Performance of TCP over ATM for Various ABR Control Policies (Extended Version)

Carlos M. D. Pazos‡*
Dirceu Cavendish Jr‡*
‡ UCLA – Computer Science Department
405 Hilgard Ave., Los Angeles, CA 90024

Vincenzo A. Signore†
Mario Gerla‡
† Politecnico di Bari – DEE
Via Orabona 4, 70125, Bari Italy

Abstract

Connectionless traffic originating from LANs is carried as Available Bit Rate (ABR) traffic in the ATM network and is subjected to ATM level flow control, either rate-based or credit-based. In addition, transport layer flow control (e.g. TCP window control) is applied at the end points. In this paper, we study the interaction between TCP and ATM flow controls. We use simulation to compare the "goodput" of different ABR flow control schemes and to evaluate the sensitivity to different parameter selections (e.g., TCP Maximum Segment Size (MSS), round trip delay, etc). The study uncovers an unfairness situation caused by use of non uniform MSS values.

1 Introduction

The use of ATM as a backbone for LAN interconnection is probably one of the main initial applications of ATM networks. Figure 1 represents a possible scenario in a large organization or campus. The same model can be used to represent the interconnection of LANs across a nationwide scale network by replacing the ATM switch with an ATM WAN.

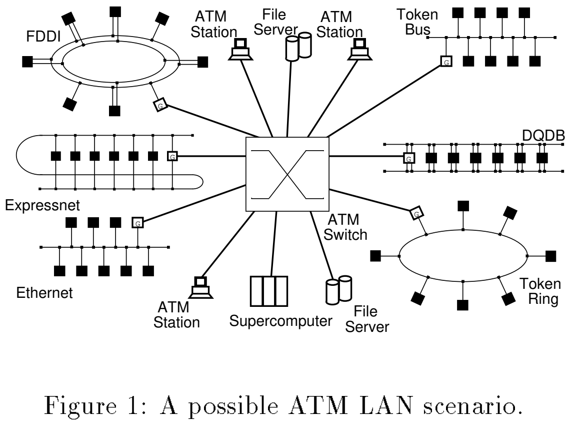

Figure 1: A possible ATM LAN scenario.

Regardless of whether the ATM network is a LAN or a WAN, it must be able to support connectionless traffic, as the traffic in a LAN typically is. This turns out to be a challenge, since ATM requires that sources provide traffic descriptors at call set up time for bandwidth negotiation and assignment. In our case, however, data sources cannot predict their traffic behavior. The ATM network thus carries connectionless traffic on ATM networks on an Available Bit Rate (ABR) basis, with minimum or no bandwidth assignment or QoS guarantee. We wil use the term ABR to refer to both ABR and UBR (Unspecified Bit Rate) traffic, assuming that in the UBR case there is no guarantee of even a minimum bandwidth. Flow control must be exercised to prevent ABR traffic from congesting the ATM network. Two classes of flow control have been proposed: credit based and rate based. The latter approach has been selected by the ATM Forum as the most desirable choice, with PRCA, EPRCA [BF95] and SP-EPRCA [MCG96] being proposals that implement rate control.

In addition to the ATM flow control mechanisms, LAN traffic is also controlled by the Transport layer flow control mechanism, for example, the TCP dynamic window scheme. ATM and TCP mechanisms interact and often reinforce each other. At times, however, this match is less than perfect, causing inefficiencies and even unfairness.

In this paper we investigate the interaction between the TCP flow control and the various ABR controls implemented at the ATM layer, namely, FCVC, PRCA, EPRCA and SP-SPRCA. One aspect of this interaction is the effect of the Maximum Segment Size (MSS), defined by the underlying LAN/MAN technology, on the performance of TCP traffic. Particularly, we consider two cases. In the first one, all sources are restricted to the same MSS value (uniform sources). In the second, The sources are partitioned in groups restricted to different MSS values (non uniform sources). We then study the impact on performance due to the MSS used and examine the unfairness caused by non uniform MSS values.

Next, we investigate the impact of different round trip delays (RTDs). The RTD directly affects TCP performance since, once in a slow start phase, an RTD has to elapse before the TCP window can be expanded. Furthermore, the RTD also affects the performance of some of the ABR controls, causing cells

*Bolsista do CNPq – Brasília/Brasil

to be dropped and segments to be retransmitted.

The performance measure we consider in this paper is the TCP goodput, i.e. the effective throughput perceived by applications running over the TCP entity. Our study is based on simulations of different variations of a tandem network topology.

In the remaining of the paper we briefly discuss the proposed flow control schemes in section 2. The problem we are studying and the simulation model used are presented in section 3. In section 4 we present our simulation results and discuss the unfairness observed. In section 5 we present our conclusions.

2 Flow Control Schemes for ABR traffic

ABR traffic will be carried in ATM networks on a Best Effort basis, with minimal or no bandwidth reservation. The idea is to use ABR traffic as fill in traffic, by using the bandwidth left over by CBR and VBR traffic. Since no UPC (User Parameter Control) can be performed on Best Effort traffic, flow control has to be implemented through feedback from the network congestion points to the traffic sources. Two ways of doing it have received considerable attention at the ATM Forum. In one of them, buffers are reserved at intermediate nodes along a VC and credit information is exchanged between consecutive intermediate nodes, from the destination all the way back to the source. One such scheme is the FCVC [KC95]. In the other approach, the source transmission rate is controlled based on congestion feedback information from the network. Examples of rate-based solutions are PRCA, EPRCA [BF95] and SP-EPRCA [MCG96].

2.1 Flow Control Virtual Circuit (FCVC)

In the FCVC scheme [KC95], each node allocates a buffer to every VC it supports. The size of a single buffer is a function of the link Round-Trip Time (RTT) and a target bandwidth. The node keeps a credit count for each VC, which decreases by one every time a cell is transmitted. Moreover the node keeps a counter for the number of cells sent through each VC. This value is inserted in a field of the credit cell, when it is sent to the upstream node. Upon receiving a credit cell, the upstream node updates its own credit count.

At a given switch output port, the FCVC scheme services a pool of buffers, one for each VC going on the associated network link. There is a cell scheduler at the port that keeps track of the link utilization. When there are no cells from connections with QoS guarantees, the scheduler sends one cell from one of the VC buffers on a round-robing fashion.

2.2 Proportional Rate Control Algorithm (PRCA)

In PRCA the source is allowed to increase its rate only when it receives an explicit indication of no congestion, otherwise it linearly decreases the rate after each cell transmission, down to the minimum value MCR (Minimum Cell Rate). The decrease step is determined by the parameter ADR (Additive Decrease Rate), depending on a negotiated Multiplicative Decrease Factor (MDF). The end-points run a closed management loop: every cell is sent with the Explicit Forward Congestion Indication (EFCI) bit set to zero, and eventually it will be marked by one of the intermediate switches in case of congestion.

The source, after transmitting a given number of data cells, specified by the parameter NRM, sends a Resource Management (RM) cell in order to probe the domain. A Congestion Indication (CI) bit and a Direction (DIR) field are sent on the RM cell payload. The DIR field merely specifies the direction of flow to which the RM cell is been used to probe the network. Once arriving at the destination, the RM cell is looped back and sent to the source.

The source sends the RM cell with the CI bit set to zero. This bit is set to 1 by the destination if the last data cell was received marked, or left zero otherwise. Upon receiving an RM cell with no congestion indication, the source is allowed not only to compensate for the rate decrease in the last cycle, but also to increase it further by an agreed upon Additive Increase Rate (AIR). This increase cannot make the source Allowed Cell Rate (ACR) to go over the Peak Cell Rate (PCR). If the congestion bit is set, the source continues reducing the rate.

2.3 Enhanced Proportional Rate Control Algorithm (EPRCA)

The enhancement of EPRCA over PRCA is accounted for the fact that an extra filed, the Explicit Rate (ER), is added to the RM cell payload. The ER field is initially set to a negotiated PCR by the source and possibly reduced by the intermediate networks (or nodes). When the source receives an RM cell back, the ER field defines the bandwidth the network can provide to that specific source. Each switch in the network has to keep track of the available bandwidth on all its out going links and the number of active ABR connections using each one of them. So, when an RM cell is received, either on the forward or backward direction, the switch divides the available bandwidth by the number of active sources and place that value on the ER field of the RM cell. With this value and given that an RM cell is received with no congestion indication, the source is allowed to increase its ACR. However, the increase cannot exceed either the PCR or the ER. If the RM cell brings a congestion indication, the source rate simply continues its additive decrease.

2.4 Smith Predictor-EPRCA (SP)

SP-EPRCA (henceforth denoted by SP only) is a closed-loop rate based congestion control algorithm in which source rates are adjusted according to VC queue levels. Its novelty is that it uses a Smith Predictor plus a simple proportional controller to overcome the large bandwidth $\times$ delay product typical of high speed wide area networks [MCG96]. The queue levels are used as feedback information, transported by RM cells.

The source rate is computed so that Λall cells stored at the most congested queue along the feedback loop plus the "inflight" cells inside the pipe are accounted for. Hence, the source rate is regulated according to the space available at the bottleneck, taking into account the number of "inflight" cells, which are computed by keeping track of the traffic pumped into the feedback loop for the last round trip delay. Of course, knowledge about the RTD of the connection is assumed.

This algorithm considers all the "inflight" cells as if they were in the bottleneck, i.e. it considers "in pipe cells" like a pseudo queue added to the bottleneck. Intuitively this can be seen as a way to overcome the problem of the delayed feedback information. The resulting behavior of the bottleneck queue level is stable, cell loss free and without oscillations.

3 The Simulation Model

We have developed a Maisie [BL90] simulation model to evaluate the system shown in Figure 2. In our model the ATM network is carrying both ABR and VBR traffic. The use of VBR sources allow us to study the TCP traffic performance when sharing a communications link with higher priority traffic.

We consider three different sets of experiments. First we compare the performance of different congestion control schemes and evaluate the effect of the MSS parameter. In this sense, all stations send data traffic to Destination 1 (see Figure 2) and they are all restricted to use the same MSS. Furthermore, we perform the study for the case of a campus ATM when we have only one switch and contrast it with the case of an ATM MAN when we have seven switches (as presented in Figure 2). In both cases, each link is considered to be 7Km long.

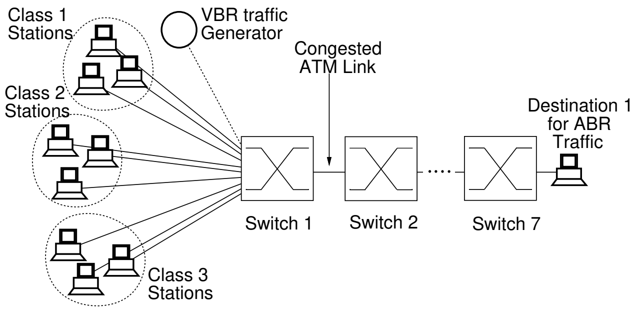

Figure 2: ATM LAN model.

In the second experiment, we study the effect of having the ATM network carry traffic from different types of LAN, as it is the case in Figure 1. Different LAN/MAN technologies put different restrictions on the maximum PDU size (MSS) that can be used by TCP entities. So, we divide the sources into three classes of sources as shown in Figure 2. Each class is restricted to a different MSS.

In the third experiment, we study the effects of having longer links, which translates into larger buffers. For this case study, we consider a cross-country distance separating the sources from the destination. We have the same configuration of seven switches interconnected by 600Km links.

We have developed our own simulator, including the code for the TCP entities. Our TCP implementation is based on the TCP description found in [Com91] with the slow-start feature proposed in [JK88]. A more efficient TCP implementation, the TCP Vegas [BOP94, ADLY95], is now available and we plan to use it in future works. However, albeit less efficient, our implementation allows us to study the performance of window flow controlled TCP traffic carried over ATM networks with ABR flow control.

For simplicity, we assume that the VBR traffic (background traffic) originates from a generator that produces the aggregated traffic of VBR sources, and this traffic is offered to the first switch through a single ATM link. The VBR traffic is always sharing the resources of the congested link with the ABR traffic, but with higher priority. To implement the VBR generator, we have used an MMDP model and adjusted the parameters such that the congested link has some bandwidth, not reserved to the VBR sources, available to be shared among the ABR sources.

All the ABR traffic is generated by a set of 15 workstations, connected possibly to different LAN/MANs, which are engaged in some connectionless communication with a common destination. Each workstation is running an application that is generating traffic simulated by an ON-OFF model. The source parameters are adjusted such that the aggragate TCP traffic requires more bandwidth than what it is left-over by the VBR traffic. Consequently, the stations are always backlogged and we can study which ABR flow control scheme is more robust and effective in using up the bandwidth left over by the VBR traffic.

In our experiments, 15 workstations generating ABR traffic are connected to the ATM switch through a gateway and that the TCP MSS is limited by the respective LAN/MAN to which the workstation is connected. For the case in which we study the traffic submitted by different classes of sources, the 15 workstations are divided into three classes of sources, with 5 workstations each. Workstations in the same class use a uniform MSS. We have chosen the values 512, 1024 and 2048 for the MSS used in each one of the classes. For the study of the TCP performance when all sources use the same MSS, we also have 15 workstations and the MSS used were 256, 1024, 2048, 8192.

For the ATM switch, we have modeled the output buffered architecture in [SNS+89]. We have implemented priority queueing for VBR and ABR cells, with ABR queue serviced only when no cells are available on the VBR queue. Furthermore, for the FCVC scheme, the ABR queue is subdivided into different per VC queues. The buffer space allowed for each queue is taken from [KC95]. For the short links the per-VC buffer accommodates 11 cells, while in the longer links used in the cross-country study, the buffers can hold 81 cells. This is also the buffering used for the SP scheme. Since in our study we have 15 ABR sources, in order to make a fair comparison we dimensioned the PRCA and EPRCA buffers to be 165 cells for the short links and 1,215 cells for the longer links.

4 Results

In this section we report and discuss the simulation results we have produced. The simulations for one switch were made to last about 8.5s, while the ones for seven switches lasted about 10s of simulated time. These simulated times were long enough to produce the 95% confidence intervals. Since these intervals are fairly narrow for the majority of our results, we assume that the values for the simulated times were more than adequate to represent a steady-state behavior. However, in order to alleviate the figures, we refrained from plotting the confidence intervals. During these simulated times, the number of cells switched is in the order of 3.5×10^6.

4.1 The TCP Performance

In Figures 3 and 4 we present the TCP goodput as a functions of the MSS when the TCP traffic is sent over connections crossing one and seven switches, respectively. In both cases, we have that the links are 7Km long. Hence, by considering the number of switches crossed we investigate the effect of increased end-to-end round trip delay, without modifying the buffering requirements. The goal here is to determine which congestion control scheme for ABR traffic provides the best performance when the ABR traffic is generated by TCP sources in a homogeneous environment (all sources use the same MSS). We evaluate the performance (goodput) as a function of TCP MSS and round-trip delay.

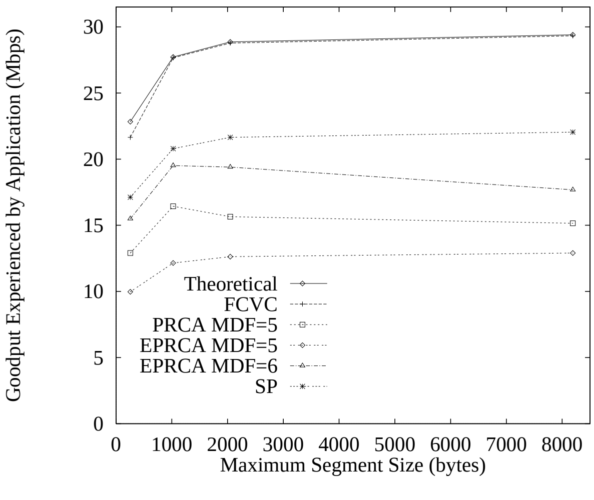

Figure 3: TCP goodput for a single-switch topology.

The first thing we observe from Figures 3 and 4 is the effect of the MSS on performance. Since the amount of protocol overhead is the same regardless of MSS, the larger the MSS the better the efficiency in the transport of data bytes. On the other hand, longer MSS means that more data has to be retransmitted in case of cell loss. For FCVC and SP, no cells are lost

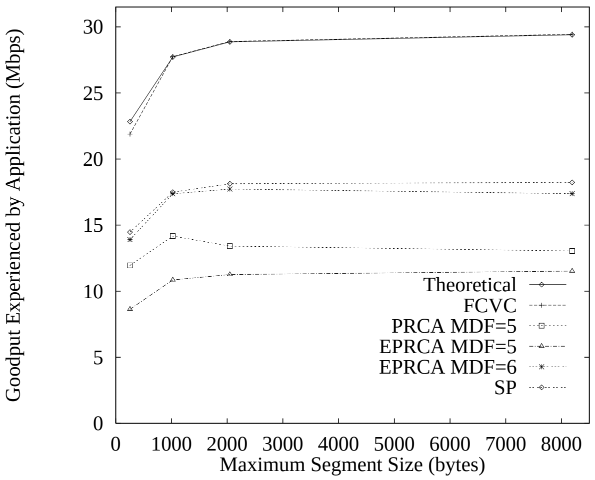

Figure 4: TCP goodput for a seven-switch topology.

and a large MSS only improves performance. That is why we observe the increase in goodput with increased MSS value. For PRCA and EPRCA, the loss of cells and consequent retransmission of segments defies the overhead efficiency of large MSS values. Hence, the decrease of goodput with MSS.

Concentrating on FCVC only, we note that this scheme is unaffected by the difference in round trip delay when considering a one-switch versus a seven-switch topology (see Figures 3 and 4). This is because FCVC employs back-pressure and buffer reservations. The reservations along the path prevent cell loss which would otherwise trigger TCP retransmissions and cause goodput degradation. Furthermore, Figures 3 and 4 also shows that the goodput experienced with FCVC is very close to a theoretical bound [GPS95].

Next we investigate the PRCA scheme for which we used an MDF value of 5. This value for the MDF was chosen to make sure that if a source started transmitting at its PCR after receiving one RM cell, its ACR would be reduced to the MCR by the time it receives the next RM cell. This way, if the source does not receive the first RM back by then, it would be forced to go on transmitting only at the MCR. More conservative values for the MDF (smaller MDF) would unnecessaryly slow down the sources even further, thus lowering the TCP goodput. An MDF value of 6 was also tested but with no performance gains since more losses occur with this more aggressive setting.

Comparing the results from Figures 3 and 4 for the PRCA scheme, we notice a reduction in the level of goodput due to the increased distance. The reduction in the level of performance is due to the fact that the feedback actually brings back already out of date congestion information. This leads to higher chances of losing cells, corrupting TCP segments which have to be retransmitted. In this case, the larger the MSS, the worst the goodput since more data must be retransmitted. Furthermore, sources having to retransmit a segment are forced by the TCP window flow control to wait for an acknowledgement before transmitting another segment. Since there is an increase in the round trip delay with seven switches, the sources waiting for the acknowledgement cannot use their share of bandwidth which is thus used by the remaining sources uneffected by losses.

For the EPRCA scheme we took the same approach as for PRCA and we simulated the cases with MDF = 5 and MDF = 6. Since the EPRCA already uses the ER to limit the increase in the ACR, MDF = 5 proves to be too conservative. With ACR being set to ER, the use of MDF = 5 implies that after transmitting 32 cells, the ACR is back to the MCR if no RM cell has returned. This is too conservative because the last RM cell guarantees that a rate ER could be supported. However, the sources do not use that rate for very long. With the $MDF = 6$, the sources reduce their ACR a bit slower and thus can use more of the available bandwidth. As a result, the goodput experienced with EPRCA with MDF = 6 is always better than the results with PRCA but still far worse than the results obtained with FCVC. Furthermore, since the ACR increase is limited by the ER, the chances of losing cells and thus the waste of resources by retransmissions is reduced. It would seem reasonable to increase MDF to 7, for an even better performance. However, MDF = 7 did not improve performance. We omitted these results in order not to overload the figures. With longer distances (i.e., seven switches), we notice the same effect of goodput reduction due to the out dated congestion notifications already observed with PRCA.

For the SP scheme, there is no parameter tuning involved and only one curve was plotted. As pointed out earlier, this scheme implements a rate based control which guarantees no loss. As a result, no bandwidth is wasted with retransmitted segments. However, the scheme relies on feedback from the network, thus it depends on round-trip delays. For these reasons, the SP performs better than the EPRCA or PRCA schemes in the topology with only one switch. As the distance is increased on the topology with seven switches, the loss free property is offset by the dependence on RTD. The results in Figures 3 and 4 clearly show the SP performance dependence on the RTD alone.

To sum up, for the two network configurations considered in this section, we notice that FCVC always outperforms both PRCA, EPRCA and SP. Despite its implementation complexity, FCVC does provide an upper bound on performance. EPRCA can provide a better performance than PRCA. However, the performance improvement is rather small.

4.2 Non Uniform Sources

In this section we are concerned with the non uniform case as illustrated in Figure 1. To simulate such an environment, the TCP sources were divided into three groups with a different MSS in each group. The aggregate goodput experienced by the sources in each class is plotted in Figures 5 and 6, for a one-switch and a seven-switch topology, respectively. Like in the uniform case, the FCVC scheme provides a better goodput than PRCA, EPRCA and SP. The small difference in goodput for different classes is due to different protocol overhead. Hence, sources using a larger MSS experience better goodput than those using a smaller MSS. This does not represent unfairness, however.

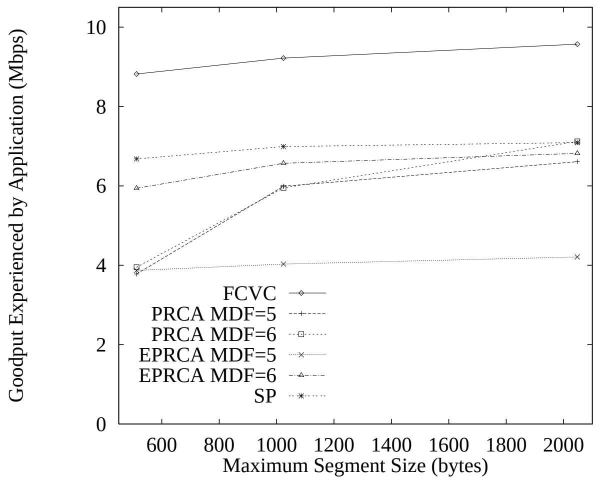

Figure 5: TCP goodput for a non uniform case, on a single-switch topology.

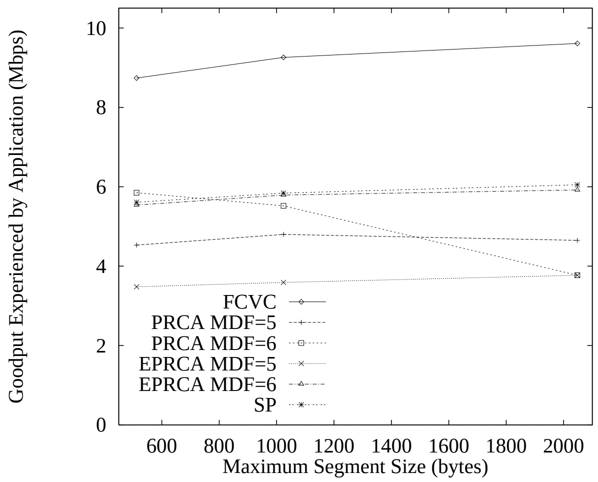

Figure 6: TCP goodput for a non uniform case, on a seven-switch topology.

For the EPRCA scheme in a heterogeneous environment, as in the homogeneous one, an MDF = 6 is less conservative and thus provides for a better utilization of resources and a better goodput than with MDF = 5. The effect of longer round trip delays has also caused here a reduction in the level of goodput experienced by the sources. However, as with FCVC, with EPRCA there is no unfairness due to different MSS. This is guaranteed by the fact that EPRCA uses ER to limit ACR, which in turn reduces the chance of cells loss.

On the SP curves of Figures 5 and 6 we show that this scheme also provides fair treatment to the non-uniform sources. As in on the uniform scenario, the SP performance is very sensitive to the RTD.

Unfairness, on the other hand, manifests itself with PRCA, as can be observed from Figures 5 and 6. Furthermore, the unfairness varies with distance. Two factors are responsible for the unfairness, and have an impact which depends on the round trip delay. The first factor is related to the retransmission of corrupted segments. The larger the MSS the more resource is wasted on retransmissions.

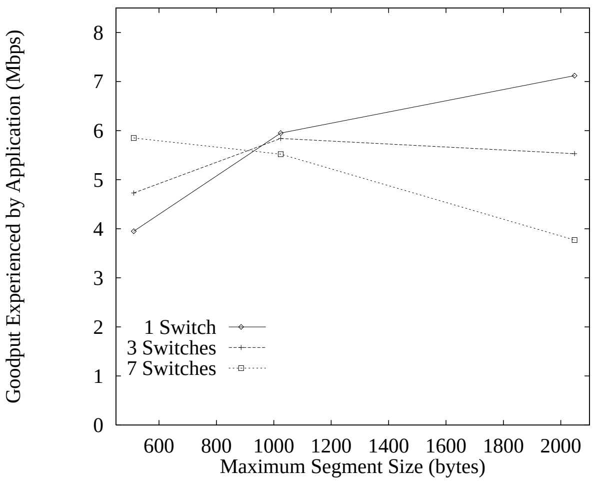

Figure 7: Unfairness behavior of PRCA with MDF 6.

The second factor has to do with the interaction of the TCP window flow control and the PRCA congestion control which penalizes sources using small MSS. When congestion occurs in the network, the RM cells are returned with congestion indication, which in turn causes the sources to reduce their ACR to MCR. Furthermore, if the sources are not able to reduce their rates fast enough, cells are dropped and the TCP window flow control moves into a slow start phase. At this point, we can maintain only one outstanding unacknowledged segment at a time. So, in the slow start phase, the source transmits one segment and then waits for an acknowledgement before sending the next one. Since we also have PRCA in effect, possibly restricting the ACR to the MCR, a source using a small MSS has to wait for the transmission of a few segments before it can send another RM cell to probe the network for congestion. For large MSS, the transmission of a single segment is enough to allow the source to send one or more RM cells. So those sources using

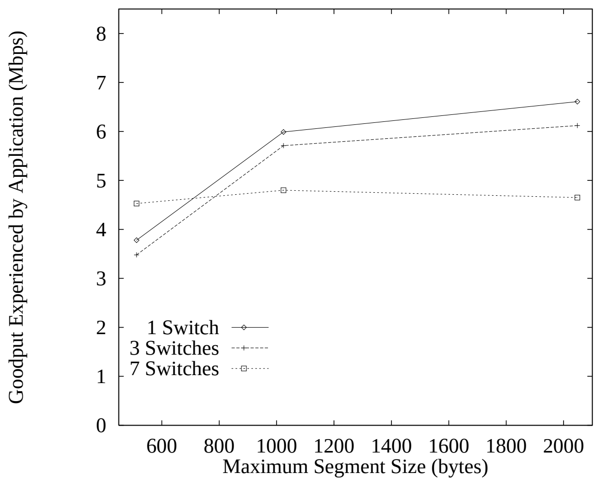

Figure 8: Unfairness behavior of PRCA with MDF 5.

larger MSS can probe the medium more often and reclaim bandwidth faster once congestion subsides, as clearly shown in Figure 5.

To better explain the unfair behavior of PRCA carrying TCP traffic, we have considered one more topology with three switches. We have plotted the results for MDF = 6 in Figure 7 and MDF = 5 in Figure 8 for the three topology configurations considered so far. For MDF = 6 we notice that increasing the number of hops traversed, and consequently the round trip delay, the unfairness changes from been towards the small MSS users towards the large MSS users as illustrated in Figure 7. With increased round trip delay, the congestion indication brought back by the RM cells is already out dated and more losses start to occur, penalizing more the sources using larger MSS. For these sources, much of the bandwidth they seize is actually wasted with the transmission of corrupted segments. Furthermore, with more losses we are more likely to have multiple retransmissions of a single segment corrupted on consecutive transmissions. This leads to doubling of the time-out timer set by the TCP entity for the arrival of acknowledgements. This silences the large MSS users allowing the small MSS user to seize a larger share of the available bandwidth.

A similar effect is observed in Figure 8. In this case, however, the use of MDF 5 is more conservative as we have already discussed. As a result, the effect of having out dated congestion information is not enough to reverse the unfairness observed but for longer round trip delays. So, for a three-switch topology, the outdated information has almost the same effect for all classes of users and hence we notice only a reduction in the level of goodput experienced by the sources as compared with a one-switch topology. For a seven-switch topology the out dated information is noticed and we experience a rather fair behavior. However, for larger round trip delays, we would again observe unfairness towards the large MSS users.

In conclusion, in non homogeneous environment, FCVC, SP and EPRCA provide for fair access to the shared resources to all sources, regardless of MSS or distances involved. For PRCA, the goodput experienced by a given source is very much dependent on the RTD and the MDF values involved. This is a very significant result since the ATM Forum has selected rate-based control as the way to do ABR flow control. However, the only recommendation so far is that switches are to implement congestion indication. There is no requirement that the switches should compute ER for the active sources. Hence, if all switches implement congestion indication only, they will actually be using PRCA which may lead to some unfairness.

4.3 Cross-country Distances and Buffering

In this section we study the effect of long propagation delays on TCP performance. In this configuration of our model, we are considering links which are 600Km long. For the seven-switch connection, we have a cross-country configuration. The results are plotted in Figure 9.

The fact that we have long links implies that we now require larger buffers as discussed in section 3. Since we have the same input traffic as in the result on Figure 4, larger buffers give the switch more capability to handle less conservative ABR controls. This is precisely what we observe in Figure 9. For the EPRCA case, the results for MDF = 6 is very poor compared to the results in Figure 4 in which MDF = 6 was the best we could get with this scheme. Because we have a very large round trip delay, the ER advertised back to the sources is much out dated. However, the large buffers allow for the accommodation of the rate mismatch between the sending rate and the actual available bandwidth. This is so much the case that even for larger MDF values, say 7 and 8, we still get an improvement in performance over the MDF = 6 case. These two latter values make the EPRCA reduce the source rate very slowly and yet the mismatch between sending and available rates is accommodated.

Another fact is that with large buffers (81 cells per VC for FCVC and 1,215 cells for PRCA and EPRCA), fewer RM cells are returned with a congestion indication. This fact is important because the rate adjustment to the ER in the source is only executed when a non-congestion RM cell is received. This allows the sources to make more use of the advertised rate on the RM cells.

Notice that for PRCA the case is exactly the same. As we have discussed before, PRCA is very aggressive and not very conservative. When an RM cell is received with no congestion indication, the ACR is often set to the PCR regardless of whether that rate can be supported by the network. However, PRCA with an MDF = 5 reduces the rates very quickly in between arrival of RM cells. So, the PRCA aggressiveness with MDF = 5, coupled with large buffers to accommodate the rate mismatches is responsible for the PRCA better performance in this scenario, as shown in Figure 9. We also notice that increasing the aggressiveness of PRCA by using an MDF = 6 did not prove fruitful.

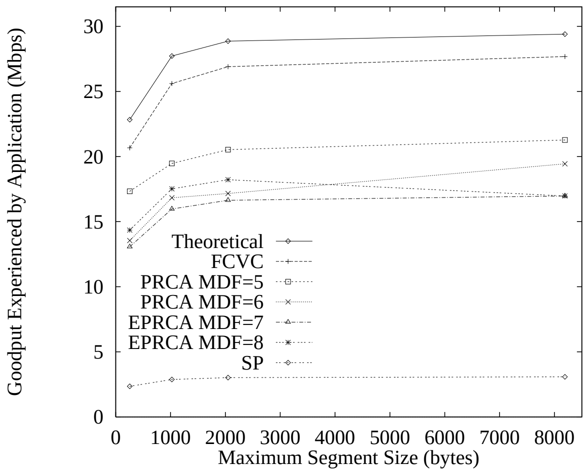

Figure 9: TCP goodput for a cross-country seven-switch topology.

The SP scheme is heavily penalized by the large RTD value involved in this experiment. Since SP does not provide a minimum cell transmission rate, the sources are forced to stop transmissions while they wait for network feedback information. This problem could only be alleviated if more buffer space were provided at the intermediate switches. An alternative remedy (which we have not yet tested) is to allow dynamic sharing of buffers by all VCs.

The results for the FCVC scheme have presented a reduction in performance for such large round trip delays. This is a rather unexpected result since buffers were dimensioned such that maximum link utilization can be achieved! This behaivior is explained by the results in Figures 3 and 4. Figures ?? and ?? show the link utilization for the cases of short 7Km and long 600Km links, respectively. In both cases we have a seven switch topology. In the 600Km-link case, the advantage of the FCVC is not taken because of TCP/ATM interaction. Since sources always start transmitting in the slow start phase, the window takes time to expand. At the ATM level, the sources are allowed to transmit their segments at full link speeds given they have ATM credits for that. However, the very large round trip delay requires the TCP entity to wait for an entire round trip delays before submitting a new set of segments, defying the FCVC ability to send more data exploiting the 81 cells worth of credits (more than enough to transmit at least a few segments). Of course, the simulated time is not long enough to allow the TCP windows to extend enough to fully utilize the available bandwidth.

In this section we see that the RTD directly affects the TCP performance, defying the benefits of an almost perfect fill in characteristic provided by the FCVC scheme. It also shows that it is not cost effective to implement a very complex control scheme to squeeze out the last ounce of bandwidth if the sources are not offering enough traffic to use that extra bandwidth because of higher-layer intervention. In this section we also observe that the rate control schemes can benefit very much from large buffers. This makes sense since the congestion generated by the sources is not realized by the source before at least a RTD.

5 Conclusions and Open Issues

In this paper we have compared the performance of TCP traffic carried as ABR traffic over ATM networks, with several different ABR flow control schemes (FCVC, PRCA, EPRCA and SP). We have also investigated the effects of TCP maximum segment size, end-to-end round-trip delay, and buffering at the ATM switches on the TCP/ATM interaction. Unquestionably, the FCVC scheme provides for the best performance for all configurations considered. However, FCVC is very complex to implement and can be fully exploited only if the sources are capable of offering enough traffic to use the available bandwidth. In this paper we considered only situations in which the TCP entities were always backloged. We concluded that even in such stress cases the potentials of FCVC cannot be fully utilized if the distances involved are very long. This situation can only get worse since the stress conditions discussed here are good to study the control scheme robustness under high load. In fact, the networks are going to be much less heavily loaded most of the time. Thus, the cost effectiveness of very robust and complex scheme is very poor. In any event, the FCVC scheme works well and provides for an upper bound on maximum performance.

We are then left with the rate-based control schemes. The ATM Forum has already ruled that rate-based control should be used for ABR traffic flow control. However there is still plenty of room for improvement. Suggesting that the switches should only provide cell marking for congestion indication, allows switch manufacturers not to implement the explicit rate computation for all VCs traversing the switches. In such case, the sources cannot implement the EPRCA scheme, leaving PRCA as the only alternative. This is not a good idea since in this paper we showed that PRCA can be very unfair in a realistic non homogeneous TCP environment.

There are still a number of open issues that need to be addressed. For example, we showed that switch buffering significantly affects ABR performance and it should be configured to accommodate different round trip delays. But, how much buffering is needed remains to be defined. Also, dynamic buffer sharing startegies should be investigated for both FCVC and SP.

Both PRCA and EPRCA provide for a certain set of parameters to control the source input rate. Further research has to be done into defining the appropriate values to be assigned to these parameters for reasonable performance. In this study, we explored the sensitivity of one such parameter, the multiplicative decrease factor (MDF), and have offered some insight into its role on performance. Much more research should be done in determining appropriate range of values for the various control parameters.

Acknowledgments

We would like to thank Abra Davini for her contributions to the simulation study presented here, for many hours of patient discussions, and for many more lines of code.

References

[ADLY95] J. S. Ahn, P. B. Danzig, Z. Liu, and L. Yan. Evaluation of TCP Vegas: Emulation and experiment. In *Proc. of SIGCOMM '95*, Cambridge, MA, 1995. ACM.

[BF95] F. Bonomi and K.W. Fendick. The rate-based flow control framework for the avaialble bit rate ATM service. *IEEE Network*, 9(2), March 1995.

[BL90] R. Bagrodia and W. Liao. Maisie: A language and optimizing environment for distributed simulation. In *1990 Simulation Multiconference: Distributed Simulation*, San Diego (CA), January 1990.

[BOP94] L. S. Brakmo, S. W. O'Malley, and L. L. Peterson. TCP Vegas: New techniques for congestion detection and avoidance. In *Proc. of SIGCOMM '94*, London, England, 1994. ACM.

[Com91] D. E. Comer. *Internetworking with TCP/IP*, volume vol. I. Prentice Hall, 1991.

[GPS95] Mario Gerla, Carlos M. D. Pazos, and Vincenzo A. Signore. Comparing ATM credit-based and rate-based controls for TCP source. In *Proc. of MILCOM '95*, SanDiego, CA, November 1995.

[JK88] Van Jacobson and Michael Karels. Congestion avoidance and control. In *Proc. of SIGCOMM '88*, Stanford, CA, August 1988. ACM.

[KC95] H. T. Kung and K. Chang. Receiver-oriented adaptive buffer allocation in credit-based flow control for ATM networks. In *Proc. INFOCOM '95*, April 1995.

[MCG96] S. Mascolo, D. Cavendish, and M. Gerla. ATM rate based congestion control using a Smith Predictor: an eprca implementation. In *Proc. INFOCOM '96*, March 1996.

[SNS+89] H. Suzuki, H. Nagano, T. Suzuki, T. Takeuchi, and S. Iwasaki. Output-buffer switch architecture for asynchronous transfer mode. In *IEEE International Conference on Communications*, Boston (USA), June 1989.