End-to-End Rate Control in ABR*

Dorgham Sisalem
Henning Schulzrinne
GMD-Fokus Berlin
{dor, hgs}@fokus.gmd.de

January 30, 1996

*This work was funded in part by the Commission of European Communities (CEC) under project R2116 TOMQAT.

Abstract

We compare between the binary congestion indication scheme and the explicit rate indication scheme proposed for congestion control of ATM's available bit rate service (ABR).

The fairness problems of the binary congestion indication scheme that arise when mixing long and short distance traffic are investigated. Our simulations suggest that the unfairness noticed in this case towards the long distance traffic mainly results from the difference in the number of the traversed hops and not from the difference in propagation delay. We also introduce a simple approach with which this unfairness can be avoided. This approach is based on setting the multiple decrease factor in an inverse relation to the number of the traversed hops.

1 Introduction

Ensuring both high utilization of the available bandwidth and a minimum cell loss rate requires efficient as well as fair congestion control schemes. These schemes should achieve three goals [1]:

1. At a minimum, the quality of service (QOS) negotiated during the connection establishment has to be satisfied for each source.
2. Unused bandwidth should be fairly distributed among the active connections.
3. During congestion the connections that are using more bandwidth than was negotiated at the connection establishment should be given the opportunity to reduce their rate before the network starts discarding traffic in excess of the negotiated QoS.

Following the work done by the traffic management subworking group (TMG) of the ATM Forum two main approaches can be distinguished for congestion control in ATM:

1. Link-by-link credit based flow control
2. End-to-end rate control

As the TMG voted for a rate control based mechanism in its September 1994 meeting, we will only describe the rate control approach.

In this scheme congestion information is collected along the VC path and is conveyed back to the source. Based on this information the source can determine the sending rate with which congestion can be avoided. In order to provide a closed end-to-end control loop, the congestion information is not sent directly to the source but is forwarded to the destination end system. This forward error congestion notification (FECN) provides the destination with a complete picture of the congestion state along the connection's path. Also, it minimizes the needed number of congestion messages compared to the backward error congestion notification (BECN) where each intermediate switch sends a congestion notification to the source when congestion is observed.

Rate information can be conveyed back to the source in one of two forms:

- Binary congestion notification using a single bit for marking the states "network congested" and "not congested".
- Explicit rate indication, with which the network notifies the source of the bandwidth share it should be using in order to avoid congestion.

Here, two algorithms are introduced. They differ in their performance, complexity and the kind of congestion information conveyed back to the source. The first one, presented in section 3, uses binary congestion notification. In section 4 we investigate the enhanced proportional rate control algorithm which uses explicit rate indication for congestion control.

In the second part of this study, see Sec. 5, we take a closer look at some of the fairness problems noticed when using binary congestion indication schemes. In this context the effects of the value of the multiplicative decrease factor in determining the throughput and fairness performance are investigated.

2 Simulation Environment

To compare the different algorithms a simple generic fairness configuration [2] was chosen. With this topology, see Fig. 1, the fairness of bandwidth allocation can be tested for the case of cascaded congested links. For a quantitative description of fairness Jain [3] suggests using the so called fairness index.

$$F = \frac{(\sum_i x_i)^2}{n \sum_i x_i^2}$$

with x_i as the ratio of the actual throughput to the fair throughput and n as the number of connections.

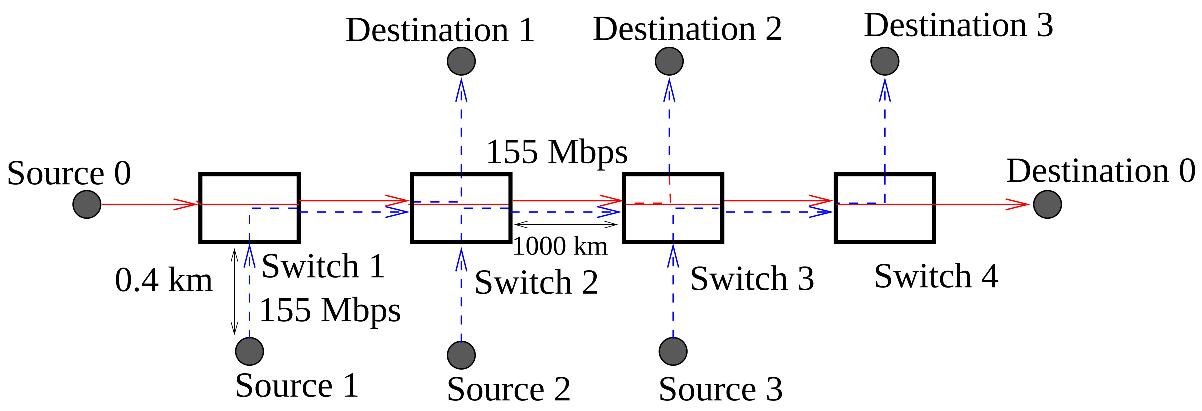

Figure 1: The generic fairness configuration

The test network consists of four switches, with each adjacent two switches connected through a 1000 km long WAN link. There are four greedy sources that send during the entire simulation with the maximum allowed rate, and four destination end systems. Except for the connection from source 0 to destination 0 that represent the transit traffic in the scheme all other connections are local traffic and pass only through one switch.

Each link is shared between two connections with the same increase and reduction factors, the same MCR and the same initial cell rate; they differ only in their round trip times. Therefore, we would expect the fair bandwidth share of each source to be half of the available link bandwidth, i.e., 77.5 Mbps.

3 Proportional Rate Control Algorithm

The proportional rate control algorithm (PRCA) is based on the positive feedback rate control scheme [4]. In this scheme the rate can only be increased if a positive indication, in this case a RM cell, was received. Otherwise, the rate is decreased after each sent cell.

Rate reductions and increases are done in proportion to the current sending rate thereby enhancing the fairness of the rate control scheme.

Parameter	value
Initial cell rate (ICR)	7.75 Mbps
Minimum cell rate (MCR)	0.155 Mbps
Peak cell rate (PCR)	155 Mbps
Additive increase rate (AIR)	1250
Link bandwidth	155 Mbps
Link length	1000 km
Source link	0.4 km
Propagation delay	5 μsec/km
Congestion threshold	50 cells

Table 1: Simulation parameters for testing the rate control algorithms

1. **Source End System Behavior**
   The source starts sending cells at an initial cell rate (ICR) with the EFCI state set to congested (EFCI=1). In the first cell and every other *Nrm** cells EFCI is set to 0. Each cell with (EFCI=0) represents an opportunity for the source to increase its rate. If these cells faced no congested links along their entire path a RM cell will be generated at the destination end system. On receiving these RM cells the source is entitled to increase its rate. This is done in such a manner as to compensate for all the decreases taken over the last *Nrm* cells and to achieve the desired overall rate increase. After the sending of each cell the rate is multiplicatively decreased by the multiplicative decrease factor (MDF).

2. **Switch Behavior**
   A congested switch can indicate congestion by:

   - setting the EFCI state in the passing cells to congested (EFCI=1),
   - removing RM cells sent by the destination end system to the source.

3. **Destination End System Behavior**
   If a cell is received with (EFCI=0) and no internal congestion is noticed the destination end system sends a RM cell to the source.

3.1 Simulation Results

To simulate the behavior of the PRCA proposal, the generic fairness topology presented in Fig. 1 was used with the parameters of Tab. 1, a multiplicative decrease factor (MDF†) of 8 and with *Nrm* set to 32.

* *Nrm* was set to 32 as was done in various contributions to the ATM Forum, see [5,6]

† Rate reduction in this algorithm is done in the form of

$$\mathrm{ACR} = \mathrm{ACR} - \frac{\mathrm{ACR}}{2^{\mathrm{MDF}}}$$

with ACR as the allowed cell rate of the source.

Using a congestion threshold of only 50 cells resulted here in a low utilization of only 53%. Due to the severe and long congestion periods noticed in this case both transit and local sources received only a small fraction of their fair bandwidth share and there was hardly no difference in the bandwidth distribution.

Congestion threshold	fairness index	utilization
50	0.95	52%
500	0.91	57%
1500	0.83	74%
3400	0.81	94%

Table 2: Fairness index and link utilization results achived with PRCA for different congestion thresholds

The high oscillations of the throughput of the PRCA proposal and the low utilization caused thereby were mainly forced by the small congestion threshold. Using a congestion threshold of 3400 cells we found that the oscillations were considerably reduced, see Fig. 2, and the utilization reached 94.5% of the available bandwidth. The bandwidth was, however, unfairly distributed. While the local traffic consumed about 76.8% of the utilized bandwidth the transit traffic bandwidth share decreased to 23.2%. This resulted in a fairness index of only 0.817.

4 Enhanced Proportional Rate Control Algorithm

As an example for an algorithm using explicit rate indication we briefly introduce in this section the enhanced proportional rate control algorithm (EPRCA).

This algorithm is based on determining the advertised rate (A-rate) [7]. This rate is computed as the capacity available (B) minus the capacity of the constrained VCs divided by the total number of VCs (N) minus the number of constrained VCs. A VC is considered to be constrained if the rate it is currently using is smaller than the fair bandwidth B_{fair} share calculated at this switch.

$$B_{\text{fair}} = \frac{B - \sum \text{constrained connections}}{N - \sum N \text{ of constrained connections}}$$

Even though the A-rate approach is accurate and fair, it is computation intensive and complex. Therefore, EPRCA estimates the value of the A-rate by computing a mean allowed cell rate (MACR) using exponential weighted averaging.

$$\text{MACR} = (1-\alpha)\text{MACR} + \alpha\text{ACR}$$

Here, α is the exponential averaging factor and is usually set to 1/6 [8]. ACR is the currently allowed cell rate of the connection.

Using the EPRCA scheme for implementing the explicit rate scheme must be seen only as one possible solution that has proven to work fine and is often used in simulation studies [6]. As the switch specifics will not be standardized by the ATM Forum the actual algorithm for calculating the fair bandwidth share can take any other shape that fulfills the fairness requirements, see [9,10].

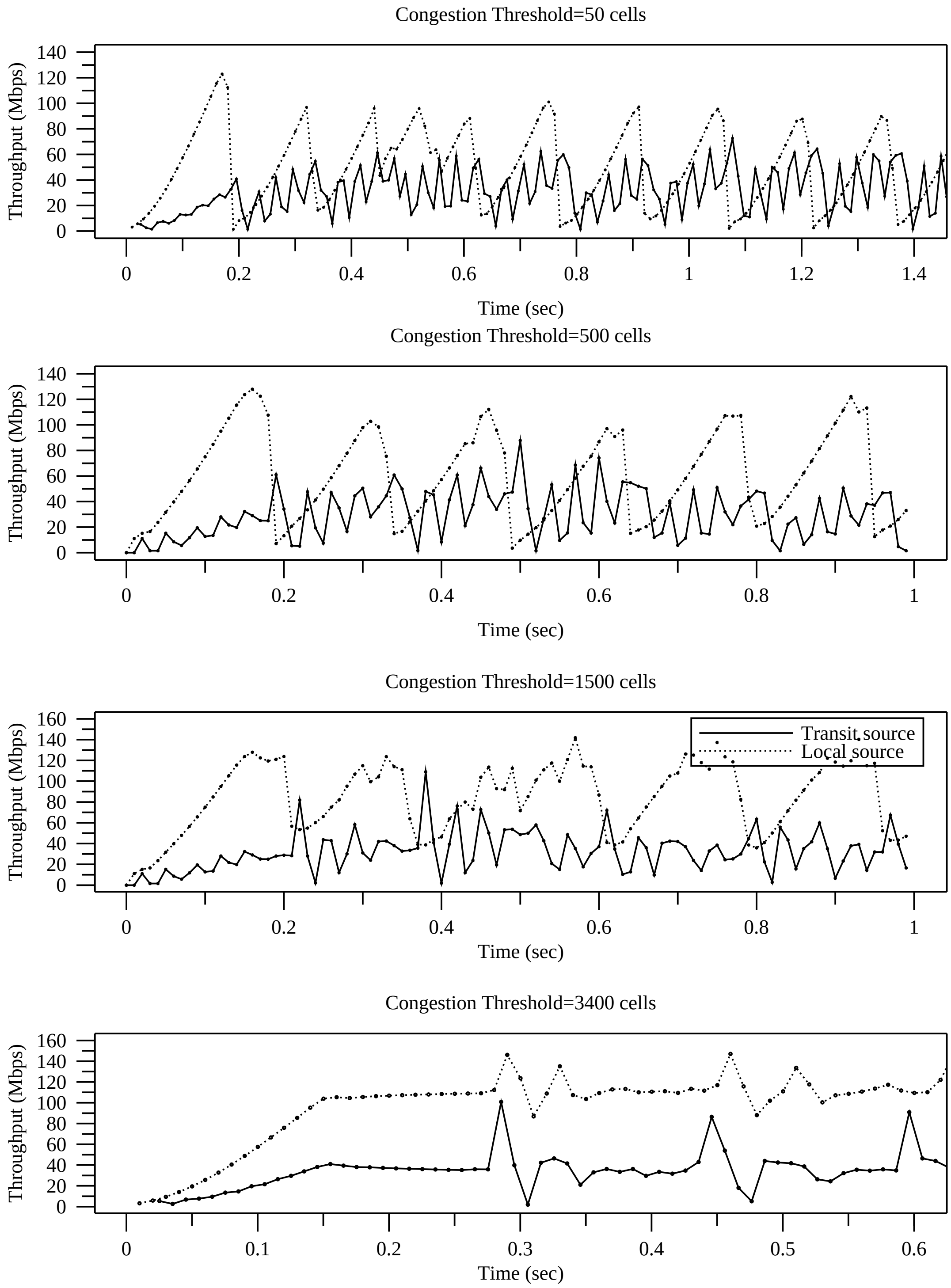

Figure 2: Throughput of PRCA as function of time

1. **Source End System Behavior**
   The source starts sending cells at an initial cell rate (ICR) with EFCI set to 0. Every Nrm sent data cells a RM cell is generated by the source. The RM cells have to at least carry the following parameters [11]:

   - **Allowed cell rate (ACR)**: This value is used mainly to selectively indicate congestion on connections passing a congested link. That is, intermediate networks can signal VCs with a high ACR to reduce their rates while allowing other connections to keep their current rates or even increase them.
   - **Explicit rate (ER)**: The ER field is initially set to the peak cell rate and can be modified downward by intermediate switches to the adequate rate needed to avoid congestion.
   - **Congestion indicator (CI)**: With the congestion indicator compatibility to older switches using EFCI marking can be provided. CI is set to congested (CI=1) in the backward sent RM cells at the destination if the EFCI state in the last received cell indicated congestion. Intermediate switches can also set CI in the backward RM cells if congestion was observed on the forward direction for that VC.
   - **DIR**: Direction of the RM cell, backward or forward relative to the source.

   EPRCA is, just like the PRCA proposal, based on the positive feedback scheme. That is, a source can only increase its sending rate when a positive indication is received from the network. Positive indications are included in the RM cells and depending on the congestion mode used can take one of the following two forms:

   - Receiving a RM cell with (CI=1) indicates that the network is congested and that the sending rate must be decreased.
   - The ER value in the received RM cells indicates the rate that the source should take on to avoid congestion.

2. **Switch Behavior**
   In the EPRCA scheme congestion can be indicated in one of the following two modes:

   (a) **Intelligent marking**: Connections which advertise a high ACR value in their RM cells are signaled to reduce their rates. This can be done through setting the CI bit in the backward RM cells for that VC. Unlike the PRCA proposal which reduces the rate of all sources, this scheme allows connections with low ACR to increase their rate and hence ensures a fair allocation of the bandwidth.

   (b) **Explicit rate indication**: If the explicit rate value noted in the RM cells is higher than the calculated fair bandwidth share at a switch, the switch can decrease the ER value to avoid congestion.

3. **Destination End System Behavior**
   On receiving a RM cell, the destination host generates a new RM cell with DIR set to backward and ER and ACR set to the values of ER and ACR in the received RM cell. To ensure compatibility to older switches using EFCI marking the destination has to keep track of the EFCI state in the received cells. If the last received cell had EFCI set to congested

then CI in the backward RM cell has to be set to 1. Otherwise, CI gets the value of CI in the received RM cell.

For a detailed description of end stations behavior and RM cells see [12].

4.1 Simulation Results

Just like with the PRCA scheme, the EPRCA proposal was tested with the generic fairness topology from Fig. 1 and the parameters of Tab. 1. In addition to that, the multiplicative decrease factor (MDF[‡]) was set to 8 and Nrm was set to 32.

Calculating the fair rate share for a connection and writing it in a RM cell is certainly more complicated than just setting a congestion bit, as was done in section 3. But it actually leads to an optimal rate allocation as well as reducing the oscillation to a minimum. Also, the impact of the initial rate can be eliminated more quickly. The source can start at link rate and the overload will last for only one round trip delay time. In a bit marking based scheme like the PRCA proposal a source would require several round trip times before converging to the optimal value.

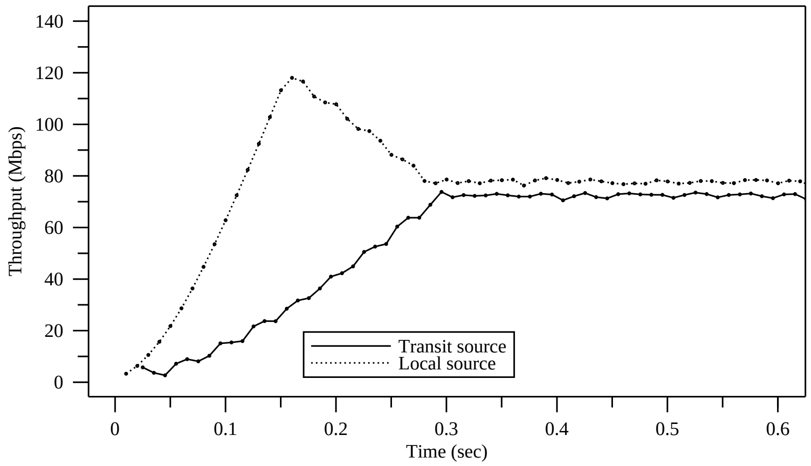

Figure 3: Throughput of EPRCA as function of time

For a congestion threshold of 50 cells the PRCA scheme achieves a network utilization of only about 50%. Using the same congestion threshold the EPRCA proposal achieves high utilization of around 97% and a fairness index of 0.99. Fig. 4 shows that for this configuration a buffer of only 70 cells suffices to avoid cell losses. Under various other topologies, see [6], similar results were achieved.

[‡]Rate reduction in this algorithm is done in the form of

$$\mathrm{ACR} = \mathrm{ACR} - \frac{\mathrm{ACR}}{2^{\mathrm{MDF}}}$$

with ACR as the allowed cell rate of the source.

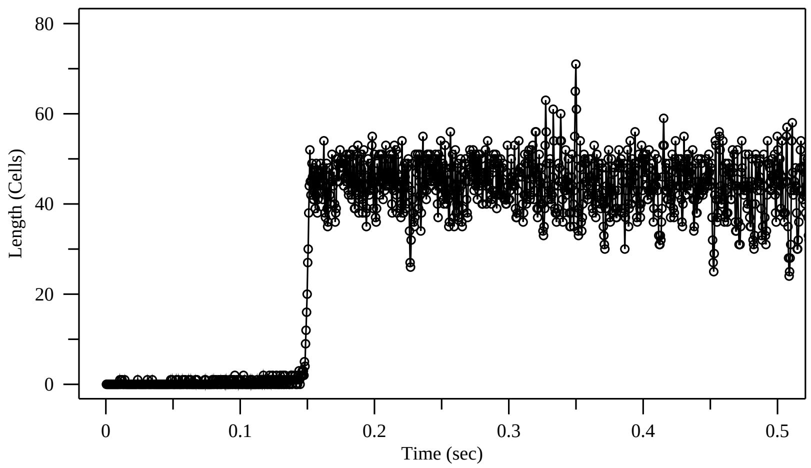

Figure 4: Buffer requirements as a function of time with the EPRCA algorithm

5 Fairness Problems with the Bit Marking Schemes

The current user network interface (UNI) [12] specifications support a binary congestion indication approach similar to the PRCA scheme introduced in section 3. However, instead of using the EFCI bit for congestion indication the source sends a RM cell every N sent data cells. The CI bit in the RM cells is then used for congestion indication.

All the connections used in our test topology, shown in Fig. 1 were persistent sources with the same initial cell rates, the same increase and decrease factors and the same minimum cell rates. As every intermediate switch was shared between two connections we could expect all the connections to receive the same bandwidth share. The only differences noted between the transit connections and the local ones were the round trip delay and the number of switches each connection passes through. In this section we will take a closer look at the effects of these two factors on the fairness performance of the PRCA scheme.

5.1 The Effect of the Round Trip Delay

For investigating the effect of having similar connections with different propagation delays on the fairness performance we used a simple configuration consisting of two connections sharing a single link, see Fig. 5. The short distance connection had a propagation delay of around 10 msec and the long distance one a delay of 30 msec. The congestion threshold was set to 3400 cells and all links had a bandwidth of 155 Mbps.

Fig. 6 shows the throughput achieved when having two connection with different round trip delays. On the shared link a utilization of 90% was achieved and the fairness index was as high as 0.98.

The only source of the slight unfairness noticed in this configuration is the difference in the propagation delay between the two sources. During congestion periods the short distance connections reduce their sending rates after 5 msec and the reduction takes effect at the switch after another 5 msec. The long distance connection, however, will only start reducing its rate after 15 msec. So, as the long distance connection reacts much slower it will suffer for a longer time period from

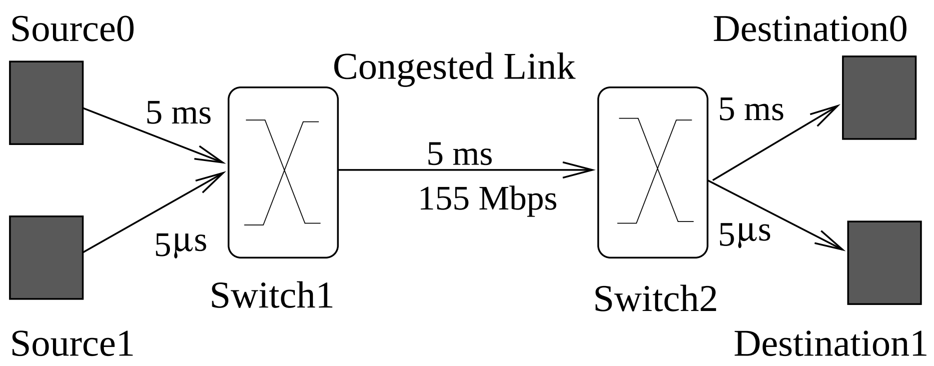

Figure 5: Test configuration for the case of similar connections with different propagation delays

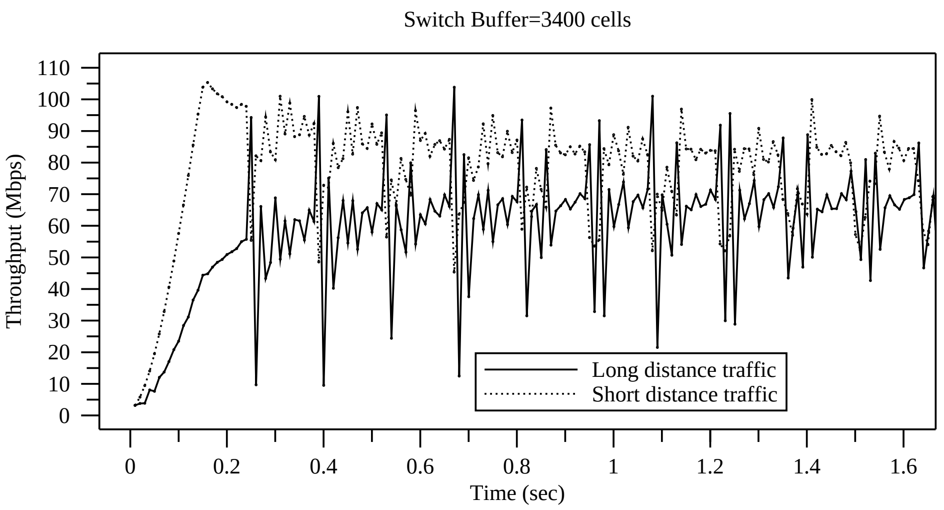

Figure 6: Throughput achieved for similar connections with different round trip delays

congestion and a larger number of cells will get marked. Receiving more marked cells causes then a more severe overall rate reduction at the source. However, this is compensated by two factors:

1. As the source reacts slower to congestion indications, it can keep sending with a higher rate for a longer time period.

2. The short distance traffic share at the switch is reduced after just 10 msec. Therefore, the long distance traffic would be passing for 20 msec, i.e., for the difference in propagation delay between the two connections, a less congested switch and the marking probability would therefore be reduced.

This results in the behavior depicted in Fig. 6. Due to the slower reaction to congestion indications the long distance traffic can reach a peak rate higher than that achieved with the short distance traffic. On the other hand, due to the longer congestion periods that the long distance traffic suffers from, it will reduce its rate to lower values than the short distance traffic.

5.2 The Effect of the Number of Switches

To investigate the effect of the number of the intermediate switches a connection has to pass through on fairness we used a simple configuration consisting of two switches and three sources, see Fig. 7. All the sources were persistent ones with the parameters noted in table 1 and Nrm set to 32 and MDF set to 8. Connection 0 traversed both switches and connection 1 and 2 traversed only one switch. We adjusted the link lengths in such a manner that all connections had the same propagation delay of 30 msec. The congestion threshold was set to 3400 cells.

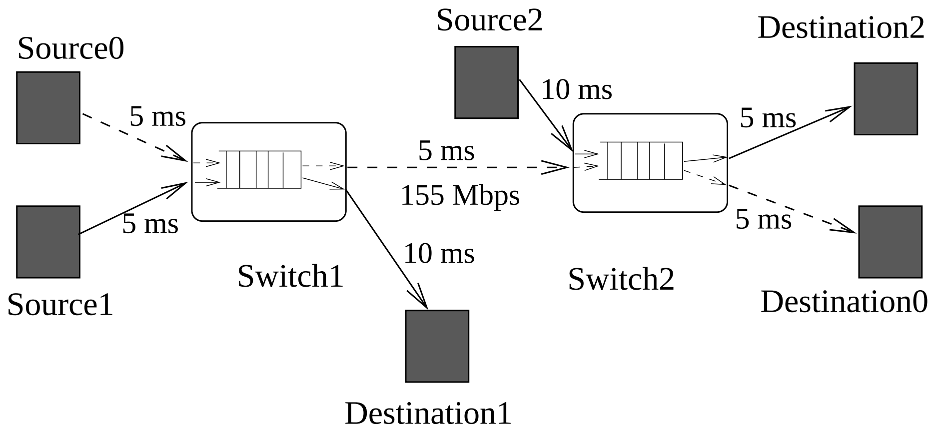

Figure 7: Test network for investigating the effects of having similar connections passing through different numbers of switches

Figure 8 depicts the throughput achieved for a local and the transit connection. A similar behavior to the one shown in Fig. 2 can be seen here as well. Here, a utilization of 80% and a fairness index of only 0.87 were reached with the local connection receiving nearly twice as much bandwidth as the transit one.

This can be explained as follows: As the transit traffic will reach the intermediate switches at different time points due to the link delay the switches will go into the congested state at asynchronous time points. Also, the uncongested state will be reached in an asynchronous manner. So, for example, if the first switch moved from the congested state to the uncongested one, it will stop marking the congestion indication bits in the passing cells. This would allow the local source to increase its rate. The transit traffic on the other hand will still have to pass other links that are still congested and more cells will be marked for this traffic. This results in longer congestion periods for the transit traffic and a more severe reduction of its rate. This problem is commonly known as the “beat down problem”.

5.3 Hop Dependent Multiplicative Reduction

To account for the unfairness of bandwidth distribution that arises when having similar connections passing through different numbers of hops and intermediate switches the transit connection could use a less aggressive reduction factor. For the PRCA proposal we have used the generic fairness configuration with a congestion threshold of 3400 cells and the MDF for the transit source set to 10 and 16 instead of 8. The throughput of the local source and transit source are depicted in figure 9.

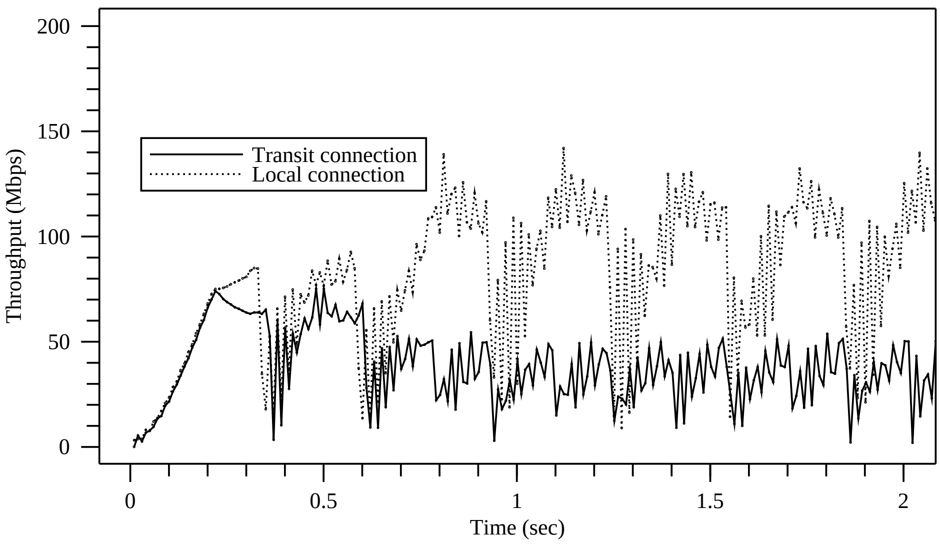

Figure 8: Throughput achieved for similar connections with different number of intermediate switches

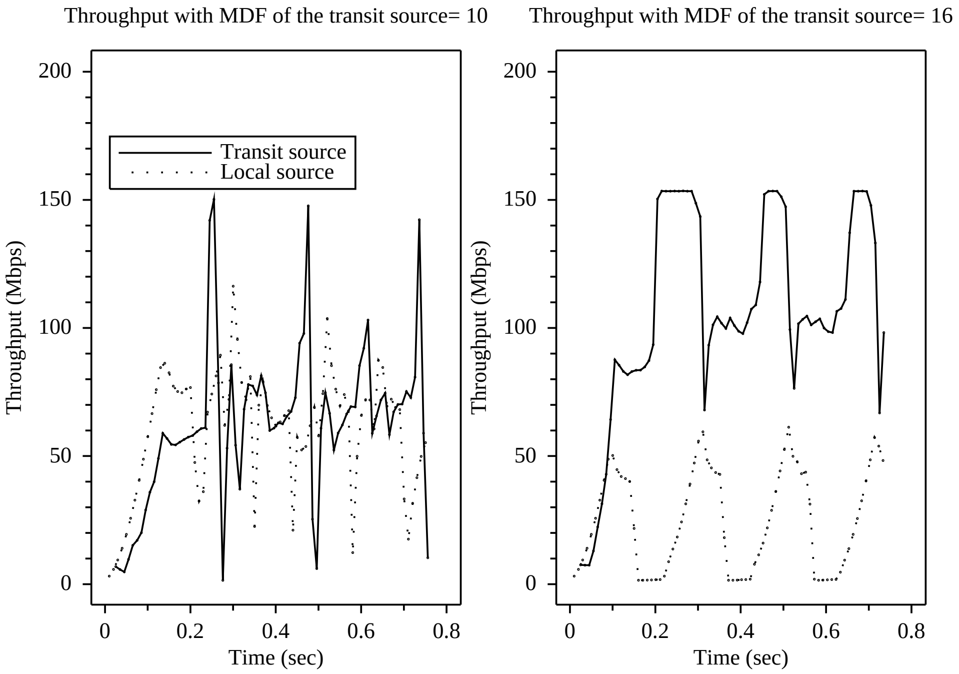

Figure 9: Throughput of a local source and the transit source with MDF= 10, 16

The transit source receives now a much higher share of the bandwidth. Setting MDF 10 results in an optimal bandwidth distribution with each source receiving 50% of the bandwidth. With MDF set to 16 the transit connection receives even four times as much as the local one with the utilization around 83%. The unfairness towards the local source indicates that we have chosen a much too large MDF and that a smaller value would have been probably enough. These results imply that setting the reduction factor in accordance to the number of hops results in improved fairness.

In order to use a hop dependent reduction factor, the RM cells should contain a congestion indication field of several bits instead of the currently used single bit. This field is initialized to 0 at the source and incremented by one at each congested switch. Now, according to the number of passed congested switches noted in the congestion indication field the source would choose the appropriate multiplicative reduction factors with the help of a lookup table. Such a table could be specified experimentally by observing the everyday traffic over a heavily used network.

While this approach seems to be promising in terms of fairness performance, its realization and applicability still need to be investigated. Specifying a lookup table for determining the appropriate reduction factor is surely a nontrivial task. Also, the impacts of such an approach on the buffer requirements at the switches, length of the transient periods for changing traffic conditions and utilization must be further tested.

6 Summary and Future Work

In this study we investigated the link utilization and fairness performance of the PRCA and the EPRCA schemes proposed for congestion control for ABR. Our results suggest the superiority of explicit rate indication schemes over the bit marking ones in terms of link utilization and fairness. However, the currently available switch algorithms for determining the per VC fair bandwidth share require per VC accounting and are therefore very computation intensive [7] or converge only very slowly to the steady state as we have seen here with the EPRCA algorithm. Therefore, further work must be done in order to implement simple yet efficient schemes for the fair rate share calculation.

For ABR to find wide acceptance it must allow for a simple and efficient integration of currently wide spread protocols such as TCP. However, various studies [13,14] show that integrating TCP and ABR might lead to performance degradation in TCP. A simple method for overcoming this problem would be to enhance TCP with a bit marking congestion control algorithm. Our investigations [14] suggest that this would not only ease the integration of TCP and ABR but might even improve TCP's performance. However, the considerable fairness problems noticed require further work.

So, besides its availability and simplicity congestion control using bit marking might actually be the key for integrating ATM in the currently available communication infrastructure and thereby accelerate its acceptance and availability. As a solution to the fairness problems noticed with this scheme using hop count dependent multiplicative decrease factors might result in significant fairness improvements. However, further investigations need to be done on the impacts of using such an approach on the buffer requirements at the switches and the response time required for accommodating changing traffic conditions with bursty connections. Also, the way to determine the appropriate reduction factor as a function of the passed hops could prove to be too difficult, especially when a wide variety of traffic need to be considered.

References

[1] N. Yin and M. G. Hluchyj, "On closed-loop rate control for ATM cell relay networks," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Toronto, Canada), June 1994.

[2] L. Wojnaroski, "Baseline text for traffic management sub-working group," Tech. Rep. 94-0394r5, ATM Forum, Oct. 1994.

[3] R. Jain, "Fairness: How to measure quantitatively?," Tech. Rep. 94-0881, ATM Forum, Sept. 1994.

[4] L. Roberts, B. Makrucki, T. Tofigh, A. Barnhart, B. Holden, J. Diagle, M. Hulchyj, H. Suzuki, G. Ramamurthy, P. Newman, N. Giroux, R. Kositpaiboon, S. Sathe, G. Garg, and N. Yin, "Closed-loop rate-based traffic management," Tech. Rep. 94-0438R1, ATM Forum, July 1994.

[5] D. Huges and P. Daley, "More ABR simulation results," Tech. Rep. 94-0777, ATM Forum, Sept. 1994.

[6] Y. Chang, N. Glomie, L. Benmohamed, and D. Su, "Simulation study on the new rate-based PRCA traffic management mechanism," Tech. Rep. 94-0809, ATM Forum, Sept. 1994.

[7] A. Charny, "An algorithm for rate allocation in a packet-switching network with feedback," Tech. Rep. MIT/TR-601, MIT, Cambridge, Massachusetts, May 1994.

[8] R. Jain, "Congestion control and traffic management in ATM networks: Recent advances and a survey." A PostScript version is available at http://www.cis.ohio-state.edu/ jain/papers.html, Jan. 1995.

[9] A. W. Barnhart, "Explicit rate performance evaluations," Tech. Rep. 94-0983, ATM Forum, Oct. 1994.

[10] R. Jain, S. Kalyanaraman, and R. Viswanthan, "The OSU scheme for congestion avoidance using explicit rate indication," Tech. Rep. 94-883, ATM Forum, Sept. 1994.

[11] H. Hsiaw, K. Fendrick, J. Alpan, L. Roberts, B. Makrucki, T. Tofigh, A. Barnhart, B. Holden, J. Diagle, M. Hulchyj, H. Suzuki, G. Ramamurthy, P. Newman, N. Giroux, R. Kositpaiboon, S. Sathe, G. Garg, and N. Yin, "Closed-loop rate-based traffic management," Tech. Rep. 94-0438R2, ATM Forum, Sept. 1994.

[12] S. S. Shirish, "ATM Forum traffic management specification version 4.0," Tech. Rep. 94-0013R6, ATM Forum, June 1995.

[13] P. P. Mishra, "Throughput degradation for TCP over ATM in the presence of traffic policing," in *Proc. of Gigabit Networking Workshop*, (Boston, Massachusetts), Apr. 1995.

[14] D. Sisalem, "Binary congestion notification in TCP," in *Conference Record of the International Conference on Communications (ICC)*, (Dallas, Texas), IEEE, June 1996. To be presented in ICC'96.

[15] D. Sisalem, "Rate based congestion control and its effects on TCP over ATM," Diplomarbeit, School of Engineering, TU Berlin, Berlin, June 1995.