequently, while

⁵Since 802.11b has significantly lower ABR than the other two interfaces, it is never chosen by the greedy AIMD scheme.

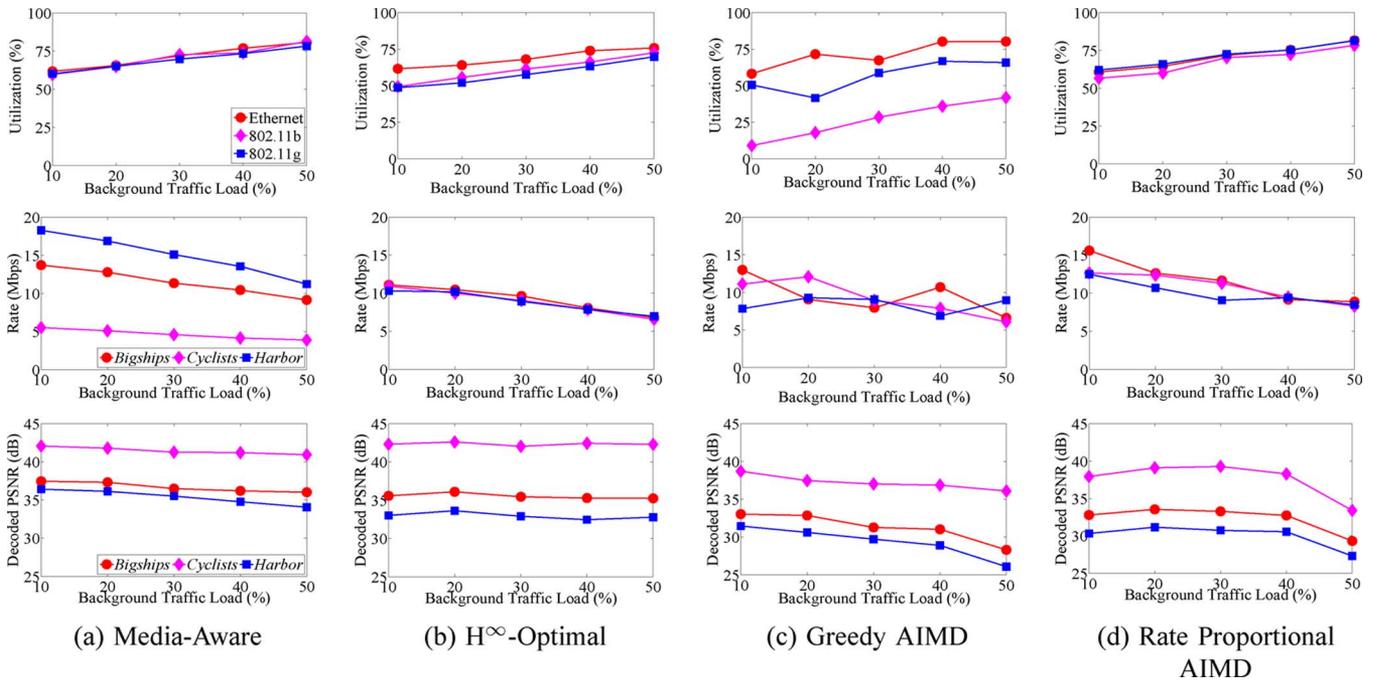


Fig. 16. Comparison of allocation results from different schemes, as the background traffic load increases.

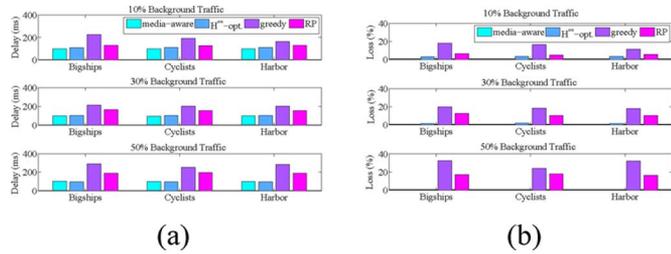


Fig. 17. (a) Average packet delivery delay and (b) packet loss ratio at the receiver for each video stream, with background traffic load at 10%, 30%, and 50%, respectively. The playout deadline is 300 ms.

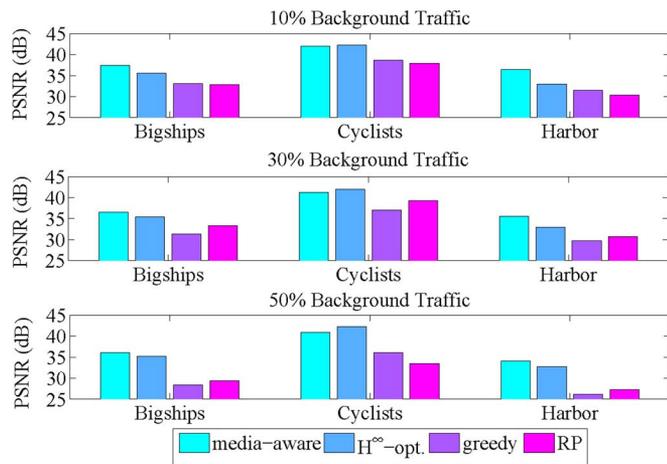


Fig. 18. Received video quality in PSNR for *Bigships*, *Cyclists*, and *Harbor*, for background traffic load at 10%, 30%, and 50%, respectively. The playout deadline is 300 ms.

the average received video quality of *Bigships* at playout deadline 300 ms is 34.0 dB and 32.8 dB from the greedy and rate proportional AIMD schemes, respectively, they are improved

to 37.3 dB with media-aware allocation, and to 36.0 dB with H^∞ -optimal control. Similar results are observed for other sequences with other playout deadlines, as shown in Fig. 22. The improvement varies between 3.3–10.7 dB in PSNR of the decoded video. The lower packet delivery delays and packet loss ratios achieved by the two proposed schemes also indicate that they are more friendly to ongoing background traffic than the two AIMD-heuristics, by virtue of more mindful congestion avoidance.

VI. CONCLUSIONS

This paper addresses the problem of rate allocation among multiple video streams sharing multiple heterogeneous access networks. We present an analytical framework for optimal rate allocation based on observed network attributes (available bit rates and round-trip times) and video DR characteristics, and investigate a suite of distributed rate allocation policies. Extensive simulation results demonstrate that both the media-aware allocation and H^∞ -optimal control schemes outperform AIMD-based heuristics in achieving smaller rate fluctuations, lower packet delivery delays, significantly reduced packet loss ratios and improved received video quality. The former benefit from proactive avoidance of network congestion, whereas the latter adjust the allocated rates only reactively, *after* detection of packet drops or excessive delays. The media-aware approach further takes advantage of explicit knowledge of video DR characteristics, thereby achieving more balanced video quality and responding to more relaxed video playout deadline by increasing network utilization.

We believe that this work has some interesting implications for the design of next generation networks in a heterogeneous, multihomed environment. Media-aware proactive rate allocation provides a novel framework for quality-of-service (QoS) support. Instead of rigidly reserving the network resources for

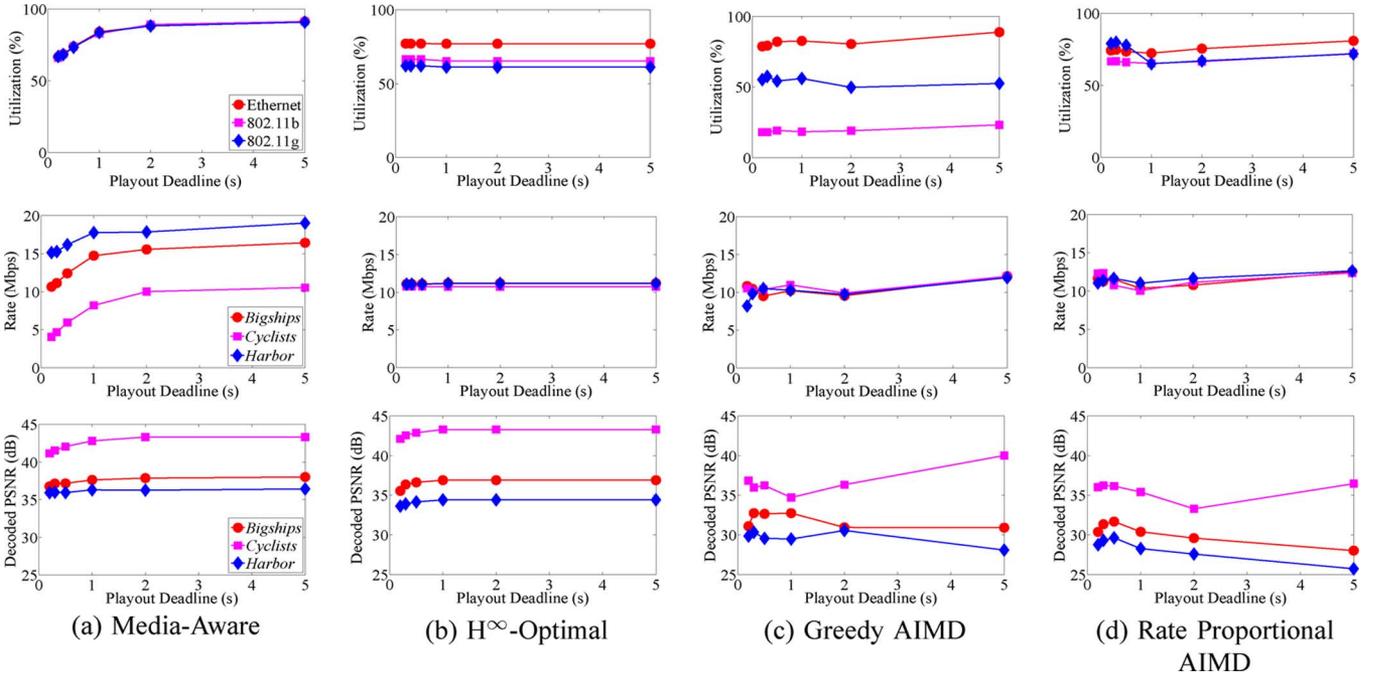


Fig. 19. Comparison of aggregated network utilization over each interface and allocated rate of each stream, as the playout deadline increases from 200 ms to 5.0 s. Background traffic load is 20%.

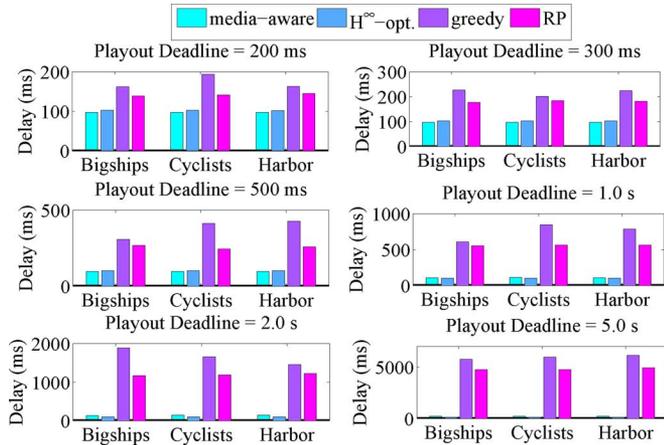


Fig. 20. Average packet delivery delays of each video stream for playout deadlines ranging from 200 ms to 5.0 s, with 20% background traffic load.

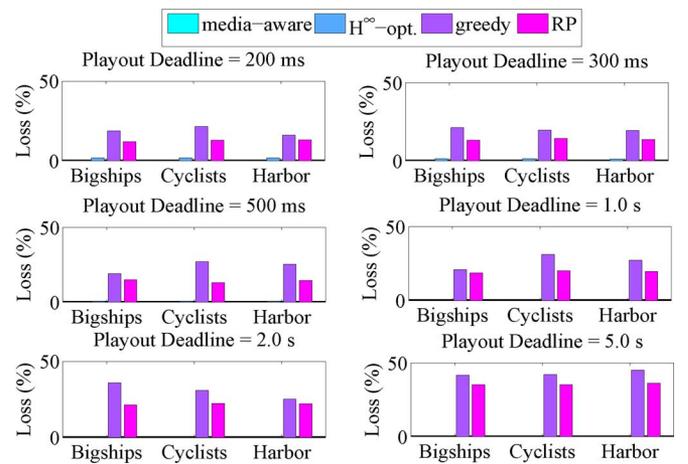


Fig. 21. Average packet loss ratios of each video stream for playout deadlines ranging from 200 ms to 5.0 s, with 20% background traffic load.

each application flow in advance, the allocation can be dynamically adapted to changes in network conditions and media characteristics. As the proposed rate allocation schemes are distributed in nature, they can be easily integrated into wireless devices. Future extensions to the current work include investigation of measures to best allocate network resources among different traffic types (e.g., web browsing versus video streaming) and to reconcile their different performance metrics (e.g., web page refresh time versus video quality) as functions of their allocated rates. In addition, our system model can be further extended to incorporate other types of access networks employing resource provisioning or admission control.

APPENDIX

We now provide the H[∞]-optimal control formulation for the general case of multiple access networks from the perspective of a single stream $s \in S$. For ease of notation, we drop the

superscript s and define $\mathbf{x} := [x_n]$, $\mathbf{r} := [r_n]$, and $\mathbf{u} := [u_n]$ for all $n \in \mathcal{N}$. The counterpart of the system (18) and (19) is given by

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{D}\mathbf{w} \\ \dot{\mathbf{r}} &= -\Phi\mathbf{r} + \mathbf{u} \end{aligned} \quad (22)$$

where $\mathbf{w} := [w_n] \quad \forall n$. Here, the matrices \mathbf{A} , \mathbf{B} , and Φ are obtained simply by multiplying the identity matrix by a , b , and ϕ , respectively.

Correspondingly, system output is

$$\mathbf{z} := \mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{u}. \quad (23)$$

The matrix \mathbf{H} represents the weight on the cost of deviation from zero state, i.e., full network utilization. We can assume that

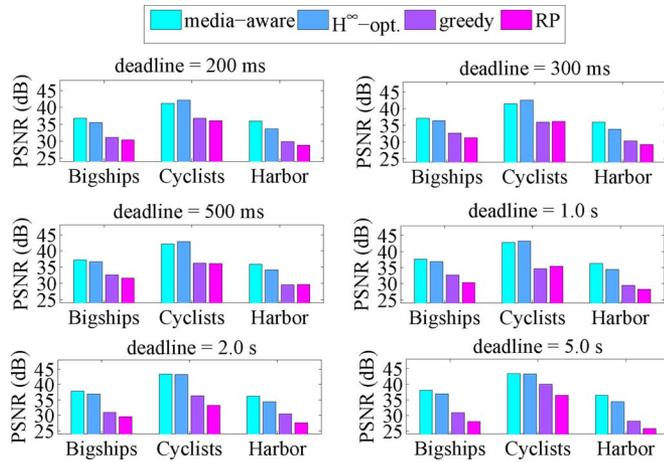


Fig. 22. Received video quality in PSNR of *Bigships*, *Cyclists*, and *Harbor* for playout deadlines ranging from 200 ms to 5.0 s, with 20% background traffic load.

$Q := H^T H$ is positive definite, in that any nonzero deviation from full utilization leads to a positive cost. Likewise, the matrix G represents the weight on the cost of deviation from zero control, i.e., constant allocated rates. We assume that $G^T G$ is positive definite, and that no cost is placed on the product of control actions and states: $H^T G = 0$.

The cost function is defined as

$$L(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \frac{\|\mathbf{z}\|}{\|\mathbf{w}\|} \quad (24)$$

where $\|\mathbf{z}\|^2 := \int_0^\infty |\mathbf{z}(t)|^2 dt$ and $\|\mathbf{w}\|^2 := \int_0^\infty |\mathbf{w}(t)|^2 dt$. Again, one can define the worst possible value for cost L as γ^* .

Similar to the solutions for the scalar system, we obtain the H^∞ -optimal linear feedback controller for the multiple network case as follows:

$$\mathbf{u} = -(G^T G)^{-1} B^T \Sigma_\gamma \mathbf{x}. \quad (25)$$

In (25), the matrix Σ_γ can be computed by solving the game algebraic Ricatti equation (GARE) as follows:

$$A^T Z + Z A - \Sigma (B (G^T G)^{-1} B^T - \gamma^{-2} D D^T) \Sigma + Q = 0. \quad (26)$$

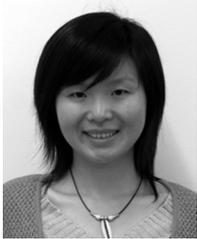
It can be verified that a unique minimal nonnegative definite solution Σ_γ exists for $\gamma > \gamma^*$, if (A, B) is stabilizable and (A, H) is detectable [6]. In our case, since the matrix B is square and negative definite and the matrix Q is positive definite, the system is both controllable and observable, hence both conditions are satisfied.

REFERENCES

[1] X. Zhu, P. Agrawal, J. P. Singh, T. Alpcan, and B. Girod, "Rate allocation for multi-user video streaming over heterogeneous access networks," in *Proc. ACM 15th Int. Conf. Multimedia*, 2007, pp. 37–46.
 [2] P. Vidales, J. Baliosion, J. Serrat, G. Mapp, F. Stejano, and A. Hopper, "Autonomic system for mobility support in 4G networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 12, pp. 2288–2304, Dec. 2005.
 [3] *IEEE 802.21*. [Online]. Available: <http://www.ieee802.org/21/>.
 [4] A. Cuevas, J. I. Moreno, P. Vidales, and H. Einsiedler, "The IMS platform: A solution for next generation network operators to be more than bit pipes," *IEEE Commun. Mag.*, Special Issue on Advances of Service Platform Technologies, vol. 44, no. 8, pp. 75–81, Aug. 2006.

[5] N. Thompson, G. He, and H. Luo, "Flow scheduling for end-host multihoming," in *Proc. 25th IEEE Int. Conf. Computer Communications (INFOCOM'06)*, Barcelona, Spain, Apr. 2006, pp. 1–12.
 [6] T. Basar and P. Bernhard, *H[∞]-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Boston, MA: Birkhäuser, 1995.
 [7] V. Jacobson, "Congestion avoidance and control," in *Proc. SIGCOMM'88*, Aug. 1988, vol. 18, pp. 314–329.
 [8] *NS-2*. [Online]. Available: <http://www.isi.edu/nsnam/ns/>.
 [9] M. Allman, V. Paxson, and W. R. Stevens, TCP Congestion Control, 1999, RFC 2581.
 [10] S. Floyd and K. Fall, "Promoting the use of end-to-end congestion control in the Internet," *IEEE/ACM Trans. Netw.*, vol. 7, no. 4, pp. 458–472, Aug. 1999.
 [11] M. Handley, S. Floyd, J. Padhye, and J. Widmer, TCP Friendly Rate Control (TFRC): Protocol Specification, 2003, RFC 3448.
 [12] Z. Wang, S. Banerjee, and S. Jamin, "Media-friendliness of a slowly-responsive congestion control protocol," in *Proc. 14th Int. Workshop Network and Operating Systems Support for Digital Audio and Video*, Cork, Ireland, 2004, pp. 82–87.
 [13] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, 1998.
 [14] H. Yaiche, R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 667–678, Oct. 2000.
 [15] T. Alpcan and T. Başar, "A utility-based congestion control scheme for Internet-style networks with delay," *IEEE Trans. Netw.*, vol. 13, no. 6, pp. 1261–1274, Dec. 2005.
 [16] T. Alpcan and T. Başar, "Global stability analysis of an end-to-end congestion control scheme for general topology networks with delay," in *Proc. 42nd IEEE Conf. Decision and Control (CDC'03)*, Maui, HI, Dec. 2003, pp. 1092–1097.
 [17] S. Shakkottai, E. Altman, and A. Kumar, "The case for non-cooperative multihoming of users to access points in IEEE 802.11 WLANs," in *Proc. IEEE INFOCOM'06*, Barcelona, Spain, Apr. 2006, pp. 1–12.
 [18] S. Shakkottai, E. Altman, and A. Kumar, "Multihoming of users to access points in WLANs: A population game perspective," *IEEE J. Select. Areas Commun.*, vol. 25, no. 6, pp. 1207–1215, Jun. 2007.
 [19] A. Szwabe, A. Schorr, F. J. Hauck, and A. J. Kessler, "Dynamic multimedia stream adaptation and rate control for heterogeneous networks," in *Proc. 15th Int. Packet Video Workshop (PV'06)*, Hangzhou, China, May 2006, vol. 7, pp. 63–69.
 [20] D. Jurca and P. Frossard, "Media-specific rate allocation in heterogeneous wireless networks," in *Proc. 15th Int. Packet Video Workshop (PV'06)*, Hangzhou, China, May 2006, vol. 7, pp. 713–726.
 [21] X. Zhu, J. P. Singh, and B. Girod, "Joint routing and rate allocation for multiple video streams in ad hoc wireless networks," in *Proc. 15th Int. Packet Video Workshop (PV'06)*, Hangzhou, China, May 2006, vol. 7, pp. 727–736.
 [22] J. P. Singh, T. Alpcan, P. Agrawal, and V. Sharma, "An optimal flow assignment framework for heterogeneous network access," in *Proc. IEEE Int. Symp. World of Wireless, Mobile and Multimedia Networks (WoWMoM'07)*, Helsinki, Finland, Apr. 2007, pp. 1–12.
 [23] T. Alpcan, J. P. Singh, and T. Basar, "A robust flow control framework for heterogeneous network access," in *Proc. 5th IEEE Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'07)*, Limassol, Cyprus, Jun. 2007, pp. 1–8.
 [24] T. Alpcan, J. P. Singh, and T. Basar, "Robust rate control for heterogeneous network access in multi-homed environments," *IEEE Trans. Mobile Comput.*, to be published.
 [25] J. P. Singh, T. Alpcan, X. Zhu, and P. Agrawal, "Towards heterogeneous network convergence: Policies and middleware architecture for efficient flow assignment, rate allocation and rate control for multimedia applications," in *Proc. Workshop Middleware for Next-Generation Converged Networks and Applications (MNCNA'07)*, Newport Beach, CA, Nov. 2007.
 [26] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 1012–32, Jun. 2000.
 [27] X. Zhu, E. Setton, and B. Girod, "Congestion-distortion optimized video transmission over ad hoc networks," *EURASIP J. Signal Process.: Image Commun.*, vol. 20, no. 8, pp. 773–783, Sep. 2005.
 [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
 [29] [Online]. Available: <http://www-iepm.slac.stanford.edu/tools/abing/>.

- [30] J. Navratil and R. L. Cottrell, "ABwE: A practical approach to available bandwidth estimation," in *Proc. Passive and Active Measurement (PAM) Workshop*, La Jolla, CA, 2003.
- [31] Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 – ISO/IEC 14496–10(AVC), ITU-T and ISO/IEC JTC 1, 2003.
- [32] x.264. [Online]. Available: <http://developers.videolan.org/x264.html>



Xiaoqing Zhu (S'01) received the B.E. degree in electronics engineering in 2001 from Tsinghua University, Beijing, China, and the M.S. degree in electrical engineering in 2002 from Stanford University, Stanford, CA, where she is currently pursuing the Ph.D. degree.

She is recipient of the Stanford Graduate Fellowship from 2001 to 2005 and has worked on several research projects granted by the National Science Foundation (NSF). For the summer of 2003, she was with I.B.M. Almaden Research Center, working on

intelligent document extraction and analysis. She spent the summer of 2006 interning at Sharp Labs of America, working on distributed resource allocation for video streaming over wireless networks, which remains her current research interest.



Piyush Agrawal received the M.S. and B.S. degrees in computer science and engineering from the Indian Institute of Technology (IIT), Kanpur, India, in 2008. He is pursuing the Ph.D. degree at the Department of Electrical Engineering at Stanford University, Stanford, CA.

He is currently working in the areas of multiview videos, free view point TV, and P2P media streaming. He was the recipient of the cadence gold medal for the best masters thesis across all engineering departments at IIT Kanpur in 2008. He also coauthored a

paper which received the best student paper award at the ACM Multimedia conference in September 2007.



Jatinder Pal Singh (M'05) received the B.S. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 2000, and the M.S. and Ph.D. degrees in electrical engineering and M.S. from Stanford University, Stanford, CA, in 2002 and 2005, respectively.

He is the Director of Research at Deutsche Telekom Inc. R&D Lab, Los Altos, CA, and a Consulting Assistant Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA. He is overseeing research and development

in the areas of clean slate internet design, services and mobile platforms, and multimedia communications and systems. He was a Senior Research Scientist with Deutsche Telekom Laboratories, Berlin, Germany, from 2006-2007. He was awarded the Stanford Graduate Fellowship and the Deutsche Telekom Fellowship. He graduated at the top of his class at the Indian Institute of Technology with the Institute Silver Medal and the A. K. Mahalanabis memorial scholarship. He worked with IBM T. J. Watson Research Center, Hawthorne, NY; Robert Bosch Corporation, Palo Alto, CA; Intel Corporation, San Jose, CA; and IBM India Research Labs, New Delhi, as an intern. He has been named Who's Who in Science and Engineering by Marquis Who's Who Publishers in 2008. His current research focus is on next generation fixed and mobile infrastructure and services, resource allocation for heterogeneous network access, and media streaming over wireless and peer to peer networks. He has published in and served as a reviewer for prominent journals and conferences in networking and communications. He has authored patents (pending) in the areas of load balancing, QoS provisioning and distributed authentication mechanisms for broadband access sharing, TCP throughput optimization, central scheduling algorithms for IEEE 802.11e networks, peer-to-peer media streaming, and heterogeneous network access.

Dr. Singh is a member of ACM.



Tansu Alpcan (S'98–M'07) received the B.S. degree in electrical engineering from Bogazici University, Istanbul, Turkey, in 1998. He received the M.S. and Ph.D. degrees in electrical and computer engineering from University of Illinois at Urbana-Champaign (UIUC) in 2001 and 2006, respectively.

His research interests include network security, game theory, control and optimization of wired and wireless communication networks, and resource allocation. He is currently a senior research scientist in Deutsche Telekom Laboratories, which is affiliated with Technische Universität Berlin, Berlin, Germany.

Dr. Alpcan has received Fulbright scholarship in 1999 and best student paper award in the IEEE Conference on Control Applications in 2003. He has authored more than 60 journal and conference articles. He was an associate editor for the IEEE Conference on Control Applications (CCA) in 2005 and has been a TPC member of several conferences including IEEE Infocom in 2007–2009. He has received the Robert T. Chien Research Award from the UIUC Department of Electrical and Computer Engineering and the Ross J. Martin Research Award from the UIUC College of Engineering in 2006. He has cochaired the Workshop on Game Theory in Communication Networks (GameComm) in 2008 and been the publicity chair of GameNets 2009 conference.



Bernd Girod (M'80–SM'97–F'98) received the M.S. degree from the Georgia Institute of Technology, Atlanta, and the Engineering Doctorate from the University of Hannover, Hannover, Germany.

He has been a Professor of electrical engineering and (by courtesy) computer science in the Information Systems Laboratory of Stanford University, Stanford, CA, since 1999. Previously, he was a Professor of telecommunications in the Electrical Engineering Department of the University of Erlangen-Nuremberg. His current research interests

are in the areas of video compression and networked media systems. He has published over 400 conference and journal papers, as well as five books. As an entrepreneur, he has been involved with several startup ventures as founder, director, investor, or advisor, among them Polycom (Nasdaq: PLCM), Vivo Software, 8x8 (Nasdaq: EGHT), and RealNetworks (Nasdaq: RNWK).

Prof. Girod is a EURASIP Fellow and a member of the German National Academy of Sciences (Leopoldina). He received the EURASIP Signal Processing Best Paper Award in 2002, the IEEE Multimedia Communication Best Paper Award in 2007, the EURASIP Image Communication Best Paper Award in 2008, as well as the EURASIP Technical Achievement Award in 2004