Addressing Heterogeneity and Scalability in Layered Multicast Congestion Control

Sergey Gorinsky K. K. Ramakrishnan Harrick Vin

Technical Report TR2000-31
Department of Computer Sciences, University of Texas at Austin
Taylor Hall 2.124, Austin, Texas 78712-1188, USA
gorinsky@cs.utexas.edu, kk@teraoptic.com, vin@cs.utexas.edu

November 24, 2000

Abstract— **In this paper, we design SIM, a protocol that integrates three distinct mechanisms – Selective participation, Intra-group transmission adjustment, and Menu adaptation – to solve the general multicast congestion control problem. We argue that only a solution that includes elements of each mechanism can scale and adapt to heterogeneity in network and receiver characteristics. In our protocol, these mechanisms operate at different time scales and distribute the responsibility of adaptation to different entities in the network. Per our knowledge, SIM is the first protocol for layered multicast that adjusts not only the subscription levels of the receivers but also the transmission rates of the layers. We show that SIM is efficient and stable in the presence of heterogeneous receivers and dynamic changes in the bottlenecks and session membership. SIM also outperforms RLM in terms of stability and efficiency.**

1 Introduction

Multicast is a network service for delivering the same data to multiple destinations. We believe that a multicast congestion control protocol should be (1) efficient (i.e., should efficiently utilize the bandwidths available in the heterogeneous network); (2) scalable (i.e., should support a large number of receivers); (3) fair (i.e., should allocate network bandwidth fairly among competing sessions); (4) responsive (i.e., should converge to a fair efficient state promptly); and (5) light-weight (i.e., should impose a low communication and computation overhead). Finally, the protocol should meet all of the above requirements while preserving privacy of the receivers. While individual objectives are achieved by some of the existing multicast congestion control protocols, none of these protocols provides the full set of the desired properties.

In this paper, we propose **SIM**, a protocol that integrates three distinct mechanisms – **S**elective participation, **I**ntra-group transmission adjustment, and **M**enu adaptation – to provide a general solution for the multicast congestion control problem for layered streams. These mechanisms operate at different time scales and place the responsibility of adaptation at different entities in the network, so as to support a scalable and timely control.

The individual mechanisms of **SIM** are not novel and appear in other layered multicast protocols. For example, RLM [13] relies on selective participation while SAMM [22] adjusts the transmission rates of the layers. However, SIM is – per our knowledge – the first protocol for layered multicast that adjusts both the group subscription levels and layer transmission rates. This feature allows SIM to use the available bandwidths efficiently in the presence of heterogeneous receivers and dynamic changes in the bottlenecks and session membership. SIM is scalable and fair. In addition, SIM provides superior efficiency and stability than such protocols as RLM where the transmission rates are fixed and the receivers keep trying to join a layer even after they reach the optimal subscription.

The challenging task of combining the intra-group transmission adjustment, selective participation, and menu adaptation mechanisms involves the design of appropriate instantiations for these mechanisms as well as the development of integration techniques such as priority marking in the network. While [8] explains a general philosophy behind our approach, this paper describes a specific protocol and conducts a thorough evaluation of its performance.

The rest of the paper is organized as follows. In Section 2, we present the related work on multicast congestion control, discuss the limitations of existing solutions, and explain the principles that guide our design. The description of our protocol is provided in Section 3. Section 4 describes the simulation environment and the results of our experiments. Finally, Section 5 summarizes our contributions.

2 Design Principles

With multicast congestion control, there is a desire to support heterogeneous receiver capabilities and different bottleneck capacities. A straightforward application of feedback-based congestion control, such as the one used in TCP [1], does provide efficient multicast delivery in heterogeneous environments. Matching the reception capability of a subset of the receivers (e.g., the slowest receiver) leads to a transmission rate that is unsatisfactory for other receivers. This observation leads us to our first design principle:

Principle 1 *A multicast congestion control protocol should adapt efficiently to the heterogeneity in the network and receiver characteristics.*

Selective participation is a congestion control mechanism that improves the efficiency of multicast delivery to heterogeneous receivers, particularly where sources are able to transmit data in multiple complementary layers. The sender of the multicast session transmits data to multiple groups. Each receiver then joins an appropriate subset of the groups based on the observed performance. Selective participation is used for multicast of replicated [7] as well as layered continuous media [13, 21]. Unfortunately, selective participation is characterized by some intrinsic inadequacies. Typically, protocols that are solely based on selective participation use a fixed number of groups that transmit data at fixed rates. Joining and leaving the multicast groups can be viewed as a way for each receiver to choose a suitable service from a menu that describes the transmission options offered by the session. If the menu is fixed while the available bandwidths vary, reliance only on selective participation can result in inefficient utilization of the available bandwidths. It can also cause a significant amount of joins and leaves by receivers trying to find the right group to subscribe to, in the light of a mismatch between the fixed transmission rates of the groups and their variable bottleneck rates. The following example illustrates this scenario.

Example 1 *Consider a multicast session with three groups and three types of receivers (denoted by A, B, and C) with receiving capabilities of 1 Mbps, 2 Mbps, and 3 Mbps respectively. Let the initial cumulative transmission rates of the groups match the bottleneck bandwidths (i.e., the cumulative transmission rates are 1 Mbps, 2 Mbps, and 3 Mbps). In this case, the receivers in A, B, and C subscribe to one, two, and three groups respectively. Now, let the receiving capabilities of the receivers in A, B, and C change to 0.9 Mbps, 1.9 Mbps, and 5 Mbps respectively. With only the selective participation mechanism, the transmission rates of the groups remain unchanged. Hence, to avoid congestion, the receivers in A will need to leave the multicast session completely; the receivers in B will need to drop the second group and consequently will receive 1 Mbps; the receivers in C will continue to subscribe to all the three groups and will receive only 3 Mbps. This illustrates that the receivers in all the groups will receive service that is significantly lower than their capabilities.* ■

The fundamental limits of selective participation can be addressed by an *intra-group transmission adjustment* mechanism. This mechanism can improve efficiency and avoid the undesirable changes in the subscription level by making the transmission rates of the groups match the bottleneck bandwidths. For instance, in the above example, the transmission adjustment mechanism can change the cumulative transmission rates for the three groups to 0.9 Mbps, 1.9 Mbps, and 5 Mbps without incurring any changes in the group memberships. Unfortunately, as the following example illustrates, the combination of selective participation and intra-group transmission adjustment is not sufficient for efficient multicast congestion control when the number of bottleneck bandwidths exceeds the number of groups in the session.

Example 2 *Consider a session with four groups. Let the transmission rate of each group be determined by the majority of the receivers, for which this group is the top subscribed group. Initially, let the receivers of the session belong to four equal-size sets (denoted by A, B, C, and D) with receiving capabilities of 1 Mbps, 2 Mbps, 2.1 Mbps, and 10 Mbps respectively. Also, let the initial cumulative transmission rates match the bottleneck bandwidths. Now, let 40% of the C receivers (denoted as set E) increase their receiving capabilities form 2.1 Mbps to 9 Mbps. Since the receivers in E represent a minority of the subscribers to the third group, the transmission rates of the groups do not change. Hence, due to the selective participation mechanism, the receivers in E periodically join and leave the fourth group. Note that the efficiency of multicast would be higher if the cumulative transmission rates were changed to 1 Mbps, 2 Mbps, 9 Mbps, and 10 Mbps with the receivers in B and C subscribing to two groups, and the receivers in E subscribing to three groups.* ■

Menu adaptation is a congestion control mechanism that allows the sender to adjust its menu and enforce the group subscriptions that improve the overall performance of the multicast session. [6] uses menu adaptation to control congestion for multicast of replicated data. However, solutions based solely on menu adaptation are insufficient because they cannot simultaneously achieve scalability and efficiency in a heterogeneous and dynamic environment.

The above discussion suggests that a multicast congestion control protocol should integrate three mechanisms – intra-group transmission adjustment, selective participation, and menu adaptation – to efficiently adapt to the heterogeneity in the network and receiver characteristics.

With the addition of these mechanisms, however, the amount of state maintained by the multicast congestion control algorithm becomes larger than what is typical for unicast. This has led researchers to observe that multicast congestion control has to operate over much slower time scales than unicast congestion control. However, the nature of applications that use layered multicast (real-time information dissemination is a dominant one) often requires even tighter reaction times for overcoming congestion. We believe that there is a need to adapt to congestion at multiple time scales: some mechanisms have to operate as fast as possible, which is on the order of a round-trip time; other mechanisms have to operate at slower time scales due to the inherent limitations on the speed of their operation (such as join and leave propagation in a multicast tree). This leads us to the following design principle:

Principle 2 *To provide a scalable and timely multicast congestion control, the control mechanisms should operate at multiple time scales.*

Scalability is a key challenge in feedback-based congestion control for multicast because the sender has a limited ability to handle feedback from many receivers. The problem of feedback implosion (excessive information flow towards the sender) can be addressed by aggregating the feedback information or by enforcing an acceptable time scale for reporting the feedback [4, 19]. However, the objective of maintaining privacy precludes designs that avoid feedback implosion by employing the receivers for feedback aggregation. For the same reasons, we find undesirable the solutions similar to RLM where the identity of a receiver is revealed to the other receivers through shared learning. This leads us to our final design principle:

Principle 3 *Mechanisms that control the time scale for reporting feedback and aggregate feedback are essential for scalable, feedback-based multicast congestion control protocol. Such mechanisms, however, should not violate privacy of the receivers.*

In the following section, we describe SIM, a layered multicast congestion control protocol designed on the above principles. Throughout this discussion, we use the term *multicast session* to describe the entire set of groups that participate in the communication of layered data. We assume that the layers are cumulative in nature. The multicast session delivers layered data from one sender to many receivers. A separate multicast group carries each layer.

3 Protocol Design

SIM integrates the following mechanisms operating at different time scales.

- The intra-group transmission adjustment mechanism addresses the need for the sender of a group to obtain the right information to achieve efficient multicast. This mechanism operates at the round-trip time scale.
- The selective participation mechanism operates at an intermediate time scale and places the responsibility with the receivers to subscribe to the appropriate groups.
- The menu adaptation mechanism activates and adapts a menu at the sender at the slowest time scale to ensure the overall efficiency of the multicast delivery.

In what follows, we first discuss the specifics of these three mechanisms and then highlight the techniques used for their integration in SIM.

3.1 Intra-group Transmission Adjustment

The intra-group transmission adjustment mechanism involves per-group congestion detection and notification, aggregation of the feedback information, and transmission adaptation.

Congestion Detection and Notification. When a link is congested, SIM attempts to resolve congestion by adjusting the transmission rate for the highest group that uses this link. To do so, routers, on detecting congestion, mark only those multicast packets that belong to the top group subscribed for the congested link. This approach is somewhat similar to priority dropping [2, 11]. However, our use of priority marking for explicit congestion notification instead of loss-based congestion detection makes our solution more effective and stable.

A naive implementation of SIM would require routers to maintain information about the multicast session associated with each group as well as the ordering of groups within the session. However, the following naming convention reduces the per-session state requirement dramatically:

- Allocate to the groups of a session contiguous multicast addresses;
- Identify the rank of a group within its session using a suffix of the multicast address. The suffix length is $\log_2(n_g)$ bits where n_g is the maximum number of groups within a session; and
- Identify the session of a group using the remaining prefix of the multicast address.

With this convention, each router needs to maintain only $\log_2(n_g)$ bits to indicate the top subscribed group for each session on each link. Since we observed that having more than $n_g = 8$ groups does not generally yield significant improvements in performance, SIM increases the routing table size only by no more than 3 bits per-link per-session. Hence, the additional state introduced into routers by SIM priority marking is minimal.

Note that SIM allows different multicast sessions to employ different numbers of groups. A scenario where session A has multiple groups while session B uses one group does not lead to unfairness or starvation of session A because, when a link becomes congested, the router marks packets from the top subscribed groups of all the sessions on this link. Hence, if sessions A and B share a congested link, the router marks as congested all packets from session B and all packets from the top subscribed group of session A. This ensures that both sessions detect and react to congestion.

We add a small number of fields to the multicast packet header. Multicast packets, as part of their address, include their session and group identifications. To carefully manage the overhead associated with the feedback of congestion information, we also introduce a field, called "feedback request number", in the packet header. As part of the network layer, the header contains a single explicit congestion notification (ECN) bit [15] that congested routers may mark. The receivers send feedback using packets that include: identification number (specifying the source of the feedback), group and session numbers, count of feedback reports aggregated in this packet, count of "congested" reports, minimum and maximum cumulative received rates, number of receivers in the aggregation subtree, and information that facilitates round-trip time computation at the aggregators and sender.

A group starts its transmission when the first receiver subscribes to it. Packets are initially marked as "uncongested" and sent with an interval that is inversely proportional to the current transmission rate of the group.

Routers detect congestion by observing the output link queue. Network routers mark a forwarded multicast packet as "congested" if the length of the queue exceeds a pre-defined threshold and the packet belongs to the top group subscribed for this link. The receivers maintain the congestion state information for its top subscribed group and transmit it to the sender in feedback packets. Once per two round-trip times, the sender adjusts the group transmission rates (based on feedback from the receivers) and requests additional feedback by incrementing the feedback request number in the multicast packet header. Upon receiving a packet with a larger feedback request number than seen before, each receiver transmits a feedback packet to its parent in the aggregation tree. In this feedback packet, the group number specifies the top subscribed group for the receiver, the total count of reports is set to 1, the count of "congested" reports is set to either 0 or 1 depending on whether the top subscribed group is "uncongested" or "congested", the minimum cumulative received rate and maximum cumulative received rate are set to the current estimate of the cumulative received rate, and the number of receivers in the aggregation subtree is set to 1.

Feedback Aggregation. The intra-group transmission adjustment mechanism requires the communication of congestion information from the routers to the receivers and thence back to the sender. To avoid the feedback implosion problem, we postulate the existence of feedback aggregation routers. These may be just the routers in the network that are a part of the multicast tree. However, *not all of the routers in the network have to necessarily perform aggregation of the feedback information*. Selected routers at the branching points in the multicast tree may perform the function of feedback aggregation. One may view the sender and the aggregation routers as forming an aggregation tree in a manner somewhat similar to the approach adopted by the Reliable Multicast Research Group (RMRG) [17] and in reliable multicast schemes such as PGM [20].

Each aggregation node maintains a cache of feedback information received from its children in the aggregation tree. Upon receiving a multicast packet with a larger feedback request number than seen before, each aggregation node empties its cache and starts a timer set to two maximum round-trip times (estimated through online measurements) between the aggregation node and the receivers in its aggregation subtree. When the timer expires or upon collecting responses from all its children in the aggregation tree, the node compiles a summary of feedback information for each group in the cache as follows: the minimum cumulative received rate is set to the minimum of the minimum cumulative received rates reported for this group, the maximum cumulative received rate is set to the maximum of reported maximum cumulative received rates, the count of "congested" reports is set to the sum of the counts of "congested" reports, and the number of receivers in the aggregation subtree is set to the sum of the reported numbers of receivers in the aggregation subtrees. Then, the node sends a feedback packet with the compiled summaries and its identification number to the parent in the aggregation tree.

To determine when to send consolidated feedback to the parent in the aggregation tree, aggregators rely on timeouts because feedback packets from the receivers may get delayed or lost. If the timeout values are not chosen carefully, feedback information from some children can arrive at the aggregator after the aggregated feedback is sent. The absence of this information from the aggregated feedback is referred to as aggregation noise and can affect the efficiency of congestion control. Thus, feedback aggregation schemes face a tradeoff between aggregation noise and the responsiveness of congestion control. Larger the timeouts, the smaller the aggregation noise but poorer the responsiveness.

For each aggregator, SIM sets the timeout value to two maximum round-trip times between the aggregator and the receivers (i.e., leaf nodes) in its aggregation subtree. Hence, the timeout value at an aggregator depends on the delay between the aggregator and the furthest receiver in its subtree. Our experiments showed that this setting reduces the probability of introducing aggregation noise dramatically, while ensuring that the sender receives a feedback within two maximum round-trip times of the session. So, SIM addresses the problem of aggregation noise carefully. While aggregation noise is, in our opinion, a significant problem, it has often been ignored in previous work.

The delay, complexity, and overhead of our approach do not increase when the number of receivers or the number of levels in the aggregation tree grows. Finally, SIM adheres to the principle of privacy: (1) receivers transmit feedback only when deemed appropriate by the sender and that receiver; and (2) the only entity that is trusted by a receiver are the aggregation routers in the network.

Transmission Adjustment. The sender adjusts the group transmission rates based on feedback from the receivers. The transmission rate of each group is adapted using an additive-increase multiplicative-decrease algorithm: if the group is "congested", its rate is reduced to a fraction of the current value; if the group is not "congested", then the rate of the group is increased by a constant amount.

The criteria for determining whether a group is "congested" vary for different groups. The bottom group (group 0) is governed by the capabilities of the slowest receiver and is considered "congested" when at least one "congested" report is received for this group. The top group (group $(n_g - 1)$, where n_g is the number of groups) satisfies the capabilities of the fastest receiver and is considered "congested" when all reports for this group are "congested". Transmission adjustment in the middle groups (groups 1 through $(n_g - 2)$) depends on the menu. If the menu is inactive, a middle group is considered "congested" when at least one "congested" report is received for this group. This allows SIM to discover the full range of available bandwidths promptly. Once the menu is active, a middle group is considered "congested" when the majority of its reports are "congested". This allows SIM to improve the overall efficiency of the multicast session when the number of bottleneck bandwidths exceeds the number of groups within the session. If the rate suggested for middle group k by the additive-increase multiplicative-decrease algorithm is such that the cumulative transmission rate for groups 0 through k exceeds the corresponding limit in the active menu, then the transmission rate for group k is reduced to the value that makes the cumulative transmission rate equal to the menu limit. The next feedback request is sent one maximum round-trip time after the transmission adjustment so that the solicited feedback could reflect the impact of the adjustment.

Note that the effectiveness of SIM depends on the ability to change the encodings of the multicast layers. It has been shown that it is possible to adjust the encodings of continuous media streams quickly and with fine granularity. For instance, modification of the quantization parameters represents a successful method of adjusting the sending rate for compressed video [9, 12].

3.2 Selective Participation

The mechanism of selective participation allows receivers to determine the right groups to join within a session. Initially, each receiver subscribes only to the bottom group. The receiver adds or drops groups if its feedback consistently fails to affect the transmitted rate in a desired fashion. If the receiver is not subscribed to all the groups, is not "congested" for a fixed number of consecutive congestion notifications, and the transmission rate for its top subscribed group is reduced (either because of some other congested receivers or as a result of menu enforcement), the receiver adds the next immediate group above its currently top subscribed group. If the receiver is subscribed to at least two groups and is "congested" for a fixed number of consecutive congestion notifications, and the transmission rate for its top subscribed group is increased, the receiver drops its top subscribed group.

3.3 Menu Structure and Adaptation

As we showed in Section 2, the combination of intra-group transmission adjustment and selective participation mechanisms is not sufficient for efficient multicast congestion control when the number of bottleneck bandwidths exceeds the the number of groups. However, when a session starts, the number of bottleneck bandwidths is not known. Further, the number of bottleneck bandwidths may change over time.

To address this, SIM supports a menu adaptation mechanism allowing the sender to discover the appropriate group transmission rates that improve the overall efficiency of the session. Using this mechanism, the sender maintains a menu that can be either active or inactive. Initially, the menu is inactive. The status and values of the menu are adjusted once per every menu adaptation interval. At menu adaptation time, the sender computes the minimum and maximum cumulative received rates based on the feedback from the receivers. The menu is marked as active once the top group of the session has seen congestion and if the difference between the minimum and maximum received rates exceeds a threshold. Requiring the difference to exceed the threshold makes the menu adaptation mechanism robust in the presence of inaccuracies in the measurements of the receiver capabilities. If the menu is active, the sender splits the range between the minimum and maximum received rates into $(n_g - 2)$ equal subranges and uses the upper boundaries of these subranges as the maximum cumulative transmission rates for corresponding middle groups.

3.4 Integration Techniques

SIM integrates the above three schemes so that the mechanisms operating on slower time scales complement and take advantage of the control provided by the faster mechanisms. Further, the aggregation algorithms are tuned to provide robust and timely feedback for both the intra-group transmission adjustment and menu adaptation mechanisms. We also discovered that priority marking is essential for the convergence of SIM to a stable efficient state.

SIM uses an exponentially-weighted estimation for measuring (1) the received rates at the receivers; (2) the maximum round-trip time at the sender and at the aggregation nodes; and (3) the queue sizes at the routers. The estimation of the received rate is updated at every congestion notification time if at least two packets have been received since the previous update. The update of the round-trip times is performed when the transmission rate is adjusted or feedback is sent to the parent in the aggregation tree. The estimate of the queue size is updated upon packet departures, and the estimate is set to 0 when the link is idle.

4 Experimental Evaluation

We evaluated SIM using NS-2 network simulator [14]. While SIM exhibited similar behaviors in all examined

d	Round-trip propagation delay
t	Time since the beginning of the session
t_f	Session duration
$n_f = \lfloor \frac{t_f}{d} \rfloor$	Number of complete intervals of duration d in a session
n_g	Maximum number of groups in a session
$n_b(t)$	Number of different bottleneck bandwidths at time t
$s(t)$	Set of receivers in a session at time t
$n(t) = \lvert s(t) \rvert$	Number of receivers in a session at time t
t_b	Interval between changes in bottleneck bandwidths
$p = \frac{t_b}{d}$	Number of round-trip delays between bandwidth changes
t_m	Interval between changes in the session membership
$m = \frac{t_m}{d}$	Number of round-trip delays between membership changes
$a_i(t)$	Available bandwidth for receiver i at time t
$r_i(t)$	Received rate for receiver i at time t
b	Ratio of the bandwidth range to the transmission range

Table 1: Definitions of the experimental variables.

topologies, we present only the results for the network topology depicted in Figure 1. This topology is a balanced tree where node 0 is the sender, the leaf nodes are the receivers, and the internal nodes aggregate feedback. We assume that all links have the same propagation delay which is selected such that the maximum round-trip propagation time d between the sender and the receivers is equal to 40 ms (roughly, round-trip propagation delay between the east and the west coasts of the United States). Thus, the propagation delay for each link in the topology with 16 leaf nodes is set to 5 ms. We refer to links by stating their end nodes, e.g., the link between nodes 0 and 1 is denoted as link 0-1. Table 1 defines the variables used in our experiments.

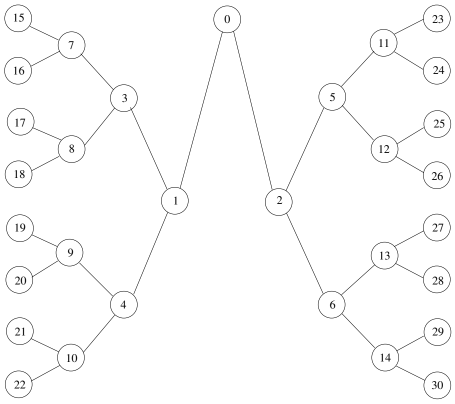


Figure 1: Network topology.

We examine the behavior of a SIM session in Section 4.2.1. Then, we evaluate SIM for heterogeneous bottleneck bandwidths and dynamic changes in the bottlenecks and session membership (Sections 4.2.2 through 4.2.4). After we study scalability (Section 4.2.5) and fairness (Section 4.2.6) of SIM, we compare SIM with RLM in Section 4.2.7.

4.1 Performance Measures

For each experiment, we measure as a function of t: (1) transmission rates for each of the groups within a SIM session, (2) received rates for each of the receivers, (3) menu of the maximum cumulative group transmission rates, and (4) group subscriptions.

To evaluate the effectiveness of bandwidth utilization and the stability of subscription levels, we introduce the following two summary performance measures:

- *Efficiency*: We define efficiency $E(t)$ as the average achieved utilization at time t. Formally,

$$E(t) = \frac{1}{n(t)} \sum_{i \in s(t)} \frac{r_i(t)}{a_i(t)}; \quad 0 \leq E(t) \leq 1$$

We compute the average efficiency $\widehat{E}$ and the deviation in efficiency $\tilde{E}$ as follows:

$$\widehat{E} = \frac{1}{t_f} \int_0^{t_f} E(t) dt; \quad \tilde{E} = \sqrt{\frac{1}{t_f} \int_0^{t_f} \left(E(t) - \widehat{E}\right)^2 dt}$$

- *Instability*: We define instability $I(t)$ as the number of changes in the subscription level per receiver during the interval $[t-d, t)$. We compute the average instability $\widehat{I}$ and the deviation in instability $\tilde{I}$ as follows:

$$\widehat{I} = \frac{1}{n_f} \sum_{j=1}^{n_f} I(jt); \quad \tilde{I} = \sqrt{\frac{1}{n_f} \sum_{j=1}^{n_f} \left(I(jt) - \widehat{I}\right)^2}$$

Instability is an important metric since it is directly related to the perceived quality of live video multicast and other continuous media applications. Smaller values of instability result in greater user satisfaction.

4.2 Experimental Results

4.2.1 Session Behavior

Consider a SIM session with 16 receivers and a maximum of 5 groups (i.e., $n(t) = 16$ and $n_g = 5$). We configured links 1-3, 9-19, 1-4, 5-11, 12-25, and 6-13 to have bandwidths of 1, 2, 3, 3, 4, and 5 Mbps respectively; all other links have a bandwidth of 6 Mbps. There are 6 different bottleneck bandwidths in this topology (i.e., $n_b(t) = 6$). It has 4 receivers behind the 1 Mbps bottleneck, 1 receiver behind the 2 Mbps bottleneck, 5 receivers behind the 3 Mbps bottlenecks, 1 receiver behind the 4 Mbps bottleneck, 2 receivers behind the 5 Mbps bottleneck, and 3 receivers behind the 6 Mbps bottleneck.

Figure 2 depicts the behavior of the SIM session. During the initial stage, the intra-group transmission adjustment and selective participation mechanisms allow the session to discover its highest available bandwidth: Figure 2(a) shows that the cumulative transmission rate for all 5 groups converges to 6 Mbps. Then, since $n_g < n_b(t)$, the SIM session activates the menu (see Figure 2(b)) and changes the rules for intra-group transmission adjustment. For instance, group 3 becomes governed not by its slowest 4 Mbps receiver but by the two receivers behind the 5 Mbps bottleneck (i.e., by the majority among the receivers subscribed exactly for 4 groups). The effect of this is shown in Figure 2(a) where the cumulative transmission rate for 4 groups increases from 4 Mbps to 5 Mbps.

After the cumulative transmission rate for 4 groups converges to 5 Mbps, the session enters a steady state of its operation. This steady state is characterized by the following behavior (see Figure 2(a)): the bottom group transmits at 1 Mbps, the 2 bottom groups transmit at the total of 2 Mbps, the 3 bottom groups transmit at the total of 3 Mbps, the 4 bottom groups transmit at the total of 5 Mbps, and all 5 groups transmit cumulatively at 6 Mbps. Figure 2(c) depicts the group subscriptions in the steady state: top group (group 4) is subscribed to by the three 6 Mbps receivers; group 3 is subscribed to by the three 6 Mbps and two 5 Mbps receivers; group 2 is subscribed to by the three 6 Mbps, two 5 Mbps, one 4 Mbps, and five 3 Mbps receivers; group 1 is subscribed to by all but the four 1 Mbps receivers; and group 0 is subscribed to by all the receivers. The 4 Mbps receiver (receiver 25) subscribes and unsubscribes to group 3, creating the fluctuations in its received rate (see the third curve from the top in Figure 2(d)) and in the number of subscribers for 4 groups (see the second curve from the bottom in Figure 2(c)).

Figures 2(a) and 2(d) demonstrate that the received rates for the six receivers with different bottleneck bandwidths track the cumulative transmission rates. This indicates that SIM results in a very small amount of packet losses in the network. Figure 2(b) shows how the menu fluctuates with the variation in the received rates for the slowest and the fastest receivers. Figures 2(e) and 2(f) demonstrate that after the SIM session reaches its steady state, its efficiency stabilizes close to the optimal value of 1 while its instability reduces dramatically.

4.2.2 Heterogeneity in Bottleneck Bandwidths

We evaluate the performance of our protocol for different values of $n_b(t)$ when $n_g = 5$, $n(t) = 16$, and $t_f = 300$ seconds. We configure the network such that all links, except the ones incident on the receivers, have a bandwidth of 6 Mbps. We uniformly divide the range between 1 Mbps and 6 Mbps into $n_b(t)$ values, and assign these bandwidths to the links incident on the receivers such that the number of receivers with each value of available bandwidth is approximately the same.

We observed that when $n_b(t) \leq n_g$, efficiency of SIM is high, and group subscriptions change only during the initial stage of convergence. Figure 3 shows that when $n_b(t) > n_g$, instability increases and efficiency drops. In either case, SIM converges to the steady state. These results indicate that, to maximize efficiency and to minimize instability, the number of groups within a SIM session should match the number of different bottleneck bandwidths.

4.2.3 Dynamically Changing Bottlenecks

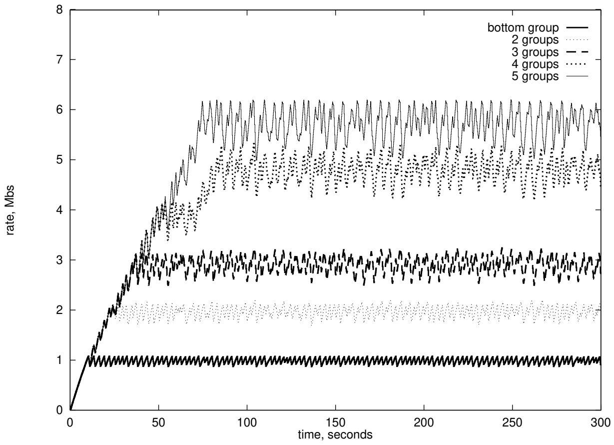
Bottlenecks can experience two types of changes: (1) in their bandwidth and (2) in their location in the network. First, we examine a scenario when the bottlenecks do not migrate but their bandwidths fluctuate. Our experiments show (we omit the corresponding graphs) that SIM maintains high efficiency and constant instability for all the examined values of the amounts and frequencies of the bandwidth fluctuations. The intra-group transmission adjustment mechanism successfully adapts to the changes in the available bandwidths, and the selective participation mechanism is not triggered. This distinguishes SIM from such solutions as RLM, in which changes in the bottleneck bandwidths can cause significant changes in the subscription levels.

Now, we consider a more general scenario where both the bandwidths and locations of the bottlenecks change. We conduct this experiment with $n_g = 3$, $n_b(t) = 3$, $n(t) = 16$, and $t_f = 300$ seconds. All the links, other than 0-2, 2-5, 1-4, and 4-9, have a bandwidth of 9 Mbps. Links 2-5 and 4-9 have bandwidths of 3 Mbps and 7 Mbps respectively. Once every p round-trip delays, the bandwidth of link 0-2 alternates between 9 and 1 Mbps while the bandwidth of link 1-4 alternates between 9 and 5 Mbps. These settings ensure that, once every p round-trip delays, (1) the bottleneck for receivers 19 and 20 migrates between links 1-4 and 4-9, (2) the bottleneck for receivers 21 and 22 migrates between links 1-4 and 0-1, (3) and the bottleneck for receivers 23 through 26 migrates between links 0-2 and 2-5. Note that these fluctuations do not change the number of bottleneck links (i.e., $\forall\, t:\ n_b(t) = 3$).

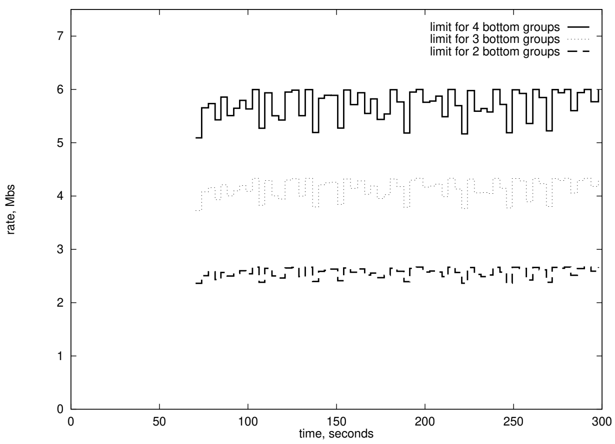
Figure 4 shows that when the bottleneck links migrate at a slower time scale than the selective participation mechanism operates, the receivers maintain high efficiency by changing their subscription levels in response to the changes in the available bandwidths. When the bottleneck migration is more frequent, SIM provides a stable, though not optimally efficient, group membership.

4.2.4 Dynamic Changes in Session Membership

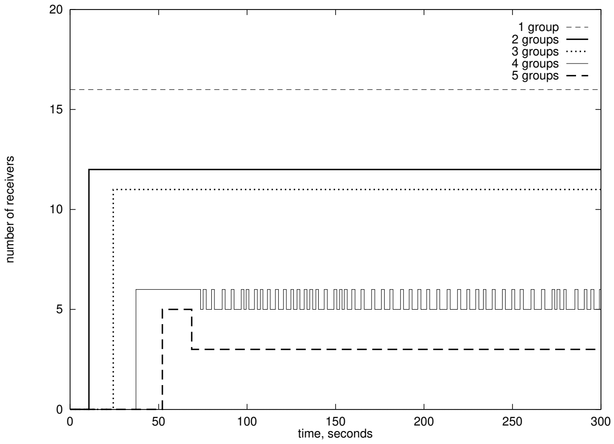
We examine the behavior of SIM when the session membership varies. Note that changes in the session subscription impact performance the most when they modify the menu. Hence, we experiment with scenarios where the receivers dominating the transmission in one of the groups join and leave the session synchronously. We conduct this experiment with a SIM session with three groups ($n_g = 3$) and $t_f = 300$ seconds. Links 2-5, 1-3, and 4-9 have bandwidths of 1 Mbps, 3 Mbps, and 5 Mbps respectively; all other links have 9 Mbps bandwidths. Once every m round-trip delays, all the receivers behind the bottlenecks with a specific bandwidth synchronously subscribe to or unsubscribe from the session.



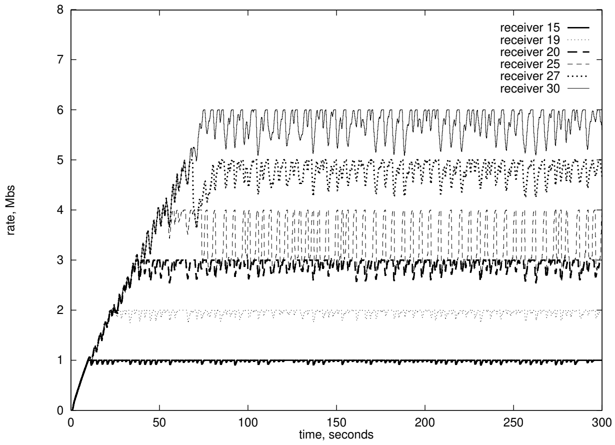
(a) Cumulative group transmission rates



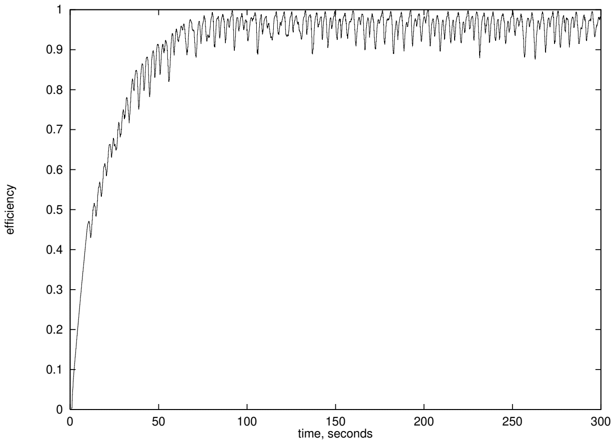
(b) Menu adaptations



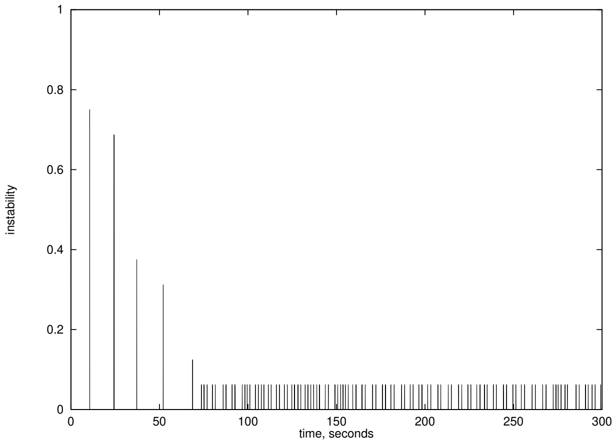
(c) Group subscriptions



(d) Received rates

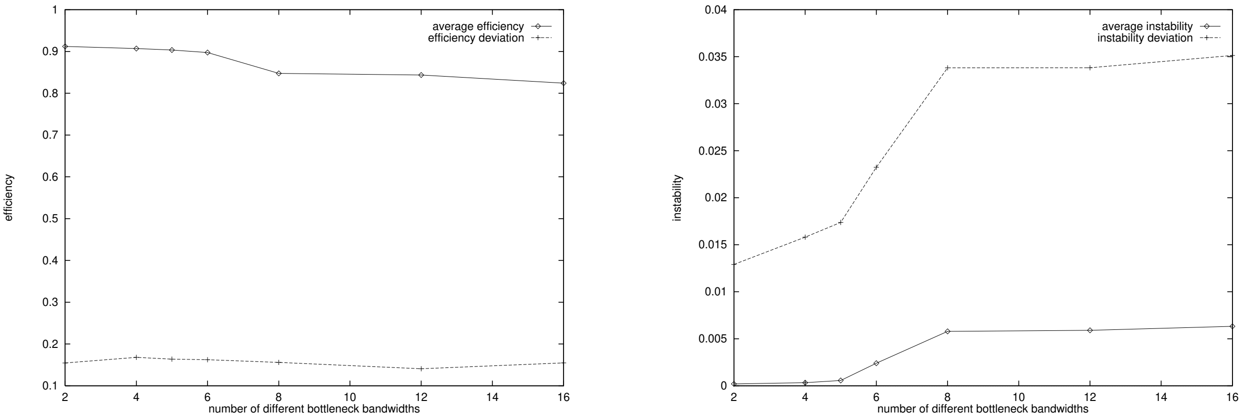


(e) Efficiency



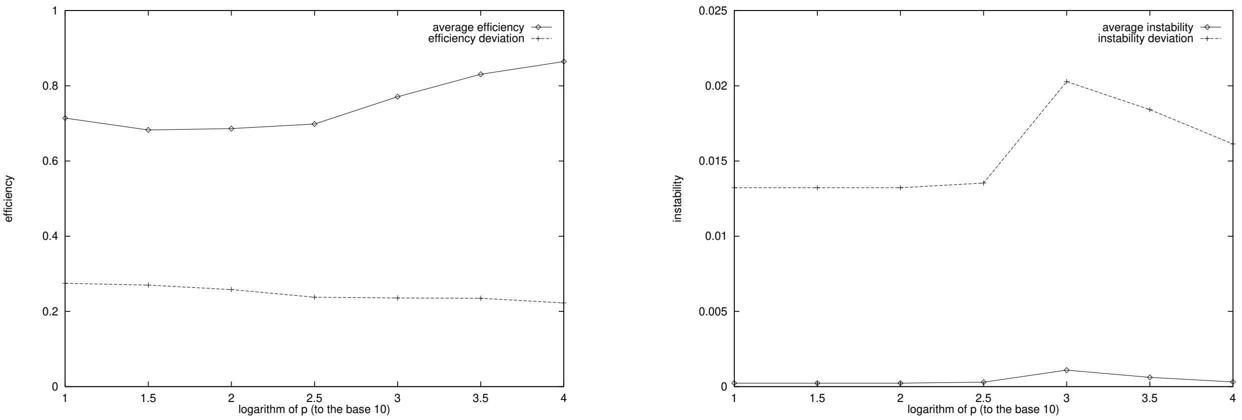
(f) Instability

Figure 2: Understanding the behavior of a SIM session with $n_b(t) = 6$, $n_g = 5$, $n(t) = 16$, and $t_f = 300$ seconds.



(a) Efficiency (b) Instability

Figure 3: Efficiency and instability for different values of $n_b(t)$ when $n_g = 5$, $n(t) = 16$, and $t_f = 300$ seconds.



(a) Efficiency (b) Instability

Figure 4: Efficiency and instability when bottleneck links migrate: $n_g = 3$, $n_b(t) = 3$, $n(t) = 16$, and $t_f = 300$ seconds.

We vary the value of m from 10 (i.e., when the receivers join and leave at the time scale of selective participation mechanism) to $1,000$ (i.e., when no receivers unsubscribe during the session).

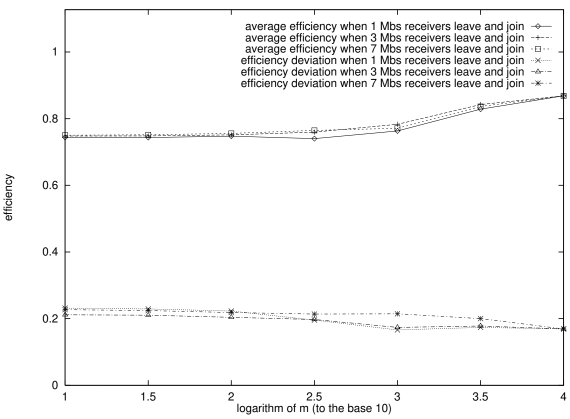
Figure 5(a) shows that SIM maintains high efficiency for a wide range of values for m (the top three curves are average efficiencies; the bottom three curves are efficiency deviations). Since dynamic join and leave of receivers to/from the session involve subscribing and unsubscribing them to/from individual groups, and possibly trigger menu adaptation, instability is higher for smaller values of m (see Figure 5(b)). Efficiency increases and instability decreases with reduction in the frequency of the changes in the session membership.

4.2.5 Scalability

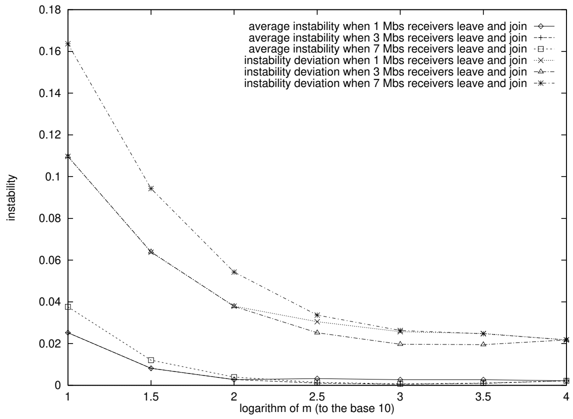
Figure 6 studies scalability of SIM when the number of receivers increases up to 1024. We configure all links to have a bandwidth of 6 Mbps, except for links 1-3, 1-4, 2-5, and 6-13 that have bandwidths of 2, 3, 4, and 5 Mbps respectively. Thus, $n_b(t) = 5$. We experiment with a SIM session with $n_g = 5$ and $t_f = 300$ seconds. Figure 6 shows that our protocol is scalable: as the number of receivers grows, instability remains the same while the average efficiency decreases just slightly.

4.2.6 Fairness

In this experiment, we study the intra-protocol fairness of SIM. We define a protocol to be fair if it provides equivalent services to the receivers of the sessions that use the same set of network resources (see [3, 10, 16, 18] for other notions of multicast fairness). We consider two sessions, denoted as Session A and Session B, with $n_b(t) = 3$, $n_g = 3$, and $n(t) = 16$. Both sessions multicast their data from node 0 to all the sixteen receivers. We configured links 0-1, 0-2, and 6-14 to have bandwidths of 6, 4, and 2 Mbps respectively; all other links have bandwidths of 10 Mbps. Thus, this topology has 8 receivers (including receiver 15) behind the 6 Mbps bottleneck link, 6 receivers (including receiver 23)

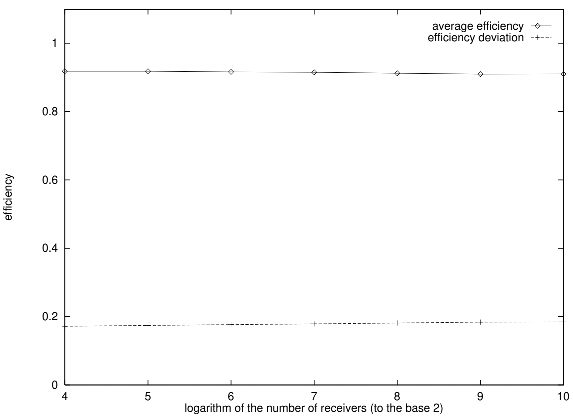


(a) Efficiency

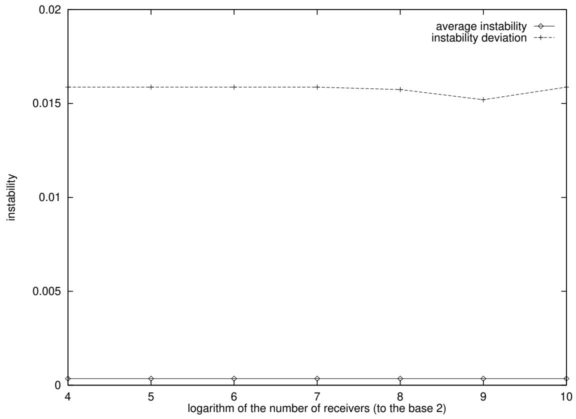


(b) Instability

Figure 5: Efficiency and instability when receivers join and leave a SIM session: $n_g = 3$ and $t_f = 300$ seconds.



(a) Efficiency



(b) Instability

Figure 6: Scalability of SIM: $n_g = 5$, $n_b(t) = 5$, and $t_f = 300$ seconds.

behind the 4 Mbps bottleneck link, and 2 receivers (including receiver 30) behind the 2 Mbps bottleneck link.

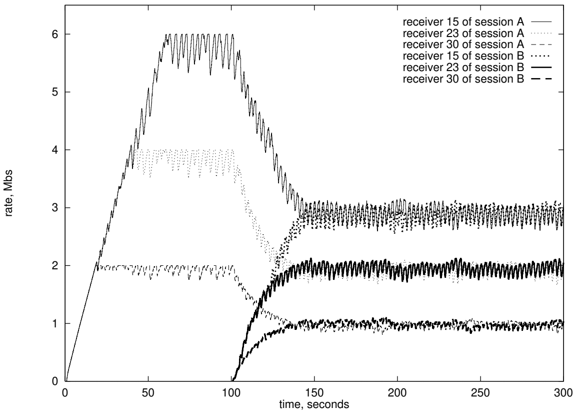
Figure 7 shows the received rates of representative receivers in scenarios when: (1) Session B starts its transmission after Session A converges to a steady state and (2) Session B becomes active before Session A reaches a steady state. In both cases, the receivers of both sessions discover their optimal subscription levels, and their received rates converge to their fair shares.

4.2.7 Comparison with RLM

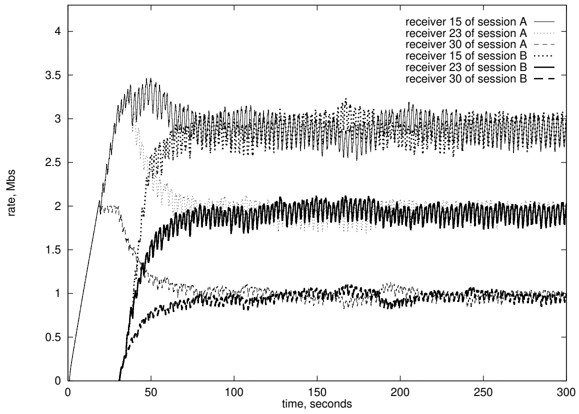
We compare SIM with RLM since RLM is a classical protocol for layered multicast. Our comparison emphasizes a fundamental property of RLM – its reliance on the mechanism of selective participation – rather than its intricate details. Hence, we believe that our results hold for other protocols (such as RLC [21] and FLID [5]) that share this fundamental property with RLM.

Since an RLM session transmits at predetermined rates, its efficiency can suffer from a mismatch between the transmission rates and the bottleneck bandwidths. We measure this effect and explore whether it can be alleviated by increasing the number of groups in the RLM session. We represent the degree of the mismatch by ratio b of the bandwidth range to the transmission range, where the bandwidth range is the difference between the highest and the lowest bottleneck bandwidths, and the transmission range is the difference between the maximum cumulative transmission rate of the RLM session and the transmission rate of its bottom group.

We compare the performance of a multicast session under SIM and under RLM for $n_b(t) = 5$, $n(t) = 16$, and $t_f = 300$ seconds. Links 1-3, 1-4, 2-5, 14-30, and 2-6 are bottlenecks listed in the increasing order of their bandwidths. These bandwidths are picked randomly, under the assumption of the uniform distribution, from an interval centered at 3 Mbps. All the other links have a bandwidth of 6 Mbps. In the case of RLM, the session uses n_g groups with cumulative transmission rates distributed uniformly from 1 to 5 Mbps,



(a) Session B starts after Session A converges



(b) Session B starts before Session A converges

Figure 7: Studying the fairness of SIM for two sessions with $n_b(t) = 3$, $n_g = 3$, and $n(t) = 16$.

i.e., its bottom group transmits at a rate of 1 Mbps while each of the upper groups sends at a rate of $\frac{4}{n_g-1}$ Mbps.

Figure 8 compares the performance of RLM and SIM for $n_g = 5$ when ratio b of the bandwidth range to the transmission range varies from 0.1 to 1. SIM consistently exhibits superior efficiency and stability. While the cumulative transmission rates in SIM converge to the bottleneck bandwidths, the efficiency of RLM suffers from the mismatch between the bottleneck bandwidths and predetermined transmission rates. RLM is especially inefficient when the bandwidth of link 1-3 is below 1 Mbps. In such scenarios (which occur for b of at least 0.8), receivers 15, 16, 17, and 18 can not sustain even the rate of the bottom group, and efficiency of RLM drops to about 60%. Moreover, instability of the RLM session is higher since its receivers keep trying, unlike in SIM, to join a group even after they reach the optimal subscription.

Figure 9 compares the performance of RLM and SIM for different numbers of groups when $b = 1$. Note that a large number of groups does not ensure a good performance for RLM. When the RLM session uses more groups, the new set of its transmission rates can be worse in terms of matching the available bandwidths. Further, a larger number of groups in the RLM session can lead to slower convergence since the receivers may need to join more groups to reach the optimal subscription levels. This can result in lower efficiency and higher instability. On the other hand, even when the maximum number n_g of groups exceeds 5, the SIM session transmits using only five groups (with cumulative transmission rates that match the bottleneck bandwidths), and its efficiency and stability remain on high and constant levels.

5 Conclusions and Future Work

In this paper, we designed SIM, a multicast congestion control protocol for dissemination of layered data to large populations of heterogeneous receivers. SIM integrates three congestion control mechanisms to provide efficient, stable and scalable multicast delivery that does not violate privacy of the receivers. Per our knowledge, SIM is the first protocol for layered multicast that adjusts not only the subscription levels of the receivers but also the transmission rates of the layers. We demonstrated that SIM is highly efficient and stable in the presence of heterogeneous receivers and dynamic changes in the bottlenecks and session membership. We also showed that SIM is scalable and fair. By adapting the cumulative transmission rates to match the bottleneck bandwidths, SIM provides better efficiency than RLM. Moreover, instability of RLM sessions is higher since their receivers keep trying, unlike SIM receivers in the steady state, to join a layer even after they reach the optimal subscription. We showed that SIM is superior to RLM in terms of efficiency and stability even when RLM employs a much larger amount of layers.

While we showed that multiple SIM sessions share the available bandwidth fairly, we defer assessment of inter-protocol fairness to our future work. We also plan to extend our approach to provide a scalable, efficient, and stable congestion control for multicast sessions with multiple senders.

References

[1] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control. RFC 2581, April 1999.

[2] S. Bajaj, L. Breslau, and S. Shenker. Uniform versus Priority Dropping for Layered Video. In *Proceedings ACM SIGCOMM'98*, August 1998.

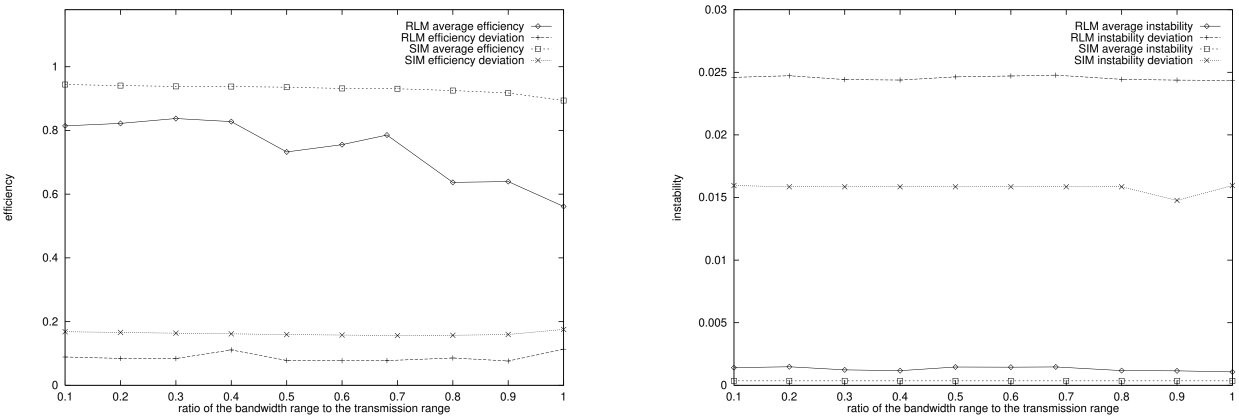
[3] S. Bhattacharyya, D. Towsley, and J. Kurose. The Loss Path Multiplicity Problem for Multicast Congestion Control. In *Proceedings IEEE INFOCOM'99*, March 1999.

[4] J-C. Bolot, T. Turletti, and I. Wakeman. Scalable Feedback Control for Multicast Video Distribution in the Internet. In *Proceedings ACM SIGCOMM'94*, October 1994.

[5] J. Byers, M. Frumin, G. Horn, M. Luby, M. Mitzenmacher, A. Roetter, and W. Shaver. FLID-DL: Congestion Control for Layered Multicast. In *Proceedings NGC 2000*, November 2000.

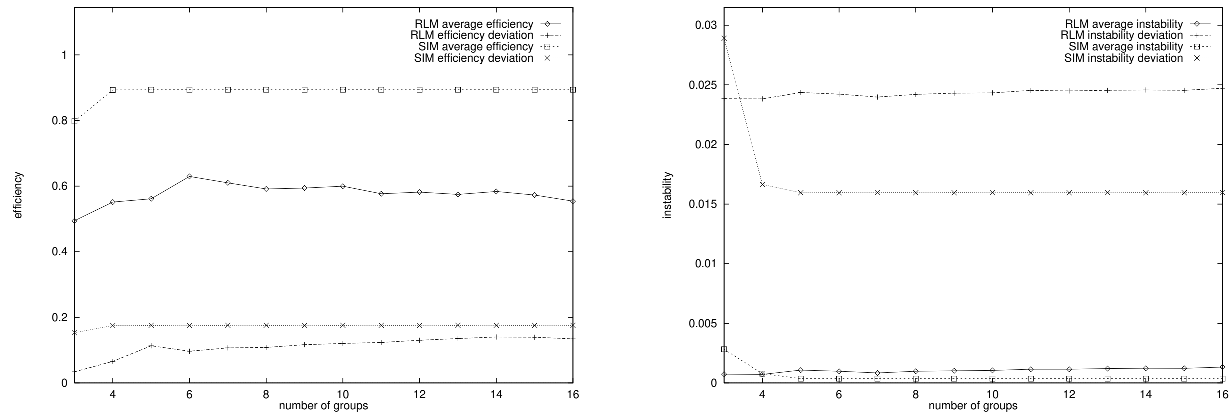
[6] S. Y. Cheung and M. H. Ammar. Using Destination Set Grouping to Improve the Performance of Window-controlled Multipoint Connections. *Computer Communications Journal*, 19:723–736, 1996.

[7] S. Y. Cheung, M. H. Ammar, and X. Li. On the Use of Destination



(a) Efficiency (b) Instability

Figure 8: Comparison of SIM and RLM for $n_b(t) = 5$, $n(t) = 16$, $n_g = 5$, and $t_f = 300$ seconds.



(a) Efficiency (b) Instability

Figure 9: Comparison of SIM and RLM when $n_b(t) = 5$, $n(t) = 16$, $b = 1$, and $t_f = 300$ seconds.

Set Grouping to Improve Fairness in Multicast Video Distribution. In *Proceedings IEEE INFOCOM'96*, March 1996.

[8] N.G. Duffield, M. Grossglauser, and K.K. Ramakrishnan. Distrust and Privacy: Axioms for Multicast Congestion Control. In *Proceedings NOSSDAV'99*, June 1999.

[9] N.G. Duffield, K.K. Ramakrishnan, and A.R. Reibman. SAVE: an Algorithm for Smoothed Adaptive Video over Explicit Rate Networks. *IEEE/ACM Transactions on Networking*, 6(6):717–728, December 1998.

[10] J. Golestani and K. Sabnani. Fundamental Observations on Multicast Congestion Control in the Internet. In *Proceedings IEEE INFOCOM'99*, March 1999.

[11] R. Gopalakrishnan, J. Griffioen, G. Hjalmtysson, C. Sreenan, and S. Wen. A Simple Loss Differentiation Approach to Layered Multicast. In *Proceedings IEEE INFOCOM'2000*, March 2000.

[12] T.V. Lakshman, P.P. Mishra, and K.K. Ramakrishnan. Transporting Compressed Video over ATM Networks with Explicit-Rate Feedback Control. *IEEE/ACM Transactions on Networking*, 7(5):710–723, October 1999.

[13] S. McCanne, V. Jacobson, and M. Vetterli. Receiver-driven Layered Multicast. In *Proceedings ACM SIGCOMM'96*, August 1996.

[14] UCB/LBNL/VINT Network Simulator NS-2. http://www-mash.cs.berkeley.edu/ns, January 2000.

[15] K.K. Ramakrishnan and S. Floyd. A Proposal to add Explicit Congestion Notification (ECN) to IP. RFC 2481, January 1999.

[16] L. Rizzo. pgmcc: A TCP-friendly Single-Rate Multicast Congestion Control Scheme. In *Proceedings ACM SIGCOMM'2000*, August 2000.

[17] Reliable Multicast Research Group. http://www.east.isi.edu/rm, January 2000.

[18] D. Rubenstein, J. Kurose, and D. Towsley. The Impact of Multicast Layering on Network Fairness. In *Proceedings ACM SIGCOMM'99*, September 1999.

[19] H. Schulzrinne and J. Rosenberg. Timer Reconsideration for Enhanced RTP Scalability. In *Proceedings IEEE INFOCOM'98*, March 1998.

[20] T. Speakman et al. PGM Reliable Transport Protocol Specification. Internet Draft draft-speakman-pgm-spec-04.txt, April 2000.

[21] L. Vicisano, L. Rizzo, and J. Crowcroft. TCP-like Congestion Control for Layered Multicast Data Transfer. In *Proceedings IEEE INFOCOM'98*, March 1998.

[22] B. Vickers, C. Albuquerque, and T. Suda. Source-Adaptive Multi-Layered Multicast Algorithms for Real-Time Video Distribution. *IEEE/ACM Transactions on Networking*, December 2000.