An Assessment of Current Layered Multicast Techniques for Real-time Video

Osman Ghazali Suhaidi Hassan
Department of Computer Science, Faculty of Information Technology
Universiti Utara Malaysia, 06010 UUM Sintok, Malaysia
e-mail: {osman|suhaidi}@uum.edu.my

Abstract- *The deployment of real-time video applications in the Internet has increased tremendously in recent years, and at the same time Layered Multicast Protocol (LMP) has achieved similar attention in the research community. Since LMP is regarded as one of the solution for real-time video applications, this paper attempts to assess the techniques employed in the current LMPs. We present a taxonomy of layered multicast techniques as well as the analysis of each facet of the taxonomy. Finally some research issues are highlighted.*

Keywords: Multicast, Layered Multicast, Transport Protocol

1. Introduction

The increasing deployment of real-time video applications in the Internet in recent years has increased the necessity for a better protocol that supports real-time video transmission over the Internet. This application has certain characteristics, which make it somewhat different from other Internet data flows. The characteristics are loss-tolerant, delay-sensitive, jitter-sensitive, and large user distribution [1]. Layered Multicast Protocol (LMP), which allows users with different network capacities to achieve different receiving rates, is regarded as one of the solution for video transmission over the best-effort Internet services.

LMP operates on the Internet Protocol (IP) multicast, where it encodes video signal into multi-rate data signals using certain encoding technique. These data signals are then transmitted into layers or channels of a multicast session, where each layer carries certain level of video quality. At the other ends, receivers adapt their reception rates by joining and dropping layers. The more layers or the higher layers the receivers join the better video quality they perceive. Each receiver may achieve different layer subscription level depending on their network capacities. This enables receivers with different network capacities to experience different video qualities.

The performance of LMPs is dependent on the techniques employed in the protocols. Given the active research undertaking and LMP is yet to reach its maturity, there are quite a number of different techniques employed in the current LMPs. The goal of this article is to assess the techniques employed in the recent LMPs based on certain classifications. This article is not intended to be exhaustive but to focus on the techniques employed in the current LMPs. The remainder of this article is organized as follows. The next section presents the overview of LMP, Section 3 presents the taxonomy of LMP techniques, Section 4 is a discussion and analysis on the LMP techniques, Section 5 highlights some research issues, and finally Section 6 concludes this paper.

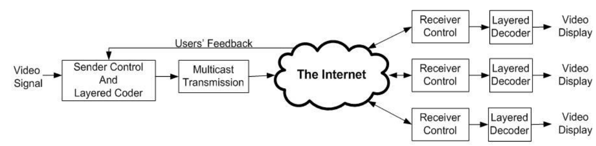

Figure 1: Layered multicast communication

2. Overview of Layered Multicast Protocol

Figure 1 shows layered multicast communication in the Internet. At the sender, video signal is encoded into multi-rate data signals using a layering scheme. Four layering schemes are currently used in LMPs, namely cumulative [2], non-cumulative [3], simulcast [4] and dynamic layering [5, 6]. Section 3.1 discusses the detail of each layering scheme. The encoded data signals are then sent to multicast layers or channels for multicast transmission.

At the other ends, receivers receive video transmission and perform video decoding. Video decoding is performed upon receiving data packets from the sender, and the decoding process is dependent on the encoding technique used at the sender.

Another important task performed in layered multicast communication is rate adaptation. The goals of rate adaptation are to control congestion, and to ensure each receiver achieves higher and smooth receiving rates. Rate adaptation can be performed either at the sender or at the receivers. Rate adaptation at the sender is performed by adjusting layers' rate according to aggregate receivers' bandwidth. To perform the sender's rate adaptation, information about receivers' network status is required. This requires feedback from all receivers. On the other hand, receivers' rate adaptation does not require feedback, but is performed by subscribing and/or dropping layers at receivers. Typically subscription level must not exceed the estimated target reception rate.

Most LMPs estimate target reception rate using TCP equation model. In order to estimate the rate using TCP equation model, two important parameters, namely loss rate and round trip time (RTT), have to be estimated. However, estimating these parameters is problematic in layered multicast communication.

3. A Taxonomy of Layered Multicast Techniques

We classify layered multicast techniques according to the influences on the design and performance of LMPs, based upon the following aspects:

i. ***Layering scheme*** – i.e. how the sender encodes video signal into layers.
ii. **Rate *adaptation location*** – i.e. the location where rate adaptation is performed.
iii. ***Target reception rate estimation*** – i.e. how the protocols estimate their target reception rate.
iv. **RTT *estimation*** – i.e. the techniques used to estimate round trip time.
v. ***Loss rate estimation*** – i.e. the techniques used to estimate loss rate.

3.1 Layering scheme

Layering scheme is the method used to encode video signals into multi-rate data flows for transmission over the IP multicast layers. Four layering schemes have been proposed, namely cumulative, non-cumulative, simulcast and dynamic layering.

Cumulative layering scheme encodes video signal into a few non-redundant layers, which composed of a base layer and enhancement layers. It encodes video signal in such a way that the upper layers are dependent on the lower layers. The base layer is the most important layer that provides the basic video quality, whereas the upper layers provide quality enhancement to the lower layers. A receiver must join the layers in cumulative order that it must first join the base layer and the consequent layers in incremental order, e.g. for a session with 4 layers the receiver must subscribe the layers in the following order: 1, 1-2, 1-2-3, and 1-2-3-4.

On the other hand, non-cumulative layering scheme encodes video signals similar to cumulative layering scheme but each layer is independent of each other and is equally important. Non-cumulative layering scheme allows a receiver to join the higher layers without the need to first subscribe the lower layers, hence mitigate the cumulative layer subscription requirement as in the cumulative layering scheme. This enables more flexible layer subscription and finer granularity.

Simulcast layering scheme encodes video signals into a few redundant layers, where the upper layer contains the video signal of the subsequent lower layer together with additional video quality. This means the upper layers contain a better video quality than the lower layers. The layers in this scheme are similar to a few single-rate multicast channels with different quality of video signals. A receiver can subscribe to any layer that fits to its target reception rate. However, the receiver has to subscribe to only one layer at one particular time. If more bandwidth is available, the receiver may achieve better video quality by dropping the currently joined layer and subscribing the subsequent higher layer.

Dynamic layering schemes [5, 6] encode data into a few layers in such a way that at the beginning of each layer the sending rate is set at the maximum rate and then gradually decreases until it reaches zero. Then, the layer will be in quiescent period for a pre-defined time period before starting sending another cycle of video signal. A receiver can join the layers at any time-point depending on its available bandwidth. The receiver can maintain certain reception rate by keep subscribing to new layers when the subscribed layers decrease their sending rates. This scheme solves the Internet Group Multicast Protocol (IGMP) latency problem, i.e. the problem of too long time taken by the IGMP to process a receiver's layer drop request. Two dynamic layering schemes have been proposed – Fair Layered Increase/Decrease with Dynamic Layering (FLID-DL) scheme [5] and Wave-like Dynamic Layering scheme [6].

3.2 Rate Adaptation Location

Typical LMPs perform rate adaptation at receivers. Some LMPs, however, allow sender-based adaptation in addition to receiver-based adaptation. Hence, LMPs can be classified into two location-based rate adaptation categories, namely receiver-based LMP [2, 3, 5-10] and hybrid LMP [11-14]. The receiver-based LMPs perform rate adaptation merely at receivers, while the hybrid LMPs perform rate adaptation at the sender and receivers.

A receiver-based LMP allows adaptive rate change at receivers. Adaptive rate change at receiver enables receivers with different network bandwidth to perform rate adaptation without the need to send feedback to the sender, hence improves the protocols scalability. Adaptive rate change also enables receivers with different network paths to achieve different receiving rates, which is desirable for the heterogeneous Internet environment. The receiver-based LMPs, however, pose problems of balancing between achieving fine granularity and protocols responsiveness. Fine granularity could be achieved by opting for many low-rate layers. However, too many layers in a multicast session result in slow response and convergence. Moreover, too many layers in a multicast session incurs more network management overhead.

Hybrid rate adaptation solves the problems in receiver-based rate adaptation by performing rate adaptation at both the sender and receivers. Hybrid LMPs achieve fine granularity by adjusting layers' rate at the sender in addition to receivers' rate adaptation. Rate adaptation at the sender is performed to satisfy target reception rates for the majority of receivers, while receivers' rate adaptations are performed to satisfy the individual receiver. Having performed dual location-based rate adaptation, fine-grained reception rate could be achieved. Moreover, performing rate adaptation at both the sender and receivers enables LMPs to achieve fast optimum throughput with minimal number of layers. Liu et al. [12] claim that using hybrid rate adaptation Hybrid Adaptation Layered Multicast (HALM) could achieve optimal throughput with only 3 layers. However, hybrid LMP poses problems of scalability and inter-receiver fairness.

In order to perform rate adaptation, the sender of a hybrid LMP requires information of the network status from all receivers. Hence, it is very important to have a feedback suppression technique that enables two-way communication (between the sender and receivers) in a scalable manner.

3.3 Target Reception Rate Estimation

LMPs perform congestion control by regulating their reception rates based on estimated target reception rate. Typically, receivers of a layered multicast session adjust their reception rate at the level that the receiving throughput must not exceed the target reception rate. Handley et al. [15] suggest that in order for a protocol to be friendly to other protocols, the throughput of its receivers should not exceed the throughput of TCP flows in the same network condition. Hence, in order to be friendly to other competing flows, almost all LMPs employed TCP equation model in estimating their target reception rate. A number of TCP equation models that imitate steady-state behaviour of certain TCPs have been proposed in [10, 16-18]. The deployment of certain TCP equation model determines the behaviour of LMPs, particularly the aggressiveness of the LMPs toward other competing flows in same network path. In line with the fact that TCP Reno is the most popular TCP flavour in the Internet[1], TCP Reno equation model [17] is the most used model in LMPs [1, 5, 6, 11-13].

Recently, a TCP Vegas equation model has been developed by Samois and Vernon [18]. The model uses queuing delay and parameter thresholds of unacknowledged delay. This enables detection of network congestion build-up before packet loss occurs. The main problem in the TCP Vegas equation model, however, is the difficulty to choose parameter thresholds that enable a protocol to fairly compete with other competing flows, especially with the more aggressive flows such as TCP Reno. This impedes its deployment in LMPs. To solve this problem, Mahanti et al. [10] propose a technique that dynamically measures parameter thresholds of the TCP Vegas equation model. Their experiment on Vegas Multi-rate Congestion Control (VMCC) protocol that employed TCP Vegas equation model exhibits TCP Reno-like characteristics, provides less oscillatory throughput, and induces no packet loss in lightly loaded bottleneck links.

Legout and Biersack [8] use packet pair queuing delay in the Packet-pair receive-driven Layered Multicast (PLM) protocol. In order to estimate available network bandwidth, PLM sends a pair of packets back to back into a multicast session. Upon receiving the packet pair, receivers measure the receiving time gap of the packets. The decision to join or drop layers is made based on the estimated available bandwidth. PLM assumes fair queuing management is implemented in the network, which is very difficult to be implemented in the current Internet. With fair queuing management, their simulation experiment demonstrates that PLM behave friendly toward other protocols. However, Puangpronpitag et al. [20] in their simulation experiment demonstrates that PLM behaves unfriendly towards other competing flows when running in a network with no fair queuing management.

Puangpronpitag et al. [9] use a combination of packet pair technique and TCP equation model [17] in Explicit Rate Adjustment Protocol (ERA) to estimate target reception rates, where the lower rate estimation of the two technique is used in join and drop decisions. Adopting both techniques, they claim ERA is able to respond quickly to congestion and friendly to the competing flows. However, by estimating two target reception rates and using the lesser rate for join and drop decisions, the protocol may behave conservatively.

3.4 Round Trip Time Estimation

There is no accepted Round Trip Time (RTT) estimation standard in LMPs. Full RTT estimation is difficult to estimate in a layered multicast session due to the implosion problem at the sender, which is the result of too many feedbacks from a large number of receivers. Consequently, different measurement techniques have been proposed.

The simplest way to assign RTT value is to use a fixed value, e.g. 1 second. This technique has been used in [3, 5, 8, 9, 21]. Assigning a fixed value to RTT has the advantage that all receivers of a multicast session use the same RTT for target reception rate estimation. This enables receivers behind the same bottleneck to estimate the same rate, and consequently achieve the same layer subscription level, which is desirable for layered multicast communication. However, a protocol that employs fixed RTT value ignores the link delay which is one of the main indicators of congestion build-up in the network.

Another RTT estimation technique is to estimate one-way transmission time (OTT) [13]. This technique requires the use of the source timestamp on each packet sent to the multicast channel. At the other ends, receivers will compute the time difference between the sending time (the source's time) and the receiving time (the receivers' time). The advantage of this technique is its simplicity, while the disadvantage is it may suffer from clock drift and may require clock synchronisation which is very difficult to do in multicast communication. Another disadvantage is it does not accurately measure RTT due to the asymmetric network path.

To solve clock synchronisation complexity, Mahanti et al. [10] use a combination of a fixed RTT value and queuing delay. Queuing delay is acquired by measuring the time difference in the observed OTT. Having been able to measure and use queuing delay for RTT, receivers would be able to respond to congestion build-up, and receivers behind the same bottleneck may achieve the same layer subscription level. This technique may solve time skew problem, however asymmetric network path problem remain unresolved.

Some researchers [11, 12, 14, 22] measure full RTT by using feedback suppression techniques. These techniques use probabilistic feedback scheme to ensure that there is no more than one receiver that sends feedback to the sender at the same time, therefore avoiding too many feedbacks at the sender. However, in the case where too many receivers join a session at the same time, each receiver has to wait for a considerable amount of time before being able to measure a new RTT. The longer a receiver has to wait the lesser the RTT accuracy.

[1] Please refer to Fall and Floyd [19] for discussion on TCP flavours.

Protocol	Adaptation Location	Layering Scheme	Target Reception Rate Estimation	RTT Estimation	Loss Rate Estimation
RLM [2]	Receiver	Cumulative	-	-	-
RLC [7]	Receiver	Cumulative	Periodic Burst	-	-
PLM [8]	Receiver	Cumulative	Packet Pair	Fixed RTT value	-
FLID-DL [5]	Receiver	Dynamic	TCP Reno equation model	Fixed RTT value	Packet loss rate
CIFL [22]	Receiver	Cumulative	TCP Reno equation model	RTT using feedback suppression	Packet loss rate
MLDA [13]	Hybrid	Cumulative	TCP Reno equation model	End-to-end one way delay estimation	Packet loss rate
WEBRC [6]	Receiver	Dynamic	TCP Reno equation model	The duration of time from join request until the arrival of the first packet.	Loss event rate
ERA [9, 21]	Receiver	Cumulative	TCP Reno equation model , Packet Pair	Fixed RTT value	Packet loss rate
HALM [11, 12]	Hybrid	Cumulative, Dynamic	TCP Reno equation model	RTT using feedback suppression	Loss event rate
SMCC [14]	Hybrid	Cumulative, Dynamic	TCP Reno equation model	RTT using feedback suppression	Loss event rate
VMCC [10]	Receiver	Cumulative	TCP Vegas equation model	Fixed rate and average queuing delay	Loss event rate
FGLM [3]	Receiver	Non-cumulative	TCP Reno equation model	Fixed RTT value	Packet loss rate

Table 1: A Taxonomy of Layered Multicast Techniques

Other researchers [6, 7] measure RTT as the time taken from the time of join request until the arrival of the first packet, where it measure the round trip time between the sender and first router that route the multicast flows. The advantage of this technique is its simplicity, while the disadvantage is the ignorance of the round trip time between the source and the first router.

Basu and Golestani [23] employ a hierarchical technique for RTT estimation. Using this technique, a group of receivers behind the same bottleneck (cluster) will select one receiver as a parent. The sender only needs to communicate with the parents. All receivers send feedbacks to the parents, and the parents aggregate the children feedbacks and send the aggregate feedbacks to the sender. This reduces the amount of feedbacks sent to the sender. RTT is calculated by combining parent-source RTT and child-parent RTT. The advantage of this technique is the accurate full RTT estimation without the need to do clock synchronisation, while the disadvantage is its complexity and processing overhead.

3.5 Loss Rate Estimation

Two loss rate estimation techniques are currently used in the current LMPs, namely packet loss rate and loss event rate, with packet loss rate is the mostly used. However, there is little discussion on the issues of loss rate estimation in LMP literature. A few researchers claim that they employs loss event rate estimation technique in their LMPs, i.e. SMCC, HALM, VMCC and WEBRC. However, the detail mechanism of the implementation is not been explained

SMCC estimate loss event rate per-layer basis which is similar to the techniques in [24]. This technique is suitable for single layered multicast but is not really suitable for typical LMP. In typical LMP, a session data packets are distributed into layers where each layer can be seen as a single layered multicast as in [24]. In contrast to single layered multicast, LMP performs aggregate loss rate estimation across all layers. It is more complicated than packet loss rate estimation and loss event rate estimation in a single layered multicast. Liu et al. [11, 12] and Luby et al. [6] also claim that they use loss event rate estimation technique to measure loss rate in HALM and WEBRC. However, little explanation is given in regards to the detail mechanism of the implementation, particularly synchronisation of loss estimation across layers. Mahanti et al. [10] estimate loss event rate using closely-spaced packet losses. Similarly, they also provide no explanation on how packet losses are spaced and how packet losses across layers are synchronised.

4.0 Discussion

We have classified the techniques of layered multicast protocols in the previous section. Table 1 provides a summary of these techniques. This section presents discussion and analysis of the techniques.

The easiest way to avoid feedback implosion at the sender is not to send feedback. That is the reason why the receiver-based LMP is more popular than the sender-based LMP. The elimination of feedback in the receiver-based LMPs has greatly improved the scalability of layered multicast protocols. However, without feedback the receivers have to adapt their receiving rates to static pre-define layer sizing. This may result to coarse granularity. Some protocols [11-14] use hybrid adaptation scheme to mitigate the problems in receiver-based LMPs. Hybrid adaptation scheme enables rate adaptation to be done at both the sender and receivers. This is possible with the use of lightweight feedback suppression techniques which allow feedback to be sent in a scalable manner. However, in the case of too many receivers in a layered multicast session, feedback suppression techniques may not work very well. When there are too many receivers, the time each user has to wait before being able to send feedback would be too long. As a result, the sender could not perform layer adaptation correctly.

Simulcast layering enables finer-granularity with less coding complexity but at the expense of low bandwidth dilution [3]. Similarly, dynamic layering enables finer granularity and solved IGMP latency problem, unfortunately its implementation is not suitable for real-time video stream. Non-cumulative layering enables receivers achieve finer-granularity but at the expense of lower bandwidth utilisation. Cumulative layering, despite its coarse granularity, is the most popular technique. This is due to its simplicity and better bandwidth utilisation. Furthermore, cumulative layering technique requires less joining and dropping coordination.

IGMP processing time may slowdown LMP responsiveness. This problem can be avoided by carefully estimating available bandwidth before joining new layers that would minimise failed join attempt. However, this has no guarantee that the newly joined layer will not cause congestion to the network, therefore they may still suffer from IGMP latency problems if the newly subscribe layer is not successful. Dynamic layering [5, 6] mitigate the IGMP latency problem. However the high joining frequency in [5] has made this technique infeasible to be implemented in the Internet. Although new dynamic layering technique [6] demonstrates better performance than the technique in [5] with lesser joining frequency, its exponentially decreasing transmission rate is not suitable for video transmission. Furthermore its complexity incurs more processing overhead.

TCP equation model is an important component in controlling network congestion. Since TCP Reno is the most deployed protocol in the Internet, TCP Reno equation model is the most popular TCP throughput estimation technique for LMP. However, a protocol that employs TCP Reno equation model is more aggressive than a protocol that employs TCP Vegas equation model, where the former is more oscillatory and induces more packet loss. This behaviour is undesirable for video applications. The protocol that employs TCP Vegas equation model, though less oscillatory and induces less packets loss have the difficulties to measure certain important parameters. Mahanti et al. [10] solve this problem by proposing a dynamic parameter measurement.

RTT value is needed to estimate receivers' target reception rate. However, measuring RTT in layered multicast communication is very complicated. It is difficult to measure RTT in a multicast session due to the large number of receivers, where feedback from receivers may cause feedback implosion at the sender. Furthermore, in a layered multicast session, it is desirable for all receivers behind the same bottleneck to have the same RTT value, thus enable receivers behind the same bottleneck to achieve the same subscription level. Moreover, reverse message transmission (message from receivers to the sender) may never occur in receiver-based layered multicast. This raises the question of whether full RTT estimation is really necessary. The complexities in measuring RTT have resulted in the proposal of a number of different RTT estimation techniques. Such techniques are fixed RTT, OTT, fixed RTT with queuing delay, full RTT with feedback suppression mechanism, and hierarchical RTT estimation. Fixed RTT enables receivers behind the same bottleneck to achieve the same throughput estimation but ignores the network queuing delay. OTT suffers from clock drift and requires clock synchronisation. Fixed RTT with queuing delay may achieve same throughput estimation and is responsive to the network condition, but not round and not very responsive to the network condition. Full RTT estimation may achieve accurate RTT, but without appropriate feedback suppression, it will cause congestion problem at the sender. Hierarchical RTT estimation is very likely to achieve accurate RTT estimation and the same target reception rate estimation for receivers behind the same bottleneck but at the expense of more processing overhead.

Loss rate is one of the most important parameter in TCP equation model [25]. Its accuracy will determine the accuracy of TCP throughput estimation. Most of the current LMPs employ packet lost rate technique to estimate loss rate. However this technique has been criticised by [26] that it does not accurately model TCP behaviour. Alternatively, they suggest that loss event rate technique models TCP behaviour more accurately than packet loss rate technique. In their study they found the protocols that employ loss event rate estimation technique, perform better than the protocols that employ packet loss rate. However the study is conducted in the unicast environment. To our knowledge there is no such study has been conducted for layered multicast environment. In bold comparison, it is easier to estimate packet loss rate than to estimate loss event rate. Loss event rate is difficult to estimate due to many layers involved in a layered multicast session. Whereby, each receiver may subscribe to different layers that make it not possible to assign session sequence number to the packets, instead each packet is assigned layer sequence number. These factors make the monitoring of loss event rate in LMP very complicated.

5.0 Research Directions

In the previous section, we discuss issues in the current LMPs. Some research directions have been identified. Two categories of location-based rate adaptation are commonly used in LMP, but which location-based rate adaptation that is more appropriate for video applications remains to be investigated. Layering schemes seem to solve certain problems but create other problems. It is desirable to have layering with fine granularity, high bandwidth utilisation, easy to coordinate, and simple to implement. However, such a layering scheme is yet to come. It is also desirable to know which TCP equation model or bandwidth estimation technique is more appropriate to be used in LMP. However, despite the aggressiveness exhibited by the protocols that employed TCP Reno equation model, most of the LMPs employ TCP Reno equation model. Techniques that respond to congestion build-up such as in [8, 9], could be alternatives for non-TCP flow like the data flow in LMP. LMP requires special RTT estimation techniques as it needs to ascertain receivers behind the same bottleneck receive the same layer subscription level. With the fact that full RTT is difficult to measure, RTT estimation techniques remain to be explored. Loss rate issue has not been discussed in detail in the current LMP literature, particularly the implementation of loss event rate. The fact that loss rate is a very important parameter in TCP equation model and packet loss rate does not accurately models TCP behaviour, the implementation of loss event rate estimation in LMP requires further exploration and explanation.

6.0 Conclusion

In this article we assess the techniques in the current LMPs. We present the taxonomy of the LMPs' techniques as well as the discussion and analysis of each facet of the taxonomy. There are some issues remain to be investigated that include location-based rate adaptation, layering scheme, target reception rate estimation, RTT estimation, and loss event rate estimation.

References

[1] S. Puangpronpitag, "Design and Performance Evaluation of Multicast Congestion Control for the Internet," *Computer Science Department*: University of Leeds, UK, 2003.

[2] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast," in the Proceedings of ACM SIGCOMM 1996, Stanford, 1996.

[3] J. Byers, M. Luby, and M. Mitzenmacher, "Fine-Grained Layered Multicast," in the Proceedings of IEEE INFOCOM 2001, Anchorage, 2001.

[4] J. Liu, B. Li, and Y.-Q. Zhang, "Adaptive Video Multicast over the Internet," *IEEE Multimedia*, vol. 10, pp. 22-31, 2003.

[5] J. Byers, M. Frumin, G. Horn, M. Luby, M. Mitzenmacher, A. Roetter, and W. Shaver, "FLID-DL: Congestion Control for Layered Multicast," in the Proceedings of ACM NGC 2000, Palo Alto, 2000.

[6] M. Luby, V. K. Goyal, S. Skaria, and G. B. Horn, "Wave and Equation Based Rate Control Using Multicast Round Trip Time," in the Proceedings of ACM SIGCOMM 2002, Pittsburgh, 2002.

[7] L. Vicisano, L. Rizzo, and J. Crowcroft, "TCP-like Congestion Control for Layered Multicast Data Transfer," in the Proceedings of IEEE INFOCOM 1998, San Francisco, 1998.

[8] A. Legout and E. Biersack, "PLM: Fast Convergence for Cumulative Layered Multicast Transmission Schemes," in the Proceedings of ACM SIGMETRICS 2000, Santa Clara, 2000.

[9] S. Puangpronpitag, R. Boyle, and S. Hassan, "Explicit Rate Adjustment for Multi-rate Multicast Congestion Control Using TCP Throughput Equation and Packet-pair Probe," in the Proceedings of APCC 2003, Penang, Malaysia, 2003.

[10] A. Mahanti, D. L. Eager, and M. K. Vernon, "Improving Multirate Congestion Control Using TCP Vegas Throughput," *available from* http://www.cpsc.ucalgary.ca/~mahanti/papers/vmrc.ps, 2004.

[11] J. Liu, B. Li, and Y. Zhang, "A Hybrid Adaptation Protocol for TCP-friendly Layered Multicast and Its Optimal Rate Allocation," in the Proceedings of INFOCOM 2002, New York, 2002.

[12] J. Liu, B. Li, and Y. Zhang, "An End-to-End Adaptation Protocol for Layered Video Multicast Using Optimal Rate Allocation," *IEEE Transactions on Multimedia*, vol. 6, 2004.

[13] D. Sisalem and A. Wolisz, "MLDA: A TCP-friendly Congestion Control Scheme," in the Proceedings of IWQoS 2000, Pittsburgh, 2000.

[14] G.-I. Kwon and J. W. Byers, "Smooth Multirate Multicast Congestion Control," in the Proceedings of IEEE INFOCOM 2003, San Francisco, 2003.

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

[16] J. Mahdavi and S. Floyd, "TCP-friendly Unicast Rate-based Flow Control," *available from* http://www.psc.edu/networking/papers/tcp_friendly.html, 1997.

[17] J. D. Padhye, V. Firoiu, D. F. Towsley, and J. F. Kurose, "Modelling TCP throughput: A Simple Model and its Empirical Validation," in the Proceedings of ACM SIGCOMM 1998, Vancouver, 1998.

[18] B. Samois and M. Vernon, "Modeling the Throughput of TCP Vegas," in the Proceedings of ACM SIGMETRICS 2003, San Diego, 2003.

[19] K. Fall and S. Floyd, "Simulation-based Comparison of Tahoe, Reno and SACK TCP," *ACM CCR 1996*, vol. 26, pp. 5-22, 1996.

[20] S. Puangpronpitag, R. D. Boyle, and K. Djemame, "Performance Evaluation of Layered Multicast Congestion Control Protocols: FLID-DL vs. PLM," in the Proceedings of SPECTS 2003, Montreal, 2003.

[21] S. Puangpronpitag, R. D. Boyle, and K. Djemame, "Explicit Rate Adjustment: an Efficient Congestion Control Protocol for Layered Multicast," in the Proceedings of ICON 2003, Sydney, Australia, 2003.

[22] I. E. Khayat and G. Leduc, "A Stable and Flexible TCP-friendly Congestion Control Protocol for Layered Multicast Transmission," in the Proceedings of IDMS 2001, Lancaster, UK, 2001.

[23] A. Basu and S. J. Golestani, "Estimation of Receiver Round Trip Times in Multicast Communications," *available from* www.bell-labs.com/user/golestani/rtt.ps, 1999.

[24] J. Widmer and M. Handley, "TCP-Friendly Multicast Congestion Control (TFMCC): Protocol Specification," *Internet draft, draft-ietf-rmt-bb-tfmcc-02.txt*, 2003.

[25] M. Vojnovic and J.-Y. L. Boudec, "On the Long-Run Behavior of Equation-Based Rate Control," in the Proceedings of ACM SIGCOMM 2002, Pittsburgh, PA, 2002.

[26] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-Based Congestion Control for Unicast Applications," in the Proceedings of ACM SIGCOMM 2000, Stockholm, Sweden, 2000.