Inter-Receiver Fair Multicast Communication Over the Internet *

Tianji Jiang Ellen W. Zegura Mostafa Ammar

Networking and Telecommunications Group
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
{tjiang,ewz,ammar}@cc.gatech.edu

Abstract

Multicast protocols target applications involving a large number of receivers with heterogeneous data reception capabilities. To accommodate heterogeneity, the sender may transmit at multiple rates, requiring mechanisms to determine the rates and allocate receivers to rates. In this paper, we develop a protocol to control the rates of a multicast session, with the goal of maximizing the inter-receiver fairness, *an intra-session measure that captures the collective "satisfaction" of the session receivers. Our target environment is the Internet, where fair sharing of bandwidth must be achieved via end-system mechanisms and fairness to TCP is important. We develop and evaluate protocols to maximize this measure by maintaining a fixed-rate base group and a variable-rate group. We show that our schemes offer improvement over single-rate sessions, while maintaining TCP-friendliness.*

1 Introduction

Multicast protocols target applications involving a large number of receivers with heterogeneous data reception capabilities. Depending on the type of application, this heterogeneity can be accommodated in one of two ways. In a *single-rate* session, the source adjusts its sending rate based on feedback it receives from the network and/or the receivers. In a typical single-rate protocol (e.g., [1]), the rate is picked to match the lowest capacity receiver (or path to a receiver). In a *multi-rate* session the sender can transmit at different rates to different receivers through layering [2, 3, 4] or destination-set splitting [5, 6] In either case, there needs to be (1) criteria for the setting of the session rate(s) and the allocation of receivers to the rates (in the case of multi-rate sessions) and (2) protocols for implementing the appropriate rate settings and allocations.

In this paper we aim to develop a protocol to control the rate of a multicast session with the goal of maximizing *inter-receiver fairness*, an intra-session measure that captures the collective "satisfaction" of the session receivers, based on the rate at which they are receiving and the data loss they are experiencing. We also strive to achieve inter-session fairness among similarly controlled multicast sessions and between multicast sessions and TCP sessions. In [7] we first introduced the concept of inter-receiver fairness together with a protocol to control the rate of an ATM multicast ABR connection in a manner that maximizes inter-receiver fairness.

This paper is concerned with developing a similar protocol within the Internet environment. The current Internet differs significantly from the ATM ABR environment in which our previous protocol operated. Our work targets the current (and near-term) best-effort Internet, in which the fair sharing of links is primarily the responsibility of end-to-end rate control protocols. In addition (and unlike ATM), because Internet routers do not currently possess the capability to aggregate feedback returned from receivers, the scalability of the feedback is of primary concern. Finally, an emerging requirement for rate adjustment protocols operating over the Internet is TCP-friendliness [8]. Our protocol encompasses features to deal with all the Internet constraints mentioned above.

In Section 2 we first give a brief overview of our proposed protocol and its features. We present the definition of inter-receiver fairness formally in Section 3. In Section 4, we discuss the protocol operation in detail, including features that are designed to make it TCP-friendly. In Section 5 we present results from simulations of our protocol to demonstrate its behavior. The paper is concluded in Section 6.

2 Protocol Overview

We assume that each receiver can estimate its *isolated rate*. Informally, this is the rate that the receiver should receive under fair sharing of the network resources, but

*This work is supported in part by a research grant from NSF under contract NCR-9628379.

without being limited by other receivers in the same session [1]. The isolated rate concept was first defined in [7], and we use recently developed formalisms for the fair rate allocation in multi-rate multicast sessions [9] to formalize our original definition as described later. Typically, the isolated rate varies over time depending on network traffic conditions.

A receiver is "unhappy" if it belongs to a group with higher or lower rate than its isolated rate. If the rate is higher than the receiver's fair share, we assume that losses will result; if the rate is lower, the receiver is artificially constrained. The extent of unhappiness at receiver i with group rate r is represented by a *fairness function*, $F_i(r)$, which achieves a maximum at a receiver's isolated rate. We define the *inter-receiver fairness* for a given session as a weighted sum of the fairness values of the receivers in the session. The main objective of a rate allocation protocol is to operate at a rate that maximizes the inter-receiver fairness within other constraints that may be imposed (e.g., being TCP-friendly).

Because a rate that maximizes inter-receiver fairness can exceed some receivers' isolated rates, such receivers will experience data loss. We assume that each receiver can specify a *loss tolerance* which allows the session rate to exceed its isolated rate as long as the operation is within the loss tolerance. Such losses, however, can be problematic within the Internet, particularly under the assumption of drop-tail, FIFO queuing in routers. This is because if a receiver is experiencing loss, it would mean that other flows sharing the receiver's bottleneck link are also experiencing loss. Whereas in our case the loss is within some receiver tolerance, no such guarantees can be made for the other flows [2].

For this reason, we attempt to operate with little or no loss experienced by the multicast session's receivers. To achieve this, our protocol requires the source to transmit on another multicast group (the "base" or B-group) at the lowest possible data rate that is acceptable to the application. A receiver that experiences losses while receiving on the original group (called the "variable" or V-group), will first inform the source of its lower isolated rate with the goal of potentially influencing the source to reduce its rate. If losses continue despite this (typically because reducing the rate will reduce the inter-receiver fairness for this session), the receiver will leave the V-group and join the B-group. If after joining the B-group the receiver is still experiencing losses, then it is made to leave the session entirely. The protocol also provides mechanisms for receivers in the B-group to try to re-join the V-group.

This scheme is essentially a special case of a destination-set grouping (DSG) protocol as described in [6]. There are, however, many differences. First, our goal here is to maximize the inter-receiver fairness and to maintain inter-session fairness and TCP-friendliness. As a consequence, the information included in the feedback to the source is different, and different techniques are required for its aggregation and processing. Second, the use of destination-set splitting is limited to the provision of the low-rate B-group which avoids some of the problems that may be associated with the use of a large number of arbitrary-rate replicated streams [3].

3 Inter-Receiver Fairness: Definitions

The inter-receiver fairness (IRF) achieved by a multicast session is a function of the individual fairness values obtained by each receiver in the session. We develop the expression for the session's IRF in three steps: (1) we formally define the isolated rate of a receiver, as required to define the individual fairness; (2) we review the definition of individual fairness for a single group session [7] and extend that definition to two groups; (3) we combine the individual fairness values into the session IRF.

3.1 Isolated Rate

In previous work, we informally defined the isolated rate to be the rate that a receiver would obtain if unconstrained by the other receivers in the group [7], assuming max-min link sharing. That is, our isolated rate definition allowed for the receivers in a single session to receive at different rates. Recently, Rubenstein et al. have formalized this notion under the term "multi-rate max-min fair allocation [9]." They define an allocation to be multi-rate max-min fair if (1) it is feasible (i.e., no link is overloaded), (2) the multicast session is a *multi-rate* session and (3) no receiver can increase its allocated rate without causing a decrease in the rate to a more constrained receiver. The reader is referred to [9] for the precise definition. Hereafter, we use the term "isolated rate" of a receiver to refer to the receiver's rate under a multi-rate max-min fair allocation.

An example of multi-rate max-min fair allocation is illustrated in Fig. 1. This network has two connections, a multicast connection, A, with one source (AS) and two receivers (AR_0, AR_1), and a unicast connection, B, with source BS and receiver BR_0. S_0, S_1 and S_2 are routers. The constrained links are $L0$ and $L1$; each is labeled with its capacity. All other links are assumed to be unconstrained. Each receiver is labeled with its isolated rate, as determined by the multi-rate max-min fair allocation scheme. Both receivers AR_0 and BR_0 will get rate 2.5, and AR_1 will get 2.0, due to the existence of the bottleneck link L_1.

[1] Of course, the isolated rate is never more than the receiver's own maximum reception rate, typically governed by its processing capacity. We use the terms "fair share" and "isolated rate" interchangeably.

[2] This was not a problem in our original ATM-based development since we assumed some form of per-flow queuing in operation at the ATM switches.

[3] For a survey of replicated-stream, and also layering, mechanisms, please refer to [10].

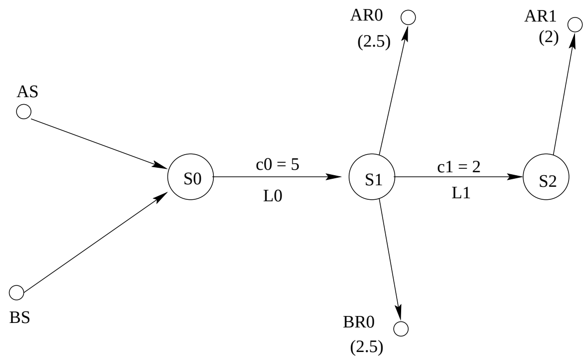

Figure 1: The Multi-rate Max-min Fair Allocation

The following definitions make use of the isolated rate for each receiver. The details of the algorithm for estimating the isolated rate in the Internet are described in section 4.2.1.

3.2 Individual Receiver Fairness

3.2.1 Single Group

Our starting point is the receiver fairness function that we previously defined, which assumes a single group session [7]. We summarize that definition here for completeness.

The *individual receiver fairness function* maps from the sending rate of the (single) source to a fairness value normalized to the range $(0.0 \ldots 1.0)$. The fairness value quantifies the notion that if the actual rate of the connection exceeds the isolated rate, then the receiver experiences performance degradation in the form of losses; if the actual rate is less than the isolated rate, then the receiver experiences performance degradation in the form of decreased throughput. In either case, there is a form of unfairness to the receiver.

For receiver i, the fairness function is denoted by $F_i(r)$, where r is the sending rate of the connection. Many different functions are possible for $F_i(r)$, depending on the application and the receiver characteristics. We assume that all fairness functions obey certain properties, including (1) $F_i(r_i) = 1.0$, where r_i is the *isolated rate* of receiver i and (2) $F_i(r)$ is non-decreasing in the range $[0, r_i)$ and non-increasing in the range $(r_i, \infty]$. In our simulation experiments, we use a particular fairness function of the form:

$$F_i(r) = \begin{cases} \frac{r}{r_i} & \text{if } r \leq r_i \\ \frac{r_i}{r} & \text{if } r > r_i \end{cases} \quad = \quad \frac{\min(r_i, r)}{\max(r_i, r)} \qquad (1)$$

A receiver is also allowed to specify the *maximum acceptable loss tolerance*, ℓ_i, indicating the maximum fraction of transmitted data that can acceptably be lost. For a single group session, the loss tolerance restricts the sending rate r to be at most $r_i/(1 - \ell_i)$.

3.2.2 Two Groups

Our scheme involves two groups per multicast session, the base (or B-) group and the variable (or V-) group. The rate of the B-group, r_b, is determined by the application and assumed to be fixed. The rate of the V-group, r_v, is selected to maximize the inter-receiver fairness and will vary over the lifetime of the session. At any point in time, each receiver is in the V-group, the B-group or no group at all. For receivers in the V-group, the fairness is $F_i(r_v)$; for receivers in the B-group, the fairness is $F_i(r_b)$; for receivers in no group, the fairness is 0.0.

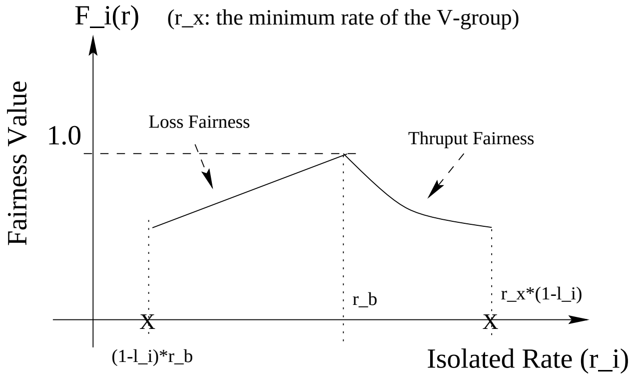

Figure 2: Fairness Under Moderate Isolated Rates

Our scheme includes a protocol to allow receivers to move between groups. Ideally, a receiver belongs to the group that maximizes its individual fairness while keeping losses below the loss tolerance. This criteria results in three cases, based on the relationship of the isolated rate r_i and loss tolerance l_i to the rates r_b and r_v:

- $r_i < (1 - l_i)r_b$

  In this case, the receiver's isolated rate is low enough that it will experience intolerable losses, even in the B-group. Such a receiver should leave the multicast session.

- $(1 - l_i)r_b \leq r_i < r_x(1 - l_i)$

  where r_x is the minimum rate of the V-group [4]. In this case, the receiver has a moderate isolated rate. The rate is sufficient to receive the B-group, but the rate is not high enough to receive the V-group without intolerable losses. Note that such a receiver may experience losses in the B-group, but the losses will be within the loss tolerance. Fig. 2 shows a graphical

[4] Selection of the minimum rate of the V-group has to ensure that benefits obtained from the two-group model can justify costs associated with it. Moreover, it should ensure that if a receiver has capability to stay in the V-group, that receiver should achieve more fairness than obtained in the B-group. Please see section 4.2.2 for details.

representation of this case. The x axis is the isolated rate of a receiver, and the y axis is the corresponding fairness value. In the figure, the straight line to the left indicates loss fairness and the curve to the right denotes throughput fairness experienced by a receiver.

- $r_x(1 - l_i) \leq r_i$

  In this case, the receiver has an isolated rate that is high enough to receive the V-group. The receiver may experience losses, but the losses will be within its loss tolerance. This case is illustrated in Fig. 3. The x axis is the sending rate of the V-group, and the y axis is the corresponding fairness value achieved by a receiver. This figure is different from Fig. 2 in that the straight line to the left indicates throughput fairness and the curve in the middle denotes loss fairness experienced by a receiver. The rightmost straight line shows throughput fairness experienced by a receiver if the receiver does not have capability to stay in the V-group and has to move to the B-group. In this scenario, the receiver achieves a constant fairness value.

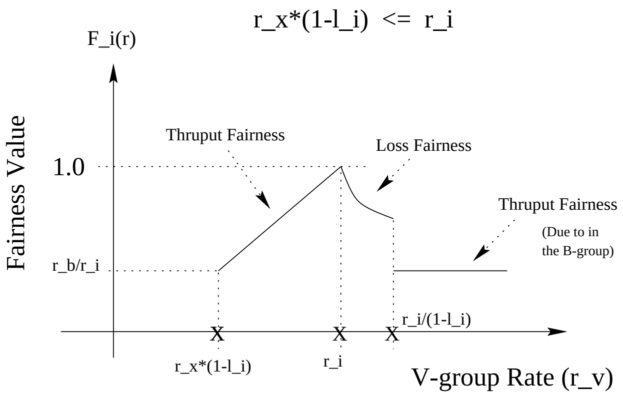

Figure 3: Fairness Under High Isolated Rates

3.3 Inter-Receiver Fairness Function in Two-group Model

In the two-group model, we define the *inter-receiver fairness* achieved by the multicast group to be the weighted sum of the individual fairness values of receivers, both in the B-group and in the V-group. That is,

$$IRF(r_b, r_v) = \sum_{i=1}^{n_b} \alpha_i F_i(r_b) + \sum_{i=n_b+1}^{n_b+n_v} \alpha_i F_i(r_v) \qquad (2)$$

$$subject\ \ to\ \ \textstyle\sum_{i=1}^{n_b+n_v} \alpha_i = 1.0,$$

$$and\ \ 0 \leq \alpha_i \leq 1.0,\ i = 1, \ldots, n_b + n_v$$

where n_b and n_v are the number of receivers in the B-group and in the V-group, respectively, $F_i()$ is the fairness functions for receiver i, and r_b and r_v are the transmission rates of the B-group and the V-group, respectively. The first term in equation (2) is the inter-receiver fairness value achieved by the receivers in the B-group and the second term is the fairness value achieved by the receivers in the V-group. Note that receivers that have left the multicast session do not figure into the IRF function. This definition allows receivers to contribute unequally to the total inter-receiver fairness, via unequal α_i's. These weights can be statically assigned or may vary over time.

4 Protocol Implementation

We describe the high-level operation of the two protocol components, followed by a discussion of four key details: estimating the isolated rate, determining the minimum rate of the V-group, approximating the IRF using samples of isolated rates, and efficiently selecting the V-group rate to maximize the IRF. These are novel issues (not addressed in [1] and [6]) which are necessitated by our goal of maximizing the inter-receiver fairness.

4.1 Protocol Components

Our protocol has two components:

1. The *intra-group* protocol used by the sender to gather and use isolated rate information from the receivers in each group. For the V-group, the isolated rates are used to adjust the group rate. For the B-group, the isolated rates are used to decide whether to invite any B-group receivers to move to the V-group. The intra-group protocol is a modified version of the probabilistic polling algorithm described in [1].

2. The *inter-group* protocol used by receivers, with assistance from the sender, to move between groups as their isolated rates change. This protocol is a simplified version of the stream change protocol proposed in [6].

4.1.1 Intra-group Protocol Overview

We now describe an algorithm that combines the scalable polling algorithm in [1] and the IRF mechanism that we proposed. This algorithm is run independently in the B-group and the V-group. The goal of the polling is to estimate the size of each group and the isolated rates for receivers in the group, so that the multicast source can adjust the V-group transmission rate. The polling algorithm consists of a series of rounds, termed an *epoch* [1].

Epochs, which are initiated periodically, consist of the following steps:

1. At the start of an epoch, both the multicast source and all the receivers generate random keys of length 16 bits.

2. When the source wishes to solicit responses from receivers, it sends out polling packets to all receivers in the group. Each polling packet contains the 16-bit random key, an indication of the number of significant bits to use in the key, and the type of the packet. There are two types of polling packets: size-solicitation packets and state-inquiry packets [1]. Size-solicitation packets are used to estimate the number of receivers in the group; state-inquiry packets are used to sample the estimated isolated rates of the receivers. Size-solicitation packets are sent out in rounds preceding state-inquiry packets.

   Initially in an epoch, all 16 bits are significant and the type of the polling packet is size-solicitation. If the multicast source does not get any reply in a round, it decreases the number of significant bits by one in the following round, thereby increasing the probability of a reply. The sender proceeds through the rounds until a "hit" is achieved, i.e., until at least one receiver replies. This allows the sender to estimate the receiver group size through a simple formula illustrated in [1] [5].

   State-inquiry packets are sent out to receivers in subsequent rounds. Responses to the state-inquiry packets from the multicast receivers indicate the estimated isolated rates of receivers,

3. When a polling packet arrives from the sender, each receiver compares the random key in the packet to its locally generated key, using the number of significant bits declared in the packet. If the relevant bits match, the receiver takes one of two different actions:

   If the packet is a size-solicitation packet, the receiver will send a response back to the sender so that the sender can estimate the receiver group size [1].

   If the packet is a state-inquiry packet and the group is large [6], only those receivers experiencing loss will reply. If the group is small, all receivers will reply. Each matching receiver responds with its isolated rate, estimated using the measured packet arrival rate and loss rate at the receiver. The isolated rate estimation is described in Section 4.2.1.

4. When the source gets both the size-solicitation and the state-inquiry responses from the receivers, it can use the difference in the index between rounds in which each hit occurs to estimate the percentage of the receivers that are experiencing loss [1] [7].

   If this probing procedure takes place in the B-group, the source will use the above information to determine whether the *inter-group* protocol should be called, inviting the receivers in the B-group with sufficient isolated rates to join the V-group.

   In the V-group, the sender uses the gathered information to decide a new rate. If no receivers are experiencing loss, the sender will increase its rate, using the loss tolerance to bound the increase. If some receivers are experiencing loss, the sender will call an optimization process to select a rate that maximizes the current inter-receiver fairness. Because the sender typically has information from just a subset of the receivers, the IRF function must be approximated. The approximation process is described in Section 4.2.3. Efficient optimization of the IRF function is described in Section 4.2.4.

4.1.2 Inter-group Protocol Overview

Each receiver continually monitors the reception rate and packet loss rate for its current group. There are three scenarios that cause a receiver to change to a different group.

1. If a receiver in the B-group experiences losses that exceed its loss tolerance, then the receiver leaves the multicast session.

2. If a receiver in the V-group experiences losses that exceed its loss tolerance, or the loss experienced by the receiver does not exceed its loss tolerance but persists for a while, then the receiver leaves the V-group and joins the lower rate B-group.

3. If a receiver in the B-group does not experience any loss, then it may join the higher V-group. This group advancement is by invitation from the source and is similar to the scheme used in [6].

[5] According to [1], if a hit of size-solicitation occurs in $round_1$, then the multicast group size n can be estimated as:

$$n \sim e^{16.25 - round_1/1.4}$$

After the estimation has been done, the group is defined to be "small" or "large".

[6] According to step 2, the size-solicitation round in which a hit is made should precede the state-inquiry round. This means that the multicast sender knows the (estimated) group size before it sends out the state-inquiry packets. Then it can inform receivers of the group size.

[7] If the "hit" rounds of size-solicitation and state-inquiry are $round_1$ and $round_2$, respectively, then the percentage experiencing loss is estimated to be

$$\frac{e^{16.25 - round_2/1.4}}{e^{16.25 - round_1/1.4}} = e^{(round_1 - round_2)/1.4}$$

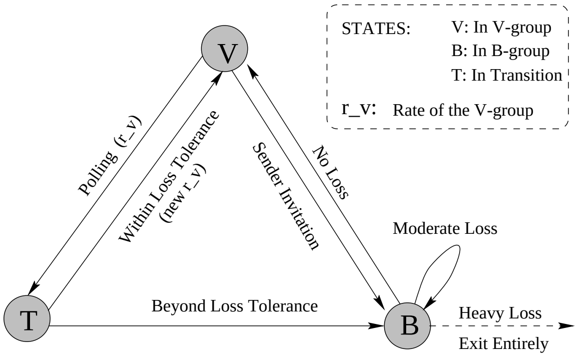

Figure 4: Receiver State-transit Diagram

4.1.3 Receiver State Diagram

In summary, the state-transit diagram of a receiver is described in Fig.4. There are three states: the V-group state (V), the B-group state (B), and the transition state (T). Each state transition is labeled with the reason for the transition. Initially, all receivers are in the V state.

If a receiver in the V state receives a polling packet from the multicast sender, the receiver enters into the T state by *probabilistically* sending a response to the sender (see section 4.1.1). This receiver in the T state continues to measure its experienced packet loss. If the measured loss exceeds the loss tolerance of the receiver, it voluntarily leaves the V-group and join the B-group by applying the inter-group protocol. This receiver's state also changes from the T state to the B state. On the contrary, if the measured loss does not exceed the loss tolerance of the receiver, the receiver continues to stay in the V-group and moves back to the V state.

If a receiver in the B state (i.e., in the B-group) does not experience any data loss, it may join the higher V-group upon receiving the group advancement invitation from the sender. While in opposition, if a receiver in the B state experiences moderate loss (less than its loss tolerance), it has to stay in the B-group even receiving an advancement invitation from the sender. Furthermore, B-state receivers experiencing heavy loss (more than their loss tolerance) should leave the multicast group entirely.

4.2 Implementation Details

Sections 4.1.1, 4.1.2 and 4.1.3 give a high-level description of the implementation of the Inter-Receiver Fairness protocol. In this section we cover four key details:

- Estimating the isolated rate
- Determining the minimum rate of the V-group
- Approximating the IRF using samples of isolated rates, and
- Efficiently selecting the V-group rate to maximize the IRF.

4.2.1 Estimation of the Receiver's Isolated rate

For the purposes of estimating isolated rate, the receivers are separated into two categories: those experiencing loss and those experiencing no loss. Let r_s denote the sending rate for the receiver's group (i.e., $r_s = r_b$ for the B-group and $r_s = r_v$ for the V-group).

If a receiver experiences loss it means there exists at least one link on the receiver's path from the sender which is fully utilized. The receiver could report the rate, $r_i = r_s * (1 - loss_i)$, as the isolated rate, where $loss_i$ is the measured loss percentage. However this behavior is sufficiently aggressive as to lead to unfriendliness when TCP connections are sharing the bottleneck link. Instead, we reduce the estimate of the isolated rate by a reduction factor, intended to be similar to TCP's rate reduction in the face of loss. In Section 5 we investigate the effect of using more conservative rate reduction on TCP sources.

If a receiver does not experience loss, then we know that the isolated rate is at least the current transmission rate, however, we don't know how much larger the isolated rate is. One possibility would be for the receiver to estimate the isolated rate by applying a relatively conservative increase function: a receiver that does not experience loss reports its estimated isolated rate $r_i = r_s/(1-l_i)$, where l_i is the loss tolerance for the receiver. The effect of this estimation is that even if the true isolated rate is exactly r_s, the sender can operate as high as the reported r_i without causing intolerable losses. This approach, however, is not scalable to large group sizes, because it requires every no-loss experiencing receiver to estimate its isolated rate based on its individual loss tolerance. Therefore, we propose an approach whereby the sender estimates the representative isolated rate for receivers that are not experiencing loss by using the current source sending rate and a group loss tolerance. The group loss tolerance is determined by the application; individual receivers may specify loss tolerances that are larger. Specifically, if r_s is the sending rate and lt is the loss tolerance for the group, then the isolated rate is estimated to be $r_s/(1 - lt)$.

In summary, the estimation of the isolated rate of a receiver is:

$$r_i = \begin{cases} r_s * (1 - loss_i) & \text{if experiencing loss} \\ r_s/(1 - lt) & \text{if experiencing no-loss} \end{cases}$$

where r_s is the transmission rate of the receiver's group, $loss_i$ is the measured loss percentage at receiver i, and lt is the loss tolerance for the group.

4.2.2 Determination of the Minimum Rate of the V-group

In the inter-receiver fair multicast protocol, the transmission rate of the V-group should be larger than a minimum value to compensate for the implementation costs associated with the two-group model. However, how to determine the minimum rate is not trivial. In this paper, we use the loss tolerance of the application to select the minimum rate.

Suppose at some time, the sending rate of the B-group in a multicast application is *fixed* at r_b and that of the V-group is r_v. Moreover, assume that the application loss tolerance is lt. We observe that the criterion for selecting the minimum rate of the V-group should guarantee: when the V-group sends at a rate higher than its minimum rate, if a receiver has capability to stay in the V-group, it should achieve more fairness value than if the receiver stays in the B-group. Based on this observation, we select the minimum rate of the V-group to be $r_v^{bot} = r_b/(1-lt)^2$.

Let $r_x = r_b/(1-lt)$. This selection can be explained in two cases:

1. At one time the V-group is sending at the rate r_v^{bot} and the estimated isolated rate r_i of receiver i is greater than r_x. This means that receiver i has capability to stay in the V-group. According to the particular single receiver fairness function defined in equation (1), if the receiver chooses to stay in the B-group, it obtains fairness value $\alpha_i * r_b/r_i$ (α_i is the IRF weight of receiver i), which is less than $\alpha_i *$ (1-lt) [8]; however, if the receiver stays in the V-group, it achieves a fairness value greater than $\alpha_i *$ (1-lt).

2. At other time the V-group is sending at a rate higher than r_v^{bot} and receiver i has capability to stay in the V-group. Obviously, the receiver achieves more fairness value when staying in the V-group than that in the B-group.

This rate selection provides an efficient way for a V-group receiver to determine its group membership: through measuring the loss percentage, a receiver chooses to stay in the V-group if the measured loss is less than its loss tolerance or join the B-group otherwise.

Moreover, we should note that this rate selection is dependent on the specific single receiver fairness function used in this paper. Different minimum rate may be selected if a different function is chosen.

4.2.3 Approximation of the Inter-Receiver Fairness Function

The core of the inter-receiver fairness optimization is the determination of the isolated rates of receivers. Ideally, if the source could obtain the (estimated) isolated rate of every receiver, then equation (2) could be used directly to maximize the IRF measure. However, to avoid feedback implosion the receivers are probabilistically selected to report their isolated rates. Thus the sender must approximate the IRF function using samples of isolated rates.

In our algorithm, the difference in index between the size-solicitation hit and the state-solicitation hit can be used to roughly determine what percentage of the multicast group members are experiencing loss; all other receivers are assumed to experience no loss [1].

The sender receives feedback from some of the receivers who are experiencing loss. The sender uses the estimated isolated rates reported by the *selected* receivers as representatives of the isolated rates for all receivers experiencing loss. Specifically, if k responses are obtained at the sender and the number of receivers experiencing loss is estimated to be m, then each response is presumed to represent the isolated rate for m/k receivers.

For example, consider a source with sending rate r_s. Within a polling epoch, the source is able to estimate the total group size, n, and the number of receivers experiencing loss, m. Suppose that during the state-solicitation process, three isolated rates, r_1, r_2 and r_3, are reported to the source. Then $m/3$ receivers are assumed to have isolated rate r_i for each $i = 1, 2, 3$. The remaining (n-m) receivers are assumed to have isolated rate $r_{nls} = r_s/(1-lt)$. Using the IRF function described earlier, the source approximates the IRF for the current receivers to be:

$$approxIRF(r_s) = \sum_{i=1}^{3} \left(\frac{m}{3}\right) \frac{\min(r_i, r_s)}{\max(r_i, r_s)} + (n-m) * \frac{\min(r_{nls}, r_s)}{\max(r_{nls}, r_s)} \tag{3}$$

The next section explains how to efficiently determine the source rate r_{opt} that maximizes the approximate IRF function [9].

We note that the *weight* of an individual receiver can be embedded in the selection probability: a receiver with a higher weight can be allowed a higher probability of matching the 16-bit source random number. For example, suppose the weight of receiver A is twice as much as that of B. Then A may generate twice as many random numbers as B in attempting to match the source random number, doubling the probability that the source hears from A.

4.2.4 Efficient IRF Optimization at Source

We have previously proven [7] that the IRF function can be efficiently optimized if individual fairness function meets several conditions. Specifically, the necessary conditions are:

1. only one point of discontinuity at $r = r_i$
2. second derivative ≥ 0 in the region $r < r_i$
3. second derivative ≥ 0 in the region $r > r_i$

[8] This can be derived through simple computation:

$$\alpha_i * \frac{r_b}{r_i} < \alpha_i * \frac{r_b}{r_x} = \alpha_i * \frac{r_b}{r_b/(1-lt)} = \alpha_i * (1-lt)$$

[9] Accuracy of the approximate algorithm has been examined in our previous work. The reader is referred to [7] for details.

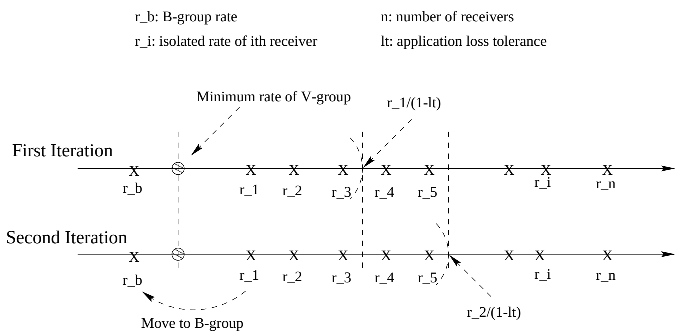

Figure 5: Two-group model optimization

where r is the source rate and r_i is the isolated rate of receiver i.

Basically, the proof shows that, given the isolated rates of receivers, the optimal value of the source rate must be one of them. Thus the source need only evaluate the IRF function for each r_i and select the best rate.

In the two-group model, the rate selected for the V-group should optimize the IRF function, with the additional complexity that receivers with low isolated rates may be pushed down to the B-group. Because of this complexity, multiple iterations are required to optimize the IRF. In iteration i, the $(i-1)$-st lowest rate receivers in the V-group are assumed to be moved to the B-group, and the IRF is optimized using the remaining receivers.

Fig.5 illustrates the optimization algorithm. This figure shows the first two iterations of the algorithm. r_b is the B-group sending rate, n is the number of receivers in the V-group and lt is the application-specific loss tolerance. $r_1, r_2, \ldots, r_n$ are the n isolated rates of the V-group receivers; they are all greater than the minimum rate of the V-group. Without loss of generality, we assume they are in non-decreasing order. Let $r_1^{lt}{=}r_1/(1-lt)$. Note that the optimal rate selected in the first iteration cannot exceed r_1^{lt}, otherwise, receiver 1 will experience intolerable loss. Thus, r_1^{lt} is the upper bound of the rate selection range for this iteration. In this example, the optimization algorithm can only select a rate from r_1, r_2, or r_3. Assume in the first iteration, the IRF protocol picks r_2 as the best rate, assuming all receivers remain in the V-group.

In the second iteration, receiver 1 is assumed to be moved to the B-group. Thus, the optimization algorithm need only evaluate the IRF function at the remaining $n-1$ isolated rates, $r_2, \ldots, r_n$, and select the best one. Now receiver 2 has the minimum isolated rate. Let $r_2^{lt}{=}r_2/(1-lt)$. Then r_2^{lt} is the upper bound of the rate selection range for this iteration, and the algorithm need only consider $r_2, \ldots, r_5$. Suppose in this iteration rate r_5 optimizes the IRF function and is selected as the best rate. Then, the inter-receiver fairness achieved under rate r_5 in the second iteration is compared with that obtained under rate r_2 in the first iteration. The better one determines the *current* optimal rate after two iterations. This optimization continues until no better inter-receiver fairness can be achieved in an iteration when compared to that in the previous iteration.

In fact, we can further reduce the complexity of the optimization algorithm by avoiding redundant computation in different iterations. For instance, in Fig.5, neither r_2 nor r_3 should be considered for the optimization in the second iteration, because they have already been examined in the first iteration. The total inter-receiver fairness achieved by the n receivers can only decrease if either r_2 or r_3 is selected, when compared with that achieved in the first iteration. Hence we need only consider rates r_4 or r_5 in the second iteration, i.e., the rates falling into the range $[r_1^{lt}, r_2^{lt}]$.

5 Simulations

In this section, we examine the performance of our scheme under a variety of network topologies and group characteristics. We implemented our protocol in the LBNL network simulator *ns-2* [11].

In the simulation, the *inter-group* protocol is called every 20 seconds. The sender invites the B-group receivers to join the V-group whenever the number of B-group receivers exceeds 10% of the total and at least one receiver in the B-group is not experiencing loss. Moreover, a receiver that leaves the V-group must remain in the B-group for at least 20 seconds before re-joining the V-group. This is used to prevent the oscillation of a receiver between the V-group and the B-group. In the intra-group protocol, the sending frequency of the polling packets is constant and set at five seconds [10].

Configurations in the simulations have the following properties:

- All connections are uni-directional. No traffic flows from the receiver(s) to the sender, except the control packets. The type of the TCP connections used in the simulation is "Tahoe" TCP.
- The type of the routers is "DropTail".
- Unless otherwise stated, the link bandwidth is 1.0 Mbps.
- All receivers in the same multicast session have the same loss tolerance, set at 10% [11]. Receivers in the same multicast session are assigned the same IRF weights.

[10] We select this time so that each epoch has sufficient time to complete before the next epoch begins.

[11] We select 10% as the loss tolerance to reflect the ability of certain real time applications which can tolerate a limited amount of random loss.

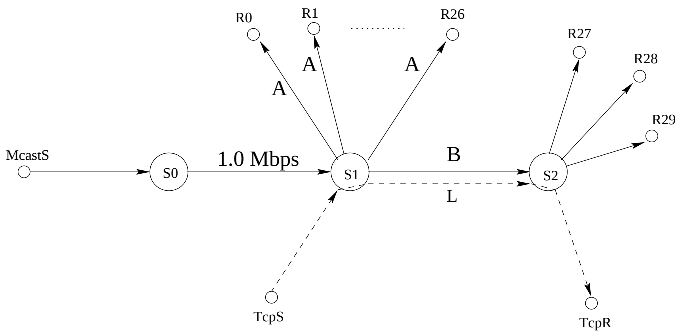

Figure 6: Simulation Topology One

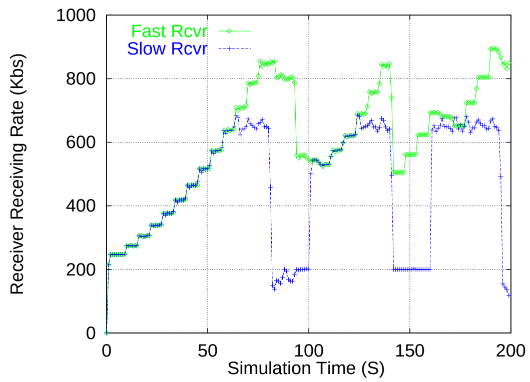

Figure 7: Topology One: B=700 Kbps

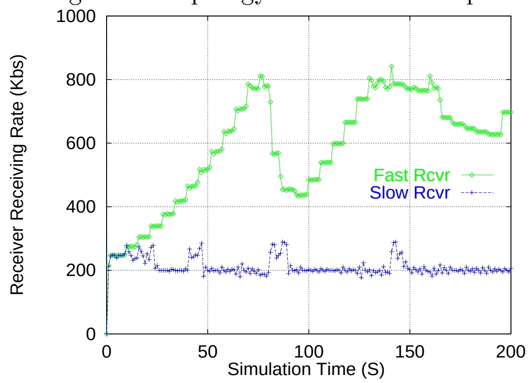

Figure 8: Topology One: B=300 Kbps

- The rate reduction factor is applied in all simulations except when we examine the effects of the rate reduction factor itself.
- Initially, all receivers are in the V-group. The B-group sending rate is 200 Kbps.
- In all simulation graphs, the x axis shows the simulation time and the y axis shows the receiver's reception rate.

5.1 Single Multicast Group

We start by examining the performance of a single IRF multicast group, without any other connections. We use the topology shown in Fig. 6, which has 30 receivers, 27 connected to router S_1 and 3 connected to router S_2. Each link is labeled with its bandwidth capacity. A is 900 Kbps. Note, in this scenario, the TCP connection (denoted as dashed line in the figure) does not exist.

Fig. 7 and 8 show the receiving rates with bandwidth B at 700 Kbps and 300 Kbps, respectively. Receivers 0–26 are *fast* receivers because they have the isolated rates 900 Kbps, while receivers 27–29 are *slow* receivers with isolated rates being B (either 700 Kbps or 300 Kbps).

Consider Fig. 7. At the beginning, because the sender's rate is low, the slow receivers can remain in the V-group. However, when the sending rate of the multicast session approaches the isolated rates of the slow receivers (around 60 seconds), the inter-receiver fairness optimization determines that it would be better if the slow receivers move to the B-group. Then, roughly at 78 second, the slow receivers move to the B-group. Now, the slow receivers receive at the B-group rate 200 Kbps and the fast ones get the leftover on link S_0–S_1, which is 800 Kbps. At 100 seconds, the receivers in the B-group receive an invitation from the sender and they return to the V-group. Later, when the slow receivers can't tolerate the loss, they move back to the B-group again.

We notice that, in Fig. 7, the rate of the fast receiver (i.e., the V-group rate) gets dropped at 90 seconds. This is because, at that time, with the slow receivers in the B-group receiving at rate 200 Kbps and the V-group sending rate at around 800 Kbps, the fast receivers do not experience any loss. Thus they try to increase their bandwidth allocation. However, this increment causes loss to those fast receivers and accordingly they will backoff. The B-group is sending at constant rate and such loss won't affect it. The same thing happens at 140 seconds again. The inter-receiver fairness of the whole multicast session improves over a single session at rate 700 Kbps. However the improvement is modest since the isolated rates are fairly similar.

If the heterogeneity increases, the IRF may be greatly improved. In Fig. 8, B is 300 Kbps; thus the isolated rates of slow receivers are 300 Kbps while those of the fast ones are still 900 Kbps. The slow receivers are quickly unable to stay in the V-group. Each time these receivers attempt to move to the V-group (e.g., at 40, 80 and 140 seconds), the losses are significant, and the receivers move back to the B-group. In this case, the inter-receiver fairness of the whole multicast session improves significantly, about 117% [12].

[12] In this scenario, the isolated rates of receivers 0-26 are 900 Kbps and receivers 27-29 are 300 Kbps. The inter-receiver fairness when

5.2 Interaction Among Multiple IRF Multicast Connections

Next we examine the interaction between different multicast connections when our scheme is used. Fig.9 shows a topology with two multicast connections, A and B. Both A and B have one sender and ten receivers. These two connections share one common bottleneck link whose bandwidth capacity is 1.0 Mbps.

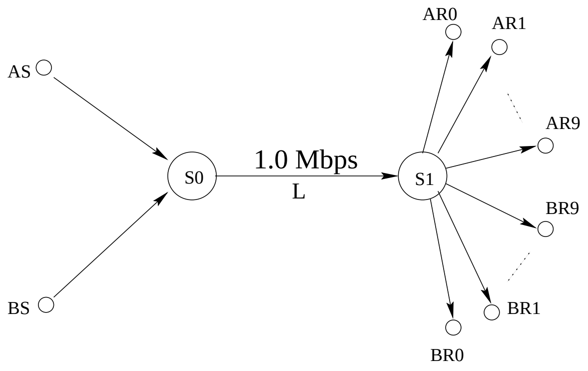

Figure 9: Simulation Topology Two

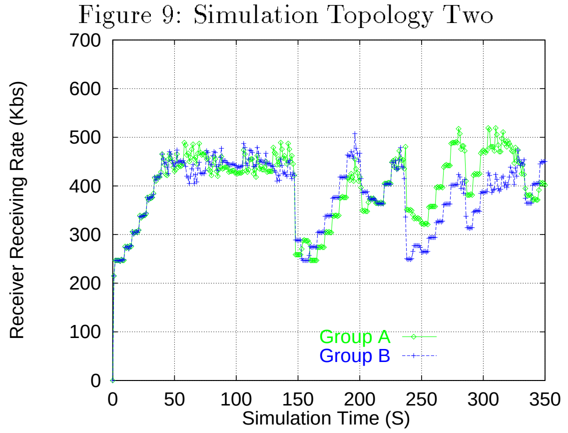

Figure 10: Topology Two: Sharing Same Bottleneck Link

The simulation result is shown in Fig. 10. From the figure, we see that both sessions have similar behavior and they roughly fairly share the bottleneck bandwidth. Initially both groups increase their sending rates without experiencing any loss, until the point (50 seconds) at which the link is saturated. After that, their respective rates oscillate for a while and then the relatively high losses force each group to backoff its sending rate (at 150 seconds). The backoff clears the bottleneck link L and eliminates loss, so the sending rates of both groups increase again. The average throughput of both groups over the whole simulation period is quite similar; A is 399.7 Kbps and B is 392.8 Kbps [13].

not using our protocol is 0.4 and when using our protocol is 0.867. The improvement is around 117%.

We also investigate the interaction between inter-receiver fair multicast when they share *different* bottleneck links. Fig.11 shows the third simulation configuration. A and B are two multicast connections, each with ten receivers. The bottleneck link of the B group is the link L_{12}, whose available bandwidth is 0.4 Mbps. The bottleneck link of the A group is the link L_{01}. Thus, the isolated rates of receivers in the B group are 0.4 Mbps and those in the A group are 0.6 Mbps.

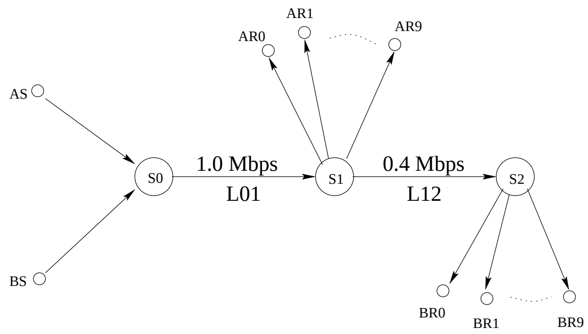

Figure 11: Simulation Topology Three

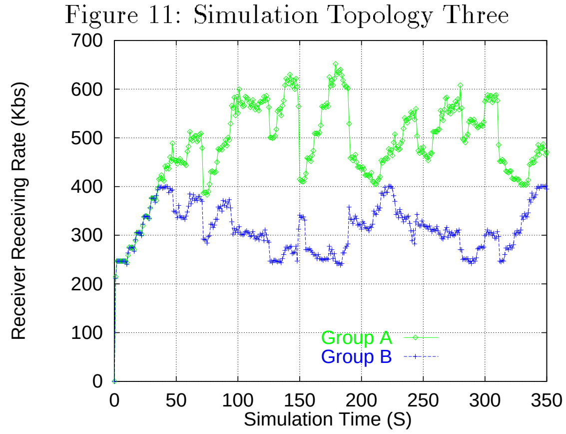

Figure 12: Topology Three: Sharing Different Bottleneck Links

Fig.12 shows the simulation results. We can see that both A and B vary approximately around the rates commensurate with their own capabilities: A varies in between $500 \sim 600$ Kbps and B varies between $300 \sim 400$ Kbps. The average throughput of group A over the whole simulation period is 503.8 Kbps and that of group B is 311.5 Kbps.

[13] The reason the sum of the average throughput of A and B groups is less than 1.0 Mbps is due to the effect of the rate reduction on loss detection and the relatively long polling epoch in the *intra-group* protocol for recovery.

5.3 Multicast Sharing With TCP

When we deploy the IRF multicast protocol in the Internet, we must consider the interaction with TCP. We use the topology shown in Fig.6 to examine the behavior of the IRF protocol with TCP connections. There are two connections, one multicast connection with 30 receivers ($R_0 \ldots R_{29}$) and one TCP connection (from node $TcpS$ to $TcpR$). They share one common link, whose bandwidth capacity B is 700 Kbps. A is 1.0 Mbps.

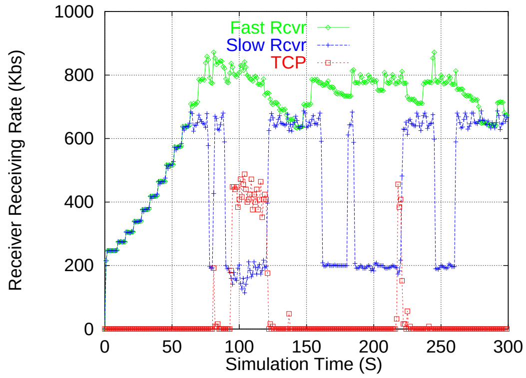

Figure 13: Sharing With TCP: Without Rate Reduction (B=700 Kbps)

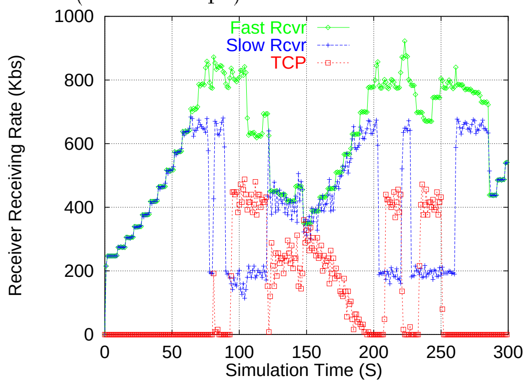

Figure 14: Sharing With TCP: With Rate Reduction (B=700 Kbps)

Fig. 13 and 14 show the simulation results without and with the rate reduction factor, respectively. The simulation period is 300 seconds and the TCP connection starts at 80 seconds and leaves at 250 seconds. As can be seen in Fig. 13, if we do not use the rate reduction factor, the TCP connection is unfairly treated and receives very little bandwidth. The average throughput of the TCP connection over its active period ([80s, 250s]) is only 77 Kbps. In Fig. 14, we see that with rate reduction, the IRF multicast can fairly share bandwidth with the TCP connection. The TCP average throughput over its active period is now improved to 220 Kbps [14].

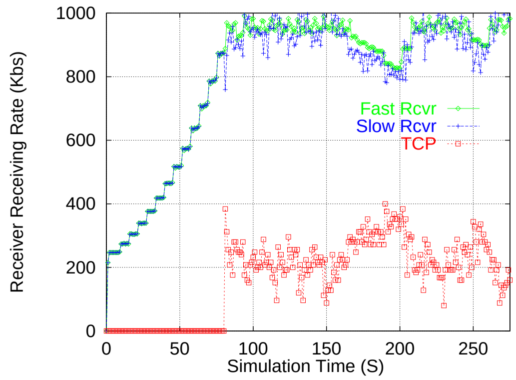

Figure 15: Sharing With TCP: Without Rate Reduction (B=1.2Mbps)

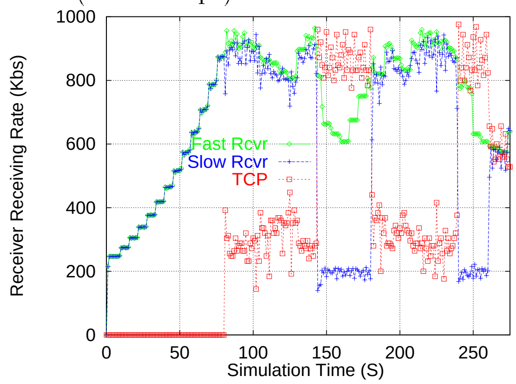

Figure 16: Sharing With TCP: With Rate Reduction (B=1.2Mbps)

To get a better understanding of the behavior of the reduction factor, we re-run the simulation by increasing the bandwidth B from 700 Kbps to 1.2 Mbps (see Fig. 15 and 16). The simulation period is 275 seconds and the TCP connection starts at 80 seconds and does not leave in the middle. In this scenario, there is extra bandwidth on the shared link and TCP can obtain it. From the Fig. 15, we can see that if we do not use the reduction factor, TCP is unfairly treated and receives only the leftover bandwidth on link L. In this case, the average throughput of the TCP connection is only 233 Kbps. In Fig. 16, with the usage of the reduction factor TCP competes fairly for the bandwidth with the IRF multicast: TCP can force the competing receivers (R_{27}, R_{28}, R_{29}) to move to the B-group. In this example, TCP achieves average throughput 488 Kbps.

[14] The reason the TCP connection does not get average bandwidth closer to 350 Mbps is because our IRF protocol is not as sensitive to the measured loss as TCP is. Our IRF multicast may ignore a small measured loss for a while.

6 Conclusions and Future Work

The issue of controlling sending rates for a multicast session over the Internet is especially challenging because of the need to consider the heterogeneity of the session's receivers and network paths. In our work we use an inter-receiver fairness measure to evaluate particular rate control mechanisms. Our focus is on the feasibility of developing protocols that can be used to maximize inter-receiver fairness.

To this end, this paper develops a protocol for the rate control of a multicast session operating over the Internet in such a way that the session's inter-receiver fairness measure is maximized. Additional goals for the rate control protocol are that it be TCP-friendly and that it exhibits good sharing with similarly-controlled multicast sessions. Because of our concerns with the effect of heavy losses on the sharing properties of the protocol, we propose the use of a low-rate B-group that can accommodate low-rate receivers. In effect, we develop a special case of the DSG protocol [6] that differs from the original work in its rate setting criteria, namely maximization of IRF, and its fair sharing goals.

Although in this paper we strive to provide fairness between multicast sessions and TCP sessions, there is no consensus on the definition of the fairness, let alone bandwidth allocation mechanisms between multicast sessions and TCP sessions. Several recent papers have examined this fairness issue [12, 13, 14]. Our ongoing work is to investigate this issue more substantially and endeavor to apply the quantitative fairness measure in this paper to multicast congestion control. Moreover, we will extend the specific two-group model to a more generalized multi-group model in the future.

References

[1] J. Bolot, T. Turletti, and I. Wakeman, "Scalable Feedback Control for Multicast Video Distribution in the Internet," in *Proceedings of SIGCOMM '94*, 1994.

[2] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast," in *Proceedings of ACM Sigcomm '96, Stanford, CA*, September, 1996.

[3] X. Li, S. Paul, P. Pancha, and M. Ammar, "Layered video Multicast with Retransmission (LVMR): Evaluation of Error Recovery Schemes," in *Proceedings of Nossdav '97, St Louis, MI*, May, 1997.

[4] X. Li, S. Paul, and M. Ammar, "Layered Video Multicast with Retransmission (LVMR): Hierarchical Rate Control Schemes," in *Proceedings of Infocomm '98, San Francisco, CA*, March, 1998.

[5] S. Cheung and M. Ammar, "Using Destination Set Grouping to Improve the Performance of Window-Controlled Multipoint Connections," *Computer Communications*, vol. 19, pp. 723–736, 1996.

[6] S. Cheung, M. Ammar, and X. Li, "On the Use of Destination Set Grouping to Improve Fairness in Multicast Video Distribution," in *Proceedings of Infocomm '96, San Francisco, CA*, March, 1996.

[7] T. Jiang, M. Ammar, and E. Zegura, "Inter-receiver Fairness: A Novel Performance Measure for Multicast ABR Sessions," in *Proceedings of ACM Sigmetrics '98, Madison, Wisconsin*, June, 1998.

[8] S. Floyd and K. Fall, "Promoting the Use of End-to-End Congestion Control in the Internet," *Submitted to IEEE/ACM Transaction of Networking*, February 1998.

[9] D. Rubenstein, J. Kurose, and D. Towsley, "The Impact of Multicast Layering on Network Fairness," in *Proceedings of Sigcomm'99*, Sept. 1999.

[10] X. Li, M. Ammar, and S. Paul, "Video Multicast over the Internet," *IEEE Network Magazine (To appear)*, 1999.

[11] S. McCanne and S. Floyd, "The LBNL *network simulator*," *Lawrence Berkeley Laboratory*, Software on-line: http://www-nrg.ee.lbl.gov/ns.

[12] H. Wang and M. Schwartz, "Achieving Bounded Fairness for Multicast TCP Traffic in the Internet," in *Proceedings of SIGCOMM'98*, 1998.

[13] A. Legout, J. Nonnenmacher, and E. Biersack, "Bandwidth Allocation Policies for Unicast and Multicast Flows," in *Proceedings of INFOCOMM'99*, March, 1999.

[14] X. Li, S. Paul, and M. Ammar, "Multi-Session Rate Control for Layered Video Multicast," in *Proceedings of Multimedia Computing and Networking'99, San Jose, CA*, Jan, 1999.