Towards Heterogeneous Network Convergence: Policies and Middleware Architecture for Efficient Flow Assignment, Rate Allocation and Rate Control for Multimedia Applications*

Jatinder Pal Singh
Deutsche Telekom AG, Laboratories
Ernst-Reuter-Platz 7
10587, Berlin, Germany
jatinder.singh@telekom.de

Tansu Alpcan
Deutsche Telekom AG, Laboratories
Ernst-Reuter-Platz 7
10587, Berlin, Germany
tansu.alpcan@telekom.de

Xiaoqing Zhu
Information Systems Laboratory
Stanford University
Stanford, CA 94305
zhuxq@stanford.edu

Piyush Agrawal
Department of Computer Science and Engineering
Indian Institute of Technology
Kanpur, India
piyushag@iitk.ac.in

ABSTRACT

The convergence in access networks brings forth the challenge of resource allocation for applications with varying utilities over networks with time-varying and heterogeneous characteristics. Owing to an ever increasing demand and popularity, multimedia applications have been a key driving force behind the efforts targeted towards architectures, industrial standards, and polices for convergence. Multimedia applications use protocols and transport dynamics that constrain the manner in which these applications can utilize access networks. In this work, we propose middleware architecture and policies that target optimal rate allocation for multimedia applications by monitoring and stochastically modeling the variation in characteristics of the access networks, using the utilities of the media application when known, and accounting for the constraints imposed by application specific transport mechanisms. The proposed measures leverage our optimal rate allocation, rate control, and interface assignment techniques [17, 6, 24] to achieve enhanced multimedia performance and efficient utilization of the access networks.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems; C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

General Terms

Design, Algorithms

*Research sponsored by Deutsche Telekom AG, Laboratories.

1. INTRODUCTION

Various developments over the past decade have contributed to an increasing drive towards convergence. Firstly, there has been the growth and evolution of several access technologies that are diverse in provisioned data rates, coverage ranges, and the nature of supported services. The fixed-line technologies (DSL, cabel modems, LANs etc.) have continued to provide Internet access in a variety of settings including residential areas. While the cellular infrastructure has traditionally catered to the voice services, it has also been providing data, multimedia services, and Internet access via existing and evolving standards and technologies [3]. While the wireless local area networks (WLANs) have primarily catered to the wireless access coverage over small ranges. Wireless mesh networks and metropolitan area access technologies [2] are also emerging to provision last mile broadband wireless access. The existence and popularity of heterogeneous access technologies has created the need for a framework where these can be intergrated and utilized efficiently.

Secondly, as applications have varying bandwidth requirements and access networks provide varying bit rates and quality of access at different sites, different networks may be of best utility at different times. Consequently multimode devices have been introduced and have become popular. Most personal computers today have a built in wireless card, a PCMCIA slots and an ethernet port. PDAs with cellular and WLAN connectivity are growing in use. The emergence of such devices necessitates their seamless operation across heterogeneous access network settings.

Thirdly, the cellular networks and Internet infrastructures are expanding to deliver voice, data, and multimedia servies. Although the data services and voice have traditionally been provisioned by Internet infrastructure and cellular networks respectively, the need for Internet access on the move and services like messaging and multimedia have made data and multimedia service provisioning an integral part of cellular servies. On the other hand, with increasing reliability, voice and multimedia services over the Internet have become popular due to low tariffs over the cellular conterpart. The progress of these two segments of infrastructures serving the same needs for the users warrants interoperability and convergence of user devices and services. For carriers with both fixed-line and mobile business, convergence entails allowing their users to utilize the voice, data and multimedia services seamlessly on multiple devices in a way that the user quality of experience is enhanced, while avoiding the threat posed by small carriers and operators offering voice and data services at competitive prices.

Due to these factors and with the realization that the heterogenous access technologies will continue to co-exist, the drive towards convergence has been gaining momentum [7, 1, 20]. Mutimedia applications continue to constitute a key area of focus in all these efforts. High definition video availability at homes, IP and peer-to-peer TV, mobile multimedia services, multimedia conferencing, constitute only a part of the whole gamut of services driving the trend towards convergence. With potential high bit rate and delay requirements, multimedia applications add another dimension of challenge to already non-trivial problem of resource allocation over heterogeneous networks. While platforms and architectures allowing video streaming applications to benefit from multi-homing capabilities have been designed, it remains to identify and evaluate measures that allow optimal utilization of multiple access technologies.

The challenges for the scenario are multi-fold. Firstly the characteristics (e.g. available bit rate and delay) of the networks are heterogeneous and time variant. For instance our measurements conducted to evaluate the available bit rate between Deutsche Telekom Laboratories in Berlin to Stanford University over Ethernet, IEEE 802.11g, and IEEE 802.11b, as summarized in Table 1, indicate a considerable difference in available bit rates and variation in these rates over time (Figure 1). The complete set of measurements which may be seen in our work [17], also show substantial variation in RTT between close to moderately-far destinations. Secondly, multimedia applications differ in their latency requirements and distortion-rate (DR) characteristics. For instance, streaming a high definition video sequence containing scenes from an action movie would require much higher data rate to achieve the same quality as that needed for streaming a static head-and-shoulder news clip for a mobile device with a low resolution display. Video streaming applications also require timely delivery of packets, failing which the packets have to be discarded by the receiver causing video quality deration due to error propagation at the decoder. Thirdly, the system design and networking challenges for enabling a multi-homed multimedia steaming system need to be addressed. For instance, as the networks that a multi-homed device has access to are typically heterogeneous in characteristics, if the underlyting transport layer used for multimedia streaming is TCP the packet reordering due to data transfer via different access networks can result in poor performance. A suitable design of a middleware that is cognizant of the needs and requirements of multimedia application and also the characteristics of access networks that a multi-homed streaming system has access to is needed.

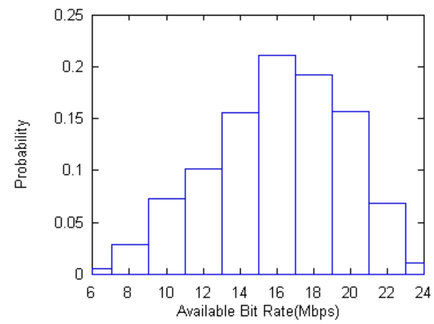


Figure 1: Available bit rate on 802.11g from T-Labs to Stanford

Table 1: Available bit rate and RTT from T-Labs to Stanford University

		ABR(Mbps)	RTT(ms)
Ethernet	Avg.	31.5	190.1
	Std. Dev.	1.7	0.03
802.11g	Avg.	15.1	193.0
	Std. Dev.	3.6	3.2
802.11b	Avg.	4.2	195.7
	Std. Dev.	0.3	0.3

In this work we propose streaming system and middleware design, and policies that allow efficient tranmission of multimedia in a setting where the multimedia system has access to networks with heterogeneous characteristics. We begin by recognizing that based on different transport layer dynamics employed to transport the multimedia content, different applications warrant different handling. Also different applications may differ in their utilities that they can draw from allocated network resources (e.g. bit rate). We propose the architecture and functionality of a middleware that when adopted by a multimedia system allows efficient transmission of multimedia content resulting in enhanced perceived user quality at the receiver end, and also better utilization of the access networks. For the case when the application data cannot be striped over multiple interfaces, we adopt the approach in our work [17] to stochastically characterize the variation in the characteristics of the access networks and maximize the utility of an application by assigning it to the optimally suited interface. For the scenario where the application data can be striped over multiple networks, we use our policies in our works [24, 6] to perform optimal rate allocation and rate control for the cases when the middleware does and does not have access to the DR characteristics of the multimedia application, by respectively utilizing a distributed approximation to a convex optimization framework that minimizes the expected distortions of all multimedia streams using the access networks, and H^{∞} optimal rate control. We also propose functional architecture, components and algorithms for such a middleware functionality.

The rest of this paper is organized as follows. We discuss related work in Section 2. We present network and middleware architecture in Section 3, and also delineate the network and system support required to have the middleware functionality. We then discuss rate allocation policies used by the middleware in Section 4. The algorithms for working of the middleware are presented in Section 5 and the work is concluded in Section 6.

2. RELATED WORK

The problem of efficient utilization of multiple networks via suitable allocation of traffic flows has been explored in the past from different perspectives. A game theoretic framework for bandwidth allocation for elastic services in networks with fixed capacities has been addressed in [23, 5, 4]. Our work, in comparison considers a scenario with access networks having time varying characteristics. In [16], a cost price mechanism is proposed to enable a mobile device to split its traffic amongst several IEEE 802.11 access points based on throughput obtained and price charged. However, the work does not take into account the existence of heterogeneous networks or the characteristics of traffic, and does not specify an operational method to split the traffic, which are addressed in this paper. A solution for addressing the handoff, network selection, and autonomic computation for integration of heterogeneous wireless networks has been presented in [20]. The work, however, does not address efficient simultaneous use of heterogeneous networks and does not consider wireline settings. Flow scheduling for collaborative Internet access in residential areas via multihomed client devices is discussed in [19]. The scheduling framework proposed in the work only 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 [18], where the authors propose an architecture to allow online measurement of network characteristics and video rate adaptation via transcoding. The rate control algorithm, on the other hand, is based on TFRC and unaware of the media content. In [10], media-aware rate allocation is achieved, by taking into account the impact of both packet loss rates and available bandwidth over each link, on the end-to-end video quality of a single stream, whereas in [25], the rate allocation problem has been formulated for multiple streams sharing one wireless network.

Rate allocation among multiple traffic flows over shared network resources is an important and well-studied problem. Internet applications typically use the TCP Congestion Control mechanism for regulating the outgoing rate [9][13]. For media streaming applications over UDP, TCP-Friendly Rate Control (TFRC) is a popular choice [8][14]. While protocols such as TCP and TFRC assume common utility of all traffic flows in the network, rate allocation for flows with different utility has also been formulated as a convex optimization problem, and can be solved using distributed solutions [11, 12]. In our work [24], the adopt the utility of each traffic flow as its expected received video quality measured in terms of mean-square-error distortion from the original uncompressed video signals, and developed the mathematical framework to target rate allocation over multiple networks simultaneously.

3. ARCHITECTURAL REQUIREMENTS

In this section we delineate the network and system architecture requirements for supporting multimedia applications over heterogeneous networks. We observe that applications using different transport mechanisms may warrant specific requirements for system design and network resource provisioning.

3.1 Heterogeneous transport mechanisms

- Multimedia applications that use transport mechanisms which warrant adherence to end-to-end semantics for in-order delivery of packets may not be able to split over heterogeneous interfaces. For instance, streaming using HTTP, or RTSP with TCP connection for media data delivery may not work well in this case, due to packet reordering caused by tranmission of data of a flow on multiple networks without the support of additional infrastructure like a proxy that aggregates user level traffic and orders the data before sending it to the destination. Moreover, since the heterogeneous networks may be provisioned by different ISPs, the aggregation of user level flows might not be a straight forward task. However, since the characteristics of networks like available bit rate and delay vary, this category of applications can still benefit from improved performance performance if an application could be dynamically switched to the best network.
- Applications (for instance RTSP that uses UDP for data connection) that can be streamed over heterogeneous networks allow better utilization of the available networks. In this case, the data belonging to the applications can be suitably striped over multiple networks.

3.2 Network Support

For applications that require the preservation of TCP like transport semantics for data transfer, the network support for a proxy that assembles the data from multihomed devices on a per flow basis is required. Such a proxy would function to preserve the application flow semantics, like packet ordering for TCP whenever used for transport layer data transfer. However, this might entail collaboration between multiple (W)ISPs that provision network access. Applications that do not have these requirements, can be suitably striped over heterogeneous access networks.

3.3 System support

The multihomed application servers and sender devices need appropriate functionality to recognize the needs and nature of the applications and accordingly strip the data onto different networks. Such a middleware functionality in demonstrated in Figure 2. The application interface in the figure recognizes all the applications currently running on the server or the device. The flow classification module then parses the flows generated by these applications. A flow is registered as the quadruple <transport_protocol, destIP, srcPort, dstPort> for the datapath of the multimedia applications, with the transport_protocol being TCP or UDP. With the network support as mentioned in the previous subsection provisioned, then all flows can be passed on to the rate allocation unit that allocates rate to the flow on all active interfaces according to the network statistics (available bit rate, delay, etc.) provided by the network statistics evaluation and storage unit. If the network support

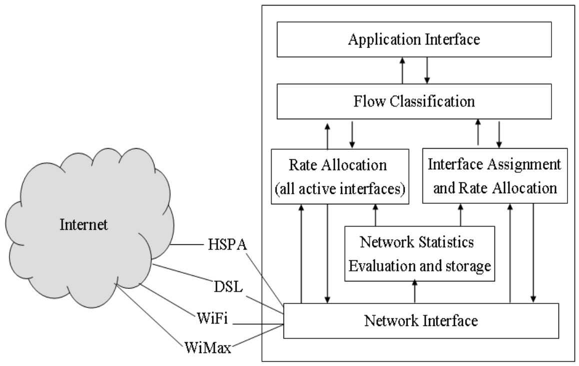


Figure 2: Middleware functionality.

as mentioned in the previous subsection is not provisioned, then only the flows that can be striped across multiple interfaces are passed to the rate allocation unit and the rest are directed to the interface assignment and rate allocation unit. The network interface component of the middleware provides statistics of the multiple networks to which the system has connectivity. These statistics can be measured by a) running a lightweight tool (e.g. [15]) to measure the available bit rate and RTT to destination, as done in our prior work [17, 24, 6], or b) using TCP parameters from existing TCP connection to estimate these characteristics as done in [21, 22] . A list of existing destinations of the streaming system is maintained (as also in [19]) by the middleware and network characteristics are periodically ascertained to these destinations.

4. FLOW ASSIGNMENT, RATE ALLOCATION, AND RATE CONTROL

Measures for optimal transport of multimedia application data have to account for a) application utilities which determine how much applications benefit from resources allocated to them b) the variation of underlying network characteristics with time and c) the constrains imposed by the transport mechanisms of the application data over the underlying networks. We address these issues below for the types of applications discussed in section 3.1.

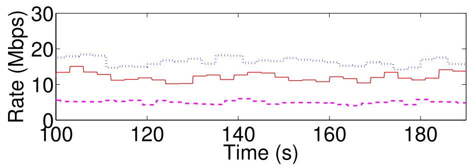
- For the flows that are constrained to traverse a single access technology, a suitable interface for communication has to be chosen. The conventional best-interface selection techniques, for example the highest capacity or lowest bit rate, may fail as these 1) represent the instantaneous characteristics of networks, which vary especially for wireless access networks 2) do not account for the utility functions of the flows, and 3) do not consider the existing traffic on the access networks.

  The optimal interface selection in this scenario warrants the stochastic characterization of the variation of network characteristics such as available bit rate and delay, and the maximization of the "reward" gained by the flow if the pertinent interface is selected for communication. The access networks can be characterized by modeling the network statistics such as available bit rate and round trip time in Figure 2 and as explained in Section 3.3. In [17] we present a modeling approach based on a Markov decision process (MDP) based control scheme that uses control Markov chains to characterize the variation in the bit rate and delays on heterogeneous access networks. The MDP is run by the device middleware and selects an access network that maximizes a discounted reward function which accounts for the utility function of the flows. The MDP based flow assignment policy is updated periodically and is dynamically consulted by the flows throughout their lifetimes to select the optimal networks. Our evaluations of high bit rate elastic video-like flows with TCP-style additive increase multiplicative decrease (AIMD) rate control using real-world traces collected on IEEE 802.11g, IEEE802.11b, and Ethernet networks, show that the MDP based flow assignment framework results in significantly better QoS for the flows in terms of lower packet delays and packet loss rates, and significantly increased tolerance for stringent playout deadlines. We refer the reader to [17] for a detailed description of the results.

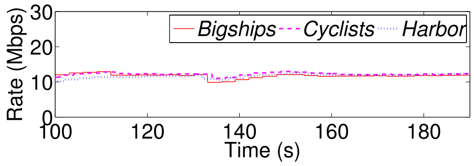
- The applications which do not mandate data transport over a single access network, allow for better striping of data over access networks in multi-homed systems, thus resulting in better QoS for the multimedia applications and better utilization of access networks. In [24], we present and evaluate distributed approaches for media streaming over heterogeneous access networks.

  1. When the streaming systems have access to DR characteristics of the video streams, the rate allocation problem can be solved via a convex optimization framework to minimize the sum of expected distortions of all participating streams. A distributed approximation (see [24]) to the optimization enables autonomous rate allocation at each streaming system and suggests the optimal rate that the flow can be allocated on each access interface. The distributed evaluation is done periodically for each flow, and the rate allocation is revised.
  2. When the media information is not accessible by the devices, H^∞ optimal rate allocation can be used for optimal bandwidth utilization on the access networks by guaranteeing a lower bound on a cost function that models the deviation of rate allocated to a stream from the rate available on a network. We have outlined the details of this policy in [6, 24].
  3. Simple heuristic-based rate allocation schemes that assign rates to streams on different networks in proportion to the available bit rate on the networks (referred to RP, rate proportional) or greedily according to the maximum available bit rate (called greedy), with each stream exercising a TCP congestion control [9] style additive increase multiplicative decrease (AIMD) rate control can also be utilized.

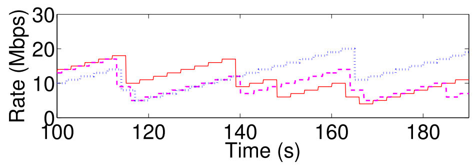
Our evaluation of the streaming of high-definition (HD) video sequences over access networks with real-world traces observed on Ethernet, IEEE 802.11g, and IEEE



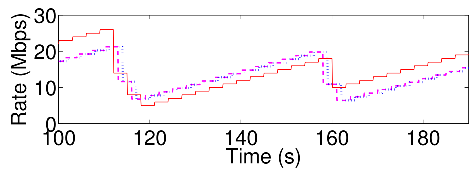
(a) Media-Aware



(b) H^{∞}-Optimal



(c) Greedy AIMD



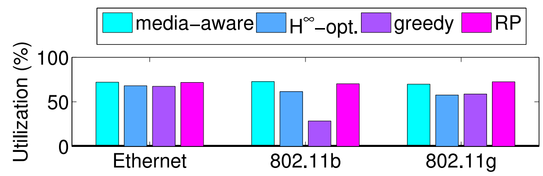
(d) Rate Proportional AIMD

Figure 3: Trace of rate allocated for HD video streams (Bigships, Cyclists, and Harbor : in order of increasing PSNRs for the same encoding rate), aggregated over three interfaces (802.11b, 802.11b, and Ethernet). Background traffic load is 20% and the playout deadline is 300 ms.

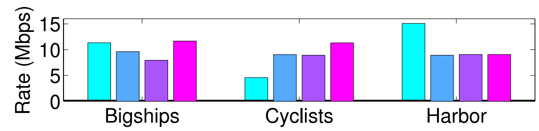
802.11b networks [24], shows that media-aware convex optimization and H^{∞} optimal control achieve significantly lower packet delays and loss rates thus resulting in better PSNRs (see Figure 4(c)) than heuristic based rate allocation for the video flows. Furthermore, the media aware scheme tends to assign higher rates for more demanding video sequences (see Figure 3 and Figure 4(b)), thus resulting in balanced video quality amongst video streams (see Figure 4(c)) and fairly even resource allocation among available access networks (see Figure 4(a)).

5. MIDDLEWARE POLICIES

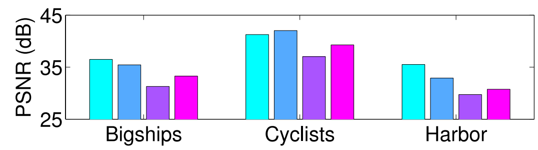
In this section we highlight the policies adopted by the middleware of Figure 2 in facilitating media streaming over heterogeneous networks. In Algorithm 1 we summarize the middleware operation in a streaming system. The application interface in Figure 2 parses the data flow(s) from the applications and determines whether the flow can be striped across multiple interfaces. If striping across heterogeneous interfaces is possible and if the DR characteristics of the application are known, then the media aware rate allocation of Section 4 is performed on all active interfaces by the rate allocation unit of Figure 2. If the DR characteristics are not known, then the H^{∞} optimal interface assignment and



(a)



(b)



(c)

Figure 4: Comparison of allocation results from different schemes for three HD video streams (Bigships, Cyclists, and Harbor), with background traffic load of 30%, and playout deadline chosen at 300 ms. (a) Aggregated network utilization over each interface; (b) Allocated video rate for each stream; (c) Corresponding received video quality in PSNR.

rate control is performed by the rate allocation unit. If the flow cannot be striped across multiple interfaces, then MDP based interface assignment and rate allocation of Section 4 is performed by the interface assignment and rate allocation unit of Figure 2.

```
parse_data_flow_from_app();
if flow can be stripped then
    if access to DR characteristics of the flow then
        media_aware_rate_allocation();
    else
        H∞_based_rate_allocation();
    end
else
    MDP_based_rate_allocation();
end
```

Algorithm 1: Middleware operation

We highlight the functionality of the network statistics evaluation and storage unit of Figure 2 in Algorithm 2. The unit monitors and updates the list of all destinations to which the system of streaming multimedia content. It measures the interface characteristics (available bit rate, RTT) to all these destinations on all interfaces, as discussed in Section 3.3. These characteristics are then stored and their Markov modeling is performed for by the middleware unit. This modeling of the interface characteristics is needed for the MDP based rate allocation.

With these policies and the architecture highlighted in Figure 2, the streaming system can accomplish the rate allcation measures described in the previous section for different multimedia applications.

dlist update_destination_list();
measure_interface_characteristics(*dlist*);
store_characteristics();
markov_model_characteristics();
Algorithm 2: Network characteristics measurement

6. CONCLUSIONS

The heterogeneity in characteristics of access networks and transport mechanisms of multimedia applications in a multi-homed environment requires specialized handling of applications for efficient utilization of access networks and enhanced quality of experience for the users. We propose the architecture and policies for a middleware in a streaming system that monitors the variation in characteristics of access networks and uses the information about the nature of media applications when available, and accordingly tailors the rate allocation, rate control and flow assignment policies for applications. Under each scenario, the middleware targets optimal allocation of access networks resources so as to maximize the utilities of the multimedia flows and utilization of the access networks.

7. ACKNOWLEDGMENTS

We would like to thank Varun Sharma for helping with statistical analysis of the network characteristics variations in Section 1.

8. REFERENCES

[1] IEEE 802.21. *http://www.ieee802.org/21/.*

[2] A comparative analysis of mobile WiMAX deployment alternatives in the access network. *White Paper, The WiMAX Forum*, May 2006.

[3] H. Akhavan, V. Badrinath, T. Geitner, H. Lennertz, Y. Sha, T. Utano, and B. West. Next generation mobile netoworks - Beyond HSPA and EVDO. *White Paper, Next Generation Mobile Networks Ltd.*, December 2006.

[4] T. Alpcan and T. Başar. Global stability analysis of an end-to-end congestion control scheme for general topology networks with delay. In *Proc. of the 42nd IEEE Conference on Decision and Control*, pages 1092 – 1097, Maui, HI, December 2003.

[5] T. Alpcan and T. Başar. A utility-based congestion control scheme for Internet-style networks with delay. *IEEE Transactions on Networking*, 13(6):1261–1274, December 2005.

[6] T. Alpcan, J. P. Singh, and T. Basar. A robust flow control framework for heterogenous network access. In *5th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Limassol, Cyprus, 2007.

[7] 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. In *IEEE Communications Magazine, issue on Advances of Service Platform Technologies*, Aug. 2006.

[8] S. Floyd and K. Fall. Promoting the use of end-to-end congestion control in the internet. *IEEE/ACM Trans. on Networking*, 7:458–472, Aug. 1999.

[9] V. Jacobson. Congestion avoidence and control. *Proc. SIGCOMM'88*, 18(4), Aug. 1988.

[10] D. Jurca and P. Frossard. Media-specific rate allocation in heterogeneous wireless networks. *Proc. Packet Video Workshop, (PV-06), Hangzhou, China*, April 2006.

[11] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of Operations Research Society*, 49(3):237–252, 1998.

[12] R. J. La and V. Anantharam. Utility-based rate control in the internet for elastic traffic. *IEEE Trans. on Networking*, 10(2):272–285, Oct. 2002.

[13] M. Allman and V. Paxson and W. R. Stevens. *TCP Congestion Control, RFC 2581*, Apr. 1999.

[14] M. Handley and S. Floyd and J. Pahdye and J. Widmer. *TCP Friendly Rate Control (TFRC): Protocol Specification, RFC 3448*, Jan. 2003.

[15] J. Navratil and R. L. Cottrell. Abing. *http://www-iepm.slac.stanford.edu/tools/abing/.*

[16] S. Shakkottai, E. Altman, and A. Kumar. The case for non-cooperative multihoming of users to acess points in IEEE 802.11 WLANs. In *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[17] J. P. Singh, T. Alpcan, P. Agrawal, and V. Sharma. An optimal flow assignment framework for heterogeneous network access. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Helsinki, Finland, 2007.

[18] A. Szwabe, A. Schorr, F. J. Hauck, and A. J. Kassler. Dynamic multimedia stream adaptation and rate control for heterogeneous networks. *Proc. Packet Video Workshop, (PV-06), Hangzhou, China*, April 2006.

[19] N. Thompson, G. He, and H. Luo. Flow scheduling for end-host multihoming. In *Proc. IEEE INFOCOM*, 2006.

[20] P. Vidales, J. Baliosion, J. Serrat, G. Mapp, F. Stejano, and A. Hopper. Autonomic sytem for mobility support in 4G networks. In *IEEE Journal on Selcted Areas in Communications'05*, volume 23, Dec. 2005.

[21] R. Wang, K. Yamada, M. Y. Sanadidi, and M. Gerla. TCP with sender-side intelligence to handle dynamic, large, leaky pipes. *IEEE Journal on Selected Areas in Communications*, 23(2):235–248, Feb 2005.

[22] K. Xu, Y. Tian, and N. Ansari. TCP-Jersey for wireless IP communications. *IEEE Journal on Selected Areas in Communications*, 22(4):747–756, May 2004.

[23] H. Yaiche, R. Mazumdar, and C. Rosenburg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. In *IEEE/ACM Transaction on Networking*, volume 8, pages 667–678, Oct. 2000.

[24] X. Zhu, P. Agarwal, J. P. Singh, T. Alpcan, and B. Girod. Rate allocation for multi-user video streaming over heterogeneous access networks. In *ACM Multimedia*, September 2007.

[25] X. Zhu, J. P. Singh, and B. Girod. Joint routing and rate allocation for multiple video streams in ad hoc wireless networks. *Proc. Packet Video Workshop, (PV-06), Hangzhou, China*, April 2006.