

\*\*\*\*\*

ATM Forum Document Number: ATM Forum/96-1267

\*\*\*\*\*

Title: ABR Switch Algorithm Testing: A Case Study with ERICA

\*\*\*\*\*

Abstract:

This contribution discusses the requirements of ATM Available Bit Rate switch algorithms, and demonstrates how each of these requirements can be tested. As a case study, the performance of the ERICA switch algorithm [1] is evaluated and the effect of some features and options of the algorithm is examined. The requirements tested include: efficiency, delay, fairness, transient and steady state performance, scalability, and adaptation to variable capacity and various source traffic models. The performance of the algorithm is evaluated for various configurations and background traffic patterns.

\*\*\*\*\*

Source:

Raj Jain, Sonia Fahmy, Shivkumar Kalyanaraman, and Rohit Goyal  
The Ohio State University  
Department of CIS

Raj Jain is now at Washington University in Saint Louis, [jain@cse.wustl.edu](mailto:jain@cse.wustl.edu) <http://www.cse.wustl.edu/~jain/>

The presentation of this contribution at ATM Forum is sponsored by NASA.

\*\*\*\*\*

Date: October 1996

\*\*\*\*\*

Distribution: ATM Forum Technical Working Group Members (Traffic Management)

\*\*\*\*\*

Notice: This contribution has been prepared to assist the ATM Forum. It is offered to the Forum as a basis for discussion and is not a binding proposal on the part of any of the contributing organizations. The statements are subject to change in form and content after further study. Specifically, the contributors reserve the right to add to, amend or modify the statements contained herein.

\*\*\*\*\*

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>ABR Congestion Control</b>	<b>3</b>
<b>3</b>	<b>Objectives of Traffic Management</b>	<b>4</b>
<b>4</b>	<b>The ERICA Algorithm</b>	<b>6</b>
4.1	The Original Algorithm . . . . .	6
4.2	Achieving Max-Min Fairness . . . . .	8
4.3	Fairshare First to Avoid Transient Overloads . . . . .	8
4.4	Other ERICA Features and Options . . . . .	8
<b>5</b>	<b>Performance Evaluation of ERICA and ERICA+</b>	<b>10</b>
5.1	Parameter Settings . . . . .	10
5.2	Efficiency . . . . .	11
5.3	Minimal Delay and Queue Lengths . . . . .	14
5.4	Fairness . . . . .	14
5.5	Transient and Steady State Performance . . . . .	17
5.6	Adaptation to the Presence of Multiple Traffic Classes (Variable Capacity) .	21
5.7	Adaptation to Various Source Traffic Models . . . . .	25
5.7.1	Bursty Traffic . . . . .	27
5.7.2	ACR Retention . . . . .	36
<b>6</b>	<b>Conclusions</b>	<b>36</b>

# 1 Introduction

This contribution describes how to test various requirements of ATM Available Bit Rate (ABR) switch algorithms. As a case study, the performance of the ABR switch algorithm ERICA is evaluated for various configurations, source models and background traffic patterns. The Explicit Rate Indication for Congestion Avoidance (ERICA) algorithm was presented at the Forum in February 1995 and the complete description of the algorithm was compiled in [1].

The goal of this study is to demonstrate the design of experiments to test all aspects and requirements of ABR switch algorithms. ABR switch algorithms need to meet the goals of efficiency, fairness, fast transient response, scalability to various distances and number of virtual circuits, and adaptability to various background traffic patterns and different source traffic models. To conduct a comprehensive evaluation of the performance of an algorithm, the algorithm must be tested in a variety of situations that ensure it meets all these goals.

The remainder of this contribution is organized as follows. First, the ABR congestion control model is briefly examined, followed by a discussion of the goals and requirements of switch algorithms. The ERICA algorithm and its enhancements and options are then outlined. The rest of the contribution presents the set of experiments conducted to ensure that ERICA meets all the requirements of switch algorithms. In the cases where the original algorithm failed to meet the requirements and an enhancement to the algorithm was deemed necessary, the performance of the basic algorithm is compared to the performance of the enhanced algorithm, and a discussion of why the enhancement was needed is presented.

## 2 ABR Congestion Control

ATM cells flow along predetermined paths called virtual circuits (VCs). End systems must set up Constant Bit Rate (CBR), Variable Bit Rate (VBR), Available Bit Rate (ABR) or Unspecified Bit Rate (UBR) virtual circuits prior to transmitting information. Bandwidth is dynamically divided among the active VCs. Data traffic, which is highly bursty and does not have strict delay requirements, is usually transported by the ABR service. Supporting the ABR traffic class requires traffic management at the intermediate switches in the network.

The ABR service periodically advises sources about the rate at which they should be transmitting. The switches compute the available bandwidth and divide it fairly among active VCs. The feedback from the switches to the sources is indicated in Resource Management (RM) cells. The RM cells are periodically sent by the sources and turned around by the destinations (see figure 1).

The RM cells contain the current cell rate (CCR) of the source, in addition to several fields that can be used by the switches to provide feedback to the sources. One of those fields, the Explicit Rate (ER) field, indicates the rate that the network can support at that time. Each switch on the path of the VC reduces the ER field to the maximum rate it can support. The

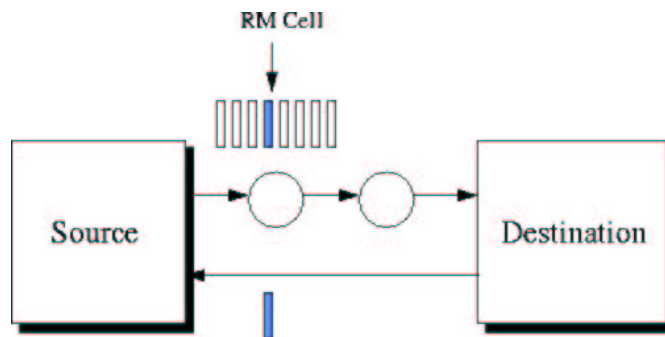


Figure 1: RM cell path

sources examine the returning RM cells and adjust their transmission rates.

The RM cells flowing from the source to the destination are called Forward RM cells (FRMs) while those returning from the destination to the source are called backward RM cells (BRMs). When a source receives a BRM, it computes its allowed cell rate (ACR) using its current ACR, two of the flags in the RM cell, and the ER field of the RM cell. For a detailed explanation of source end system behavior, see [8].

### 3 Objectives of Traffic Management

Traffic is a difficult problem because traffic management schemes attempt to achieve a wide variety of objectives, such as:

1. *Efficiency and minimal delay.* There is a tradeoff between the link utilization and the end-to-end delay. For low utilization, the queues at the switches are small and the delay is small. Once utilization is very high, the queues grow and cells are dropped when the queue size exceeds the available buffer size. The delay varies according to the load, but there is always a non-zero queuing delay. Figure 2 shows the throughput and delay with varying load in the network. The operating point which has a utilization close to 100% and moderate delays is called the *knee* of the delay-throughput curve. This is a good choice for an optimal operating point, and congestion control schemes which operate at this point are called *congestion avoidance schemes*. For a detailed description of congestion avoidance, the reader is referred to [15].
2. *Fairness.* The optimal operation of a distributed shared resource is given by criterion called the *max-min allocation* [9]. Mathematically, it is defined as follows. Given a configuration with  $n$  contending sources, suppose the  $i$ th source is allocated a bandwidth  $x_i$ . The allocation vector  $\{x_1, x_2, \dots, x_n\}$  is feasible if all link load levels are less than or equal to 100%. The total number of such feasible vectors is infinite. Given any allocation vector, the source that is getting the least allocation is, in some sense, the “unhappiest source”. We need to find the feasible vectors that give the maximum allocation to this unhappiest source. The number of such vectors is also infinite. Now

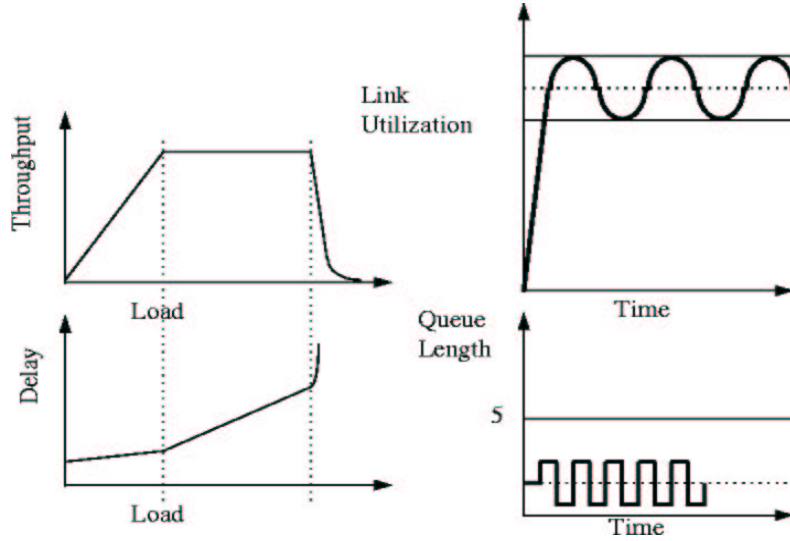


Figure 2: Throughput versus delay

we remove this “unhappiest source” and reduce the problem to that of the remaining  $n - 1$  sources operating on a network with reduced link capacities. Again, we find the unhappiest source among these  $n - 1$  sources, give that source the maximum allocation and reduce the problem by one source. We keep repeating this process until all sources have been allocated the maximum that they could get.

If  $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$  is the max-min allocation, then the fairness of a scheme that allocates  $\{x_1, x_2, \dots, x_n\}$  is given by:

$$\text{Fairness Index} = \frac{(\sum y_i)^2}{n \times \sum y_i^2}$$

where:  $y_i = \frac{\hat{x}_i}{x_i}$

3. *Good steady state as well as transient response characteristics.* Steady state characteristics can be tested using persistent sources which always have cells to send. Most schemes attempt to exhibit minimal oscillations during steady state conditions. In the real world, the transient response is almost as important as steady state performance. Jain and Routhier [14] point out that most real world traffic is bursty because most sources are transient. Transient response can be tested using transient sources which start after other sources have started and/or stop before the other sources have stopped. Good schemes must be able to respond rapidly to these load transients and achieve optimal performance in the steady state.
4. *Adaptation to the presence of multiple traffic classes.* Both VBR and CBR traffic are delay sensitive and have a higher priority than ABR traffic. When VBR and CBR sources are active, the available ABR capacity is reduced as follows:

$$\text{ABR Capacity} \leftarrow \text{Link Capacity} - \text{VBR Capacity} - \text{CBR Capacity}$$

Moreover, ABR capacity is not fixed, but varies according to the VBR and CBR load. Congestion control schemes must respond quickly to dynamic changes in the ABR capacity.

5. *Good performance for various source traffic patterns.* The real world traffic may not correspond to any single traffic model. In general, ABR congestion control schemes needs to perform optimally with high variance in the demand and different source traffic models. Congestion control schemes must adapt to persistent sources which always have data to send, as well as to bursty sources which alternate between active periods and idle periods. In addition, the schemes must behave optimally for bottlenecked sources which cannot utilize the full capacity they are allocated, and they should be able to rapidly react to the overload that can arise if the bottlenecked sources suddenly start fully utilizing their allocations.

We found increased complexity of the source models in the following sequence: persistent cell traffic, bursty cell traffic, source bottleneck, persistent TCP sources and bursty TCP sources. The problems become even more complicated when the ABR capacity is variable due to the presence of VBR background. The scheme should perform well for any combination of traffic conditions.

6. *Scalability to ATM networks that cover a wide range of speeds, distances, number of switches and number of VCs.* The same scheme should perform well for local area networks (LANs) as well as wide area networks (WANs). The time taken for feedback to reach the source from a switch clearly depends on the distance between them. The scheme must exhibit optimal behavior under widely varying conditions without need for parameter adjustment.

## 4 The ERICA Algorithm

The ERICA algorithm aims at achieving a fair and efficient allocation of the available bandwidth to contending sources. It requires monitoring the available capacity and the current demand on the resources. The ERICA algorithm is applied to each output port (or link). In this section, we first briefly describe the original ERICA algorithm, and then summarize its features, enhancements and options. For a complete description of the ERICA algorithm, including pseudocode and flowcharts, see [1].

### 4.1 The Original Algorithm

The switch periodically monitors the load on each link and determines a load factor,  $z$ , the available capacity, and the number of currently active VCs.

The load factor is calculated as the ratio of the measured ABR input rate at the port to the target ABR capacity of the output link.

$$LoadFactor(z) \leftarrow \frac{ABR\ Input\ Rate}{Target\ ABR\ Capacity}$$

where,

$$Target\ ABR\ Capacity \leftarrow Target\ Utilization \times Link\ Bandwidth$$

The Input Rate is measured over an interval called the switch averaging interval. The above steps are executed at the end of the switch averaging interval.

Target utilization is a parameter which is set to a fraction (close to, but less than 100 %) of the available capacity.

The fair share of each VC, *FairShare*, is also computed as follows:

$$FairShare \leftarrow \frac{ABR\ Capacity}{Number\ of\ Active\ Sources}$$

The switch allows each source sending at a rate below the *FairShare* to rise to *FairShare* every time it sends a feedback to the source. If the source does not use all of its *FairShare*, then the switch allocates the remaining capacity to the sources which can use it. For this purpose, the switch calculates the quantity *VCShare* as follows:

$$VCShare \leftarrow \frac{CCR}{z}$$

If all VCs change their rate to their *VCShare* values then, in the next cycle, the switch would experience unit overload ( $z$  equals one).

Hence, *VCShare* aims at bringing the system to an efficient operating point, which may not necessarily be fair, and *FairShare* allocation aims at achieving fairness, possibly leading to overload (inefficient operation). A combination of these two quantities is used to rapidly reach optimal operation as follows:

$$ER\ Calculated \leftarrow \text{Max} (FairShare, VCShare)$$

This is called the “maximum formula.”

The calculated ER value cannot be greater than the ABR Capacity which has been measured earlier. Hence, we have:

$$ER\ Calculated \leftarrow \text{Min} (ER\ Calculated, ABR\ Capacity)$$

To ensure that the bottleneck ER reaches the source, each switch computes the minimum of the ER it has calculated and the ER value in the RM cell. This value is indicated in the ER field of the RM cell as follows:

$$ER\ in\ RM\ Cell \leftarrow \text{Min} (ER\ in\ RM\ cell, ER\ Calculated)$$

## 4.2 Achieving Max-Min Fairness

To achieve max-min fairness, the basic ERICA algorithm is extended by remembering the highest allocation given during an averaging interval and ensuring that all sources are given this high allocation. We add a variable *MaxAllocPrevious* which stores the maximum allocation given in the previous interval.

For  $z > 1 + \delta$ , where  $\delta$  is a small fraction, we use the basic ERICA algorithm and allocate the source Max (FairShare, VCShare). But, for  $z \leq 1 + \delta$ , we attempt to make all the rate allocations equal, by giving the allocation Max (FairShare, VCShare, MaxAllocPrevious). This method equalizes the allocations of the VCs during underload.

## 4.3 Fairshare First to Avoid Transient Overloads

The inter-RM cell time is a factor in determining the transient response time when load conditions change. With the basic ERICA scheme, it is possible that a source which receives feedback first can keep getting rate increase indications, purely because it sends more RM cells before competing sources can receive feedback. This results in unnecessary spikes (sudden increases) in rates and queues with the basic ERICA scheme.

This problem can be alleviated by incorporating the following change to the ERICA algorithm. When the calculated ER is greater than the fair share value, and the source is increasing from a CCR below *FairShare*, we limit its increase to *FairShare*.

## 4.4 Other ERICA Features and Options

This subsection highlights some of the features and options of the ERICA algorithm. The effect of these options will be discussed when the performance of ERICA is examined in the next section. The options and features of ERICA that are most interesting can be summarized as follows:

- **ABR Operation with VBR and CBR in the Background.**

ATM links will be used by constant bit rate (CBR) and variable bit rate (VBR) traffic along with ABR traffic. CBR and VBR have a higher priority. Only the capacity left unused by VBR and CBR is allocated to ABR sources. Hence, we need to measure the CBR and VBR usage along with the input rate. The ABR capacity is then calculated as follows:

$$\text{Target ABR Capacity} \leftarrow \text{Target Utilization} \times \text{Link Bandwidth} - \text{VBR Usage} - \text{CBR Usage}$$

This formula is consistent with one of the goals of ERICA, which is keeping the link utilization close to the target utilization. The ABR capacity is ensured to be nonzero through scheduling.



- **Bi-directional Counting of Bursty Sources.**

A bursty source sends data in bursts during its active periods, and remains idle during other periods. It is possible that the BRM cell of a bursty source is traveling in the reverse direction, but no cells of this source are traveling in the forward direction. An enhancement to the counting algorithm is to also count a source as active whenever a BRM of this source is encountered in the reverse direction. We refer to this as the “bidirectional counting of active VCs”.

- **Averaging of the Number of Sources.**

Another technique to overcome problems of underestimating the number of active sources is to use averaging to decay the contribution of each VC to the number of active sources count. The main motivation behind this idea is that if a source is inactive during the current interval, but was recently active, it should still contribute to the number of active sources. This is because this source might be sending its data in bursts, and simply happened to be idle during the current interval.

- **Per-VC CCR Measurement Option.**

Inaccurate CCR estimates can result in inaccurate ER values because the CCR value is used in the calculation of the *VCShare* quantity. A solution to this problem is to measure the CCR of every VC during the same averaging interval as the load factor. This requires the switch to ignore the CCR field of the RM cell, count the number of cells received per VC during every averaging interval and update the estimate.

- **The ERICA+ Algorithm.**

ERICA+ is a further modification of ERICA that involves the following enhancements:

1. The link utilization is no longer targeted at a constant *Target Utilization* as in ERICA. The *Target Utilization* is set to one, and the total ABR capacity is measured given the link bandwidth and the CBR and VBR usage in that interval.

$$Total\ ABR\ Capacity = Link\ Bandwidth - VBR\ Usage - CBR\ Usage$$

2. The target ABR capacity is a fraction of the total ABR capacity and this fraction is a function of the queueing delay  $T_q$  at the switch.

$$Target\ ABR\ Capacity \leftarrow f(T_q) \times Total\ ABR\ Capacity$$

A sample queue control function is shown in figure 3. The function has four parameters:  $a$  and  $b$  (the intercepts of the two hyperbolas in the function), the target delay  $T_0$ , and the queue drain limit factor (*QDLF*).

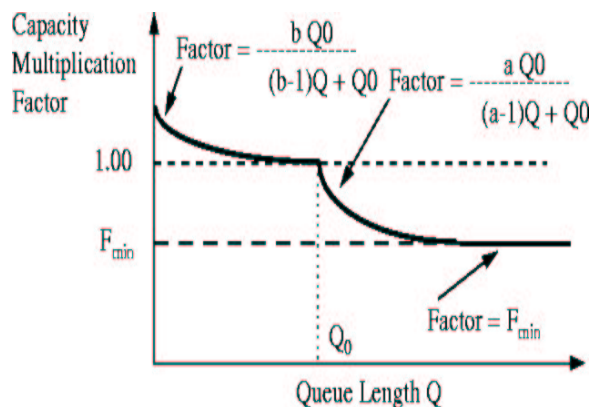


Figure 3: A sample queue control function in ERICA+

## 5 Performance Evaluation of ERICA and ERICA+

In order to assert that ERICA meets the goals of ABR congestion control schemes previously outlined in section 3, ERICA has been tested for a variety of networking configurations using several performance metrics. Its performance in the presence of various background traffic patterns and various source models has also been examined. We present simulation results for several configurations, which have been specifically selected to demonstrate particular aspects of the scheme. We prefer to use simple configurations when applicable because they are more instructive in finding problems. The results are presented in the form of four graphs for each configuration:

1. Graph of allowed cell rate (ACR) in Mbps over time for each source
2. Graph of ABR queue lengths in cells over time at each switch
3. Graph of link utilization (as a percentage) over time for each link
4. Graph of number of cells received at the destination over time for each destination

Our test strategy will attempt to examine whether ERICA meets each of the objectives mentioned in section 3. We will examine the efficiency and delay requirements, the fairness of the scheme, its transient and steady state performance, and finally its adaptation to variable capacity and various source traffic models. The experiments will also be selected such that they have varying distances and number of connections to examine the scalability requirement.

### 5.1 Parameter Settings

Throughout our experiments, the following parameter values are used:

1. All links have a bandwidth of 155.52 Mbps.

2. All LAN links are 1 Km long and all WAN links are 1000 Km long.
3. All VCs are bidirectional.
4. The source parameter Rate Increase Factor (RIF) is set to one, to allow immediate use of the full Explicit Rate indicated in the returning RM cells at the source.
5. The source parameter Transient Buffer Exposure (TBE) is set to large values to prevent rate decreases due to the triggering of the source open-loop congestion control mechanism. This was done to isolate the rate reductions due to the switch congestion control from the rate reductions due to TBE.
6. The switch target utilization parameter was set at 95% for LAN simulations and at 90% for WAN simulations.
7. The switch averaging interval was set to the minimum of the time to receive 50 cells and 1 ms for LAN simulations, and to the minimum of the time to receive 100 cells and 1 ms for WAN simulations.
8. The ERICA+ parameters are set as follows. The parameters  $a$  and  $b$  (intercepts of the two hyperbolas in the queueing delay function) are set to 1.15 and 1.05 respectively. The target delay  $T_0$  is set at 100 for LAN simulations and 500 for WAN simulations, and the queue drain limit factor ( $QDLF$ ) is set at 0.5.
9. All sources, including VBR sources are deterministic, i.e., their start/stop times and their transmission rates are known. The bursty traffic sources send data in bursts, where each burst starts after a request has been received from the client.

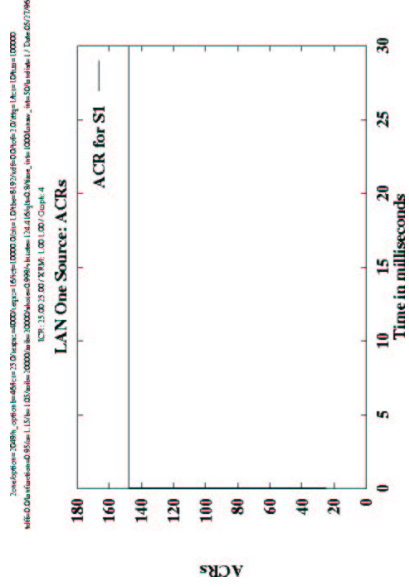
## 5.2 Efficiency



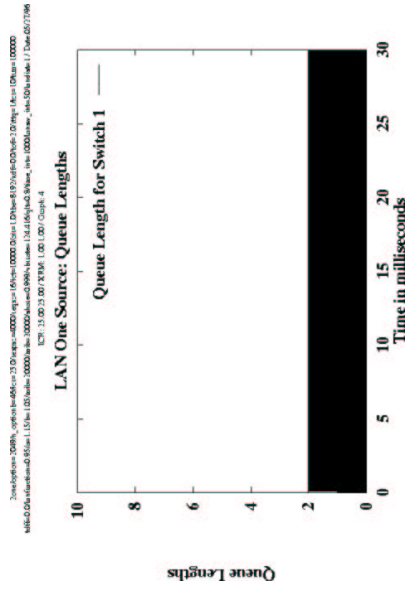
Figure 4: One source configuration

The very first test to verify efficient operation is to use a single source configuration as shown in figure 4. Any scheme that does not work for this simple configuration is not worth further analysis. The source is active over the entire simulation period. Figure 5 illustrates that ERICA achieves the required efficiency, since the source rate rises to almost fully utilize the link. Observe that there are no rate oscillations in the steady state, and that utilization is at the target utilization goal (95%).

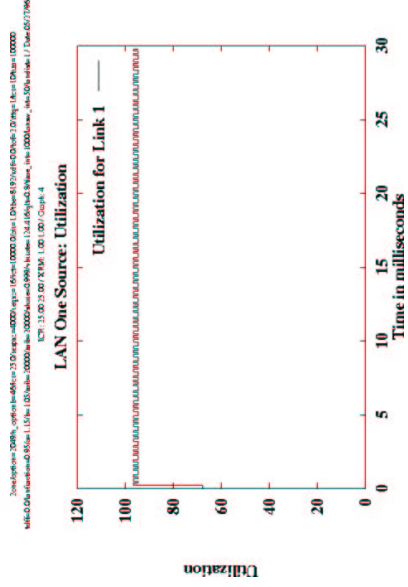
The same configuration has also been simulated to examine the efficiency of ERICA+. As seen in figure 6, the source rate rises to fully utilize the link (100% utilization) with no oscillations and minimal queues.



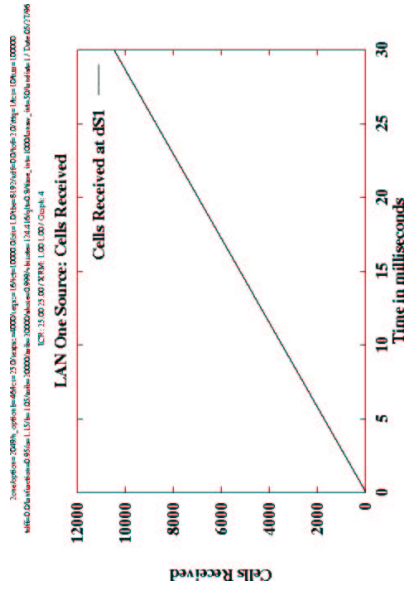
(a) Allowed cell rate



(b) Queue length

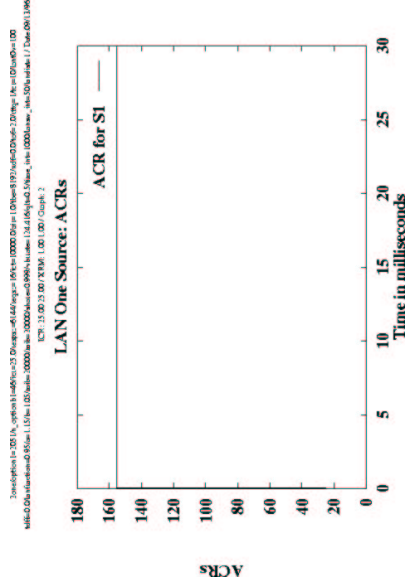


(c) Link utilization

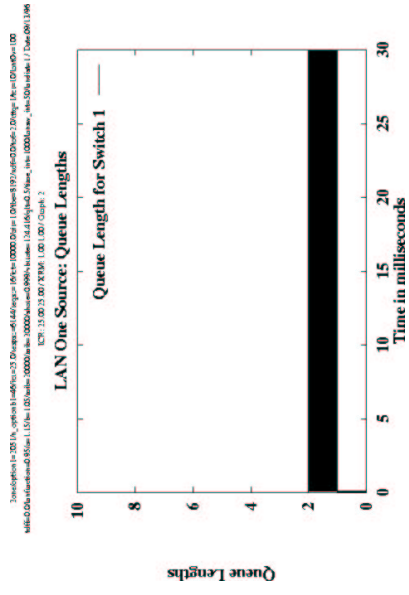


(d) Cells received

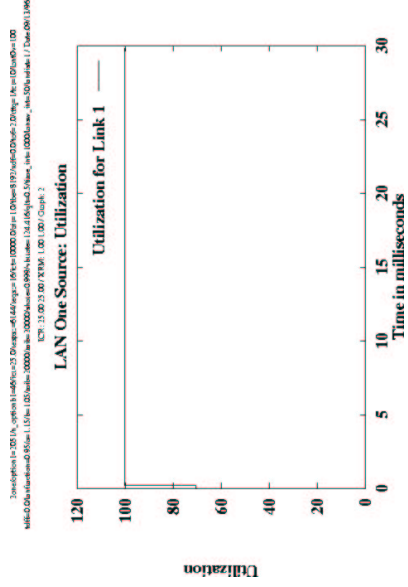
Figure 5: Results for a one source configuration in a LAN (ERICA)



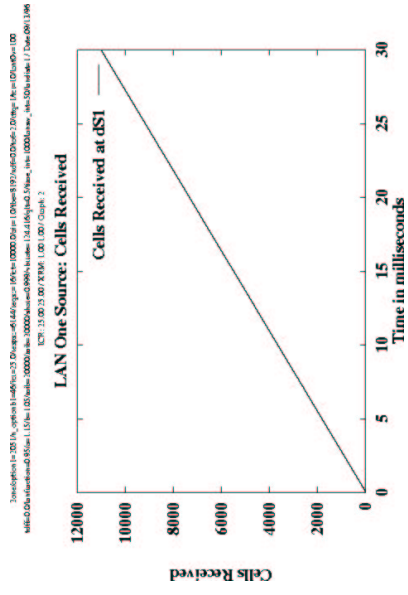
(a) Allowed cell rate



(b) Queue length



(c) Link utilization



(d) Cells received

Figure 6: Results for a one source configuration in a LAN (ERICA+)

### 5.3 Minimal Delay and Queue Lengths

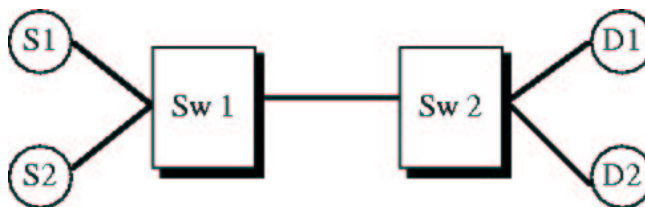


Figure 7: Two source configuration

To test for minimal delays and short queue lengths, we use a multiple source configuration. The simplest such configuration is the two source configuration, where two sources share a link as illustrated in figure 7. Each source must converge to almost half of the link rate ( $1/2 \times \text{Target Utilization}$ ), which is the max-min optimal allocation.

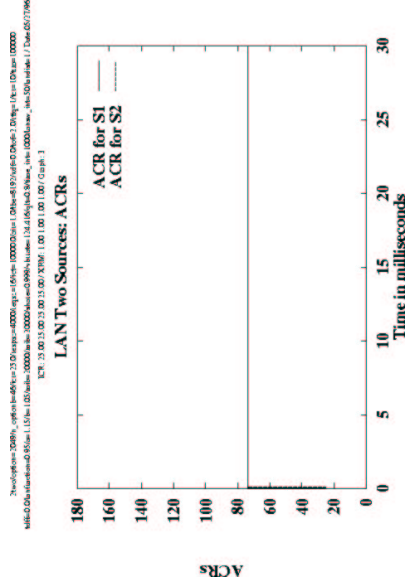
Figure 8 shows that the convergence is fast, the queue lengths are small (hence, the delay is minimal) and steady state performance is good. For ERICA+, the two sources rapidly converge to their optimal rates as seen in figure 9, and the queue length rises to reach the limit corresponding to its target delay parameter (100 microseconds corresponds to approximately 30 cells at 155.52 Mbps). There is a slight rate oscillation seen in figure 9(a) to allow the queues to reach the target value, but the steady state has no rate oscillations and 100% link utilization (figure 9(c)).

### 5.4 Fairness

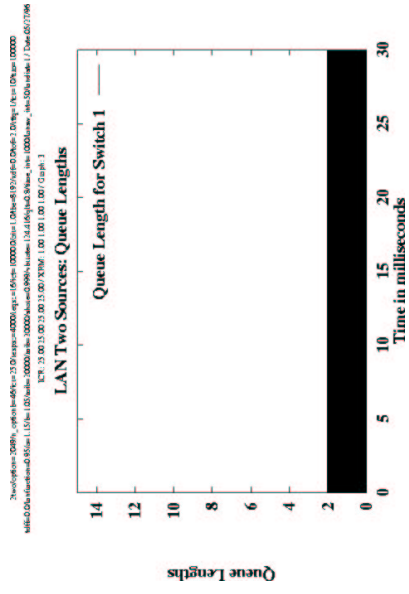
Two configurations are used for studying the fairness of the scheme: the parking lot configuration and the upstream configuration. The parking lot configuration and its name were derived from theatre parking lots, which consist of several parking areas connected via a single exit path. At the end of the show, congestion occurs as cars exiting from each parking area try to join the main exit stream.

For computer networks, an  $n$ -stage parking lot configuration consists of  $n$  switches connected in series. There are  $n$  VCs. The first VC starts from the first switch and goes through all the remaining switches. For the remaining VCs, the  $i^{\text{th}}$  VC starts at the  $i - 1^{\text{st}}$  switch. The link between the last two switches is the bottleneck link. The max-min allocation for each VC is  $1/n$  of the bandwidth. A 3-switch parking lot configuration is shown in figure 10. Figure 11 illustrates that ERICA achieves the desired max-min allocation, and figure 12 shows that ERICA+ also satisfied the max-min fairness criterion. The second link is fully utilized (figure 12(c)) and all rate allocations are equal (figure 12(a)).

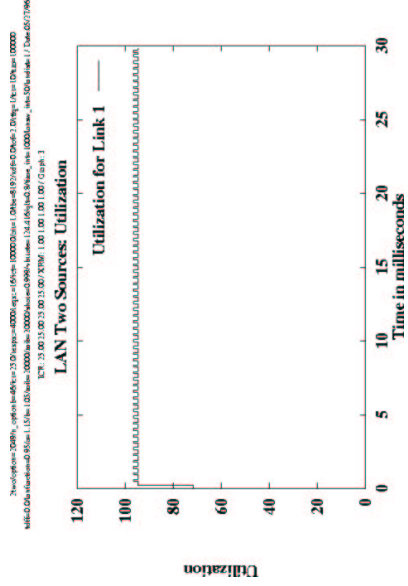
Although the parking lot configuration had been believed to test fairness, we discovered that it is not sufficient to demonstrate max-min fairness, and a more stringent test is required. We had observed that the original ERICA algorithm does not converge to max-min fairness in certain situations. Such situations arise when the ERICA algorithm is executed in a state where some of the sources cannot fully utilize their allocated bandwidth on a link



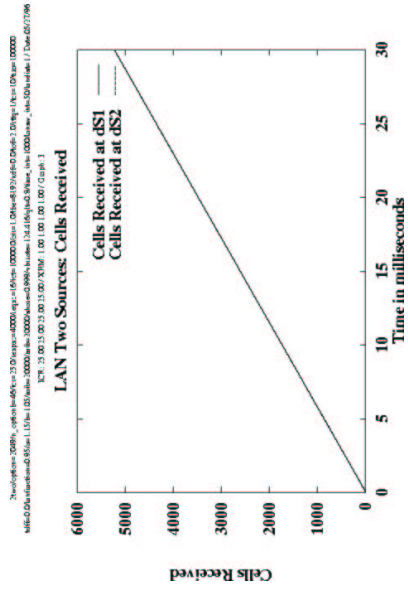
(a) Allowed cell rate



(b) Queue length

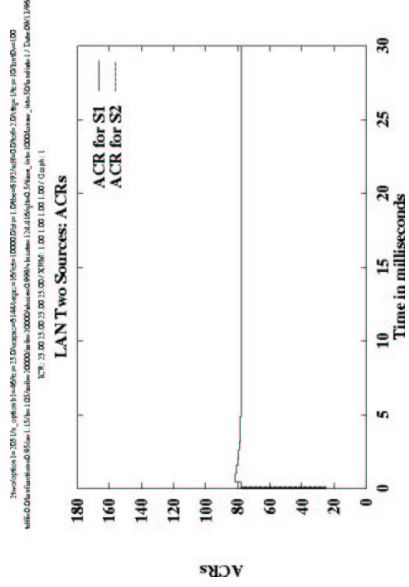


(c) Link utilization

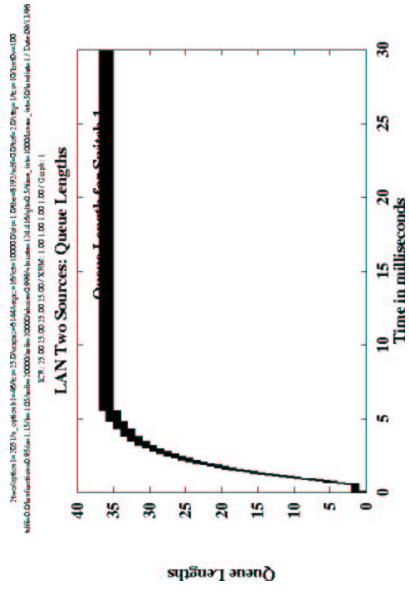


(d) Cells received

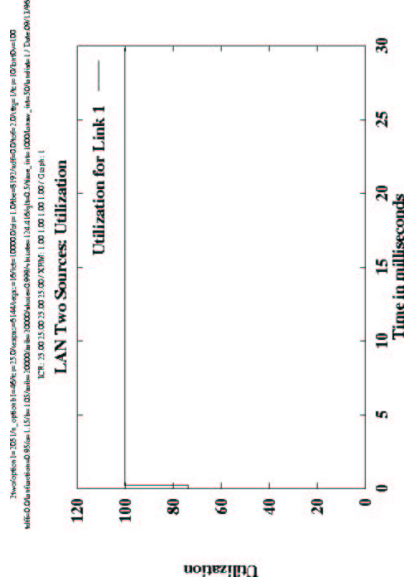
Figure 8: Results for a two sources configuration in a LAN (ERICA)



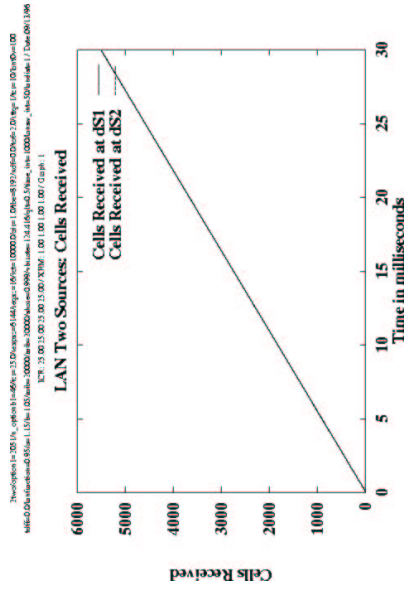
(a) Allowed cell rate



(b) Queue length



(c) Link utilization



(d) Cells received

Figure 9: Results for a two sources configuration in a LAN (ERICA+)



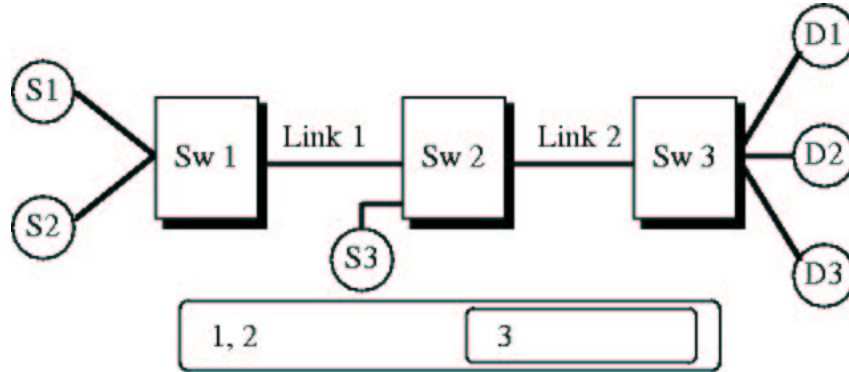


Figure 10: Parking lot configuration

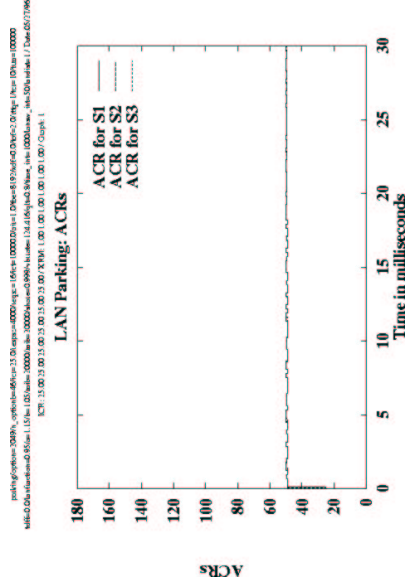
(for example, because they are bottlenecked on another link), and the rest of the sources contending for bandwidth have unequal CCR values, which are greater than the fair share value (first term in the maximum formula). The ERICA algorithm does not converge to max-min fairness in these situations because, after  $z$  converges to one, the second term in the maximum formula becomes  $CCR_i/1 = CCR_i$ , and the first term is constant. The maximum of the two terms for the contending sources is the second term, because there are sources that are not fully utilizing their allocated bandwidth. Hence, the sources do not change their rates.

An example of this situation can be illustrated by an upstream configuration (see figure 13). The upstream configuration consists of three switches and 17 VCs. The second link is shared by  $VC_{15}$ ,  $VC_{16}$ , and  $VC_{17}$ . Because there are 15 VCs on the first link,  $VC_{15}$  is limited to a throughput of less than  $1/15$  the link rate.  $VC_{16}$  and  $VC_{17}$  should, therefore, each converge to a little less than  $7/15$  of the second link rate. The configuration is called an upstream configuration because the bottleneck link is the first link (upstream link). A WAN configuration (1000 Km links) is used in this situation to illustrate the scalability of ERICA to long distances.

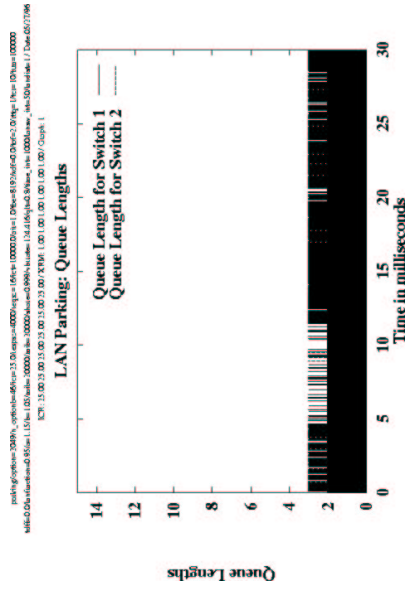
Figure 14 shows that the original ERICA algorithm was unfair in this situation, and figure 15 shows that ERICA, after the modification discussed in section 4.2, is fair. As seen in figure 15, the modified algorithm converges to max-min allocations. Regardless of the initial load factor value, after a short transient period, all sources contending for bandwidth are allocated equal rates, and the two curves in figure 15(b) (number of cells received at the destination) have the same slope (compare with figure 14(b)). The transient response is slightly worse than the original ERICA algorithm due to the temporary over-allocation needed to equalize the shares, but the steady state performance is as good as with the original ERICA algorithm.

## 5.5 Transient and Steady State Performance

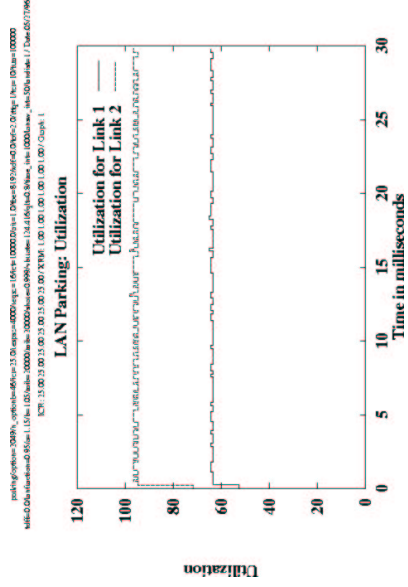
To test the transient response of the system, we use a modified two source configuration. The configuration is similar to the two source configuration because two sources share the same link, but one of the sources is only active from 10 ms to 20 ms while the other source is active



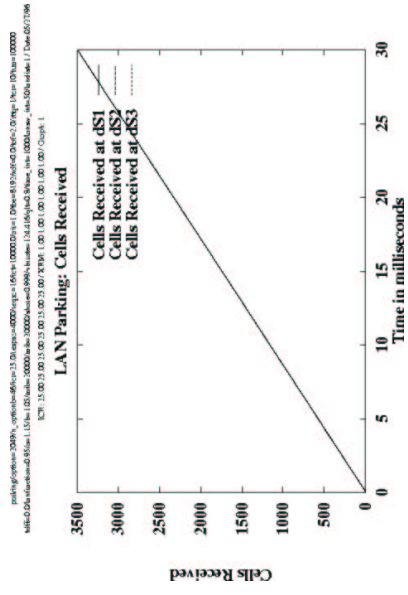
(a) Allowed cell rate



(b) Queue length

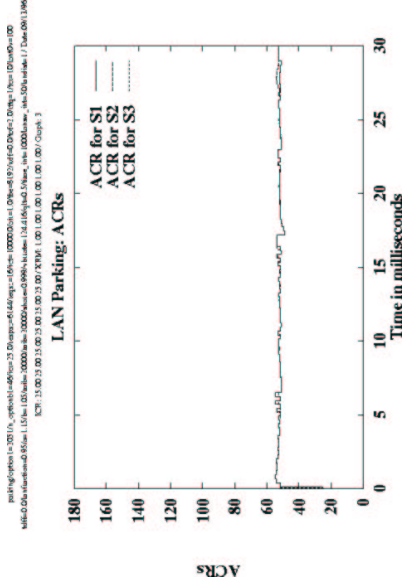


(c) Link utilization

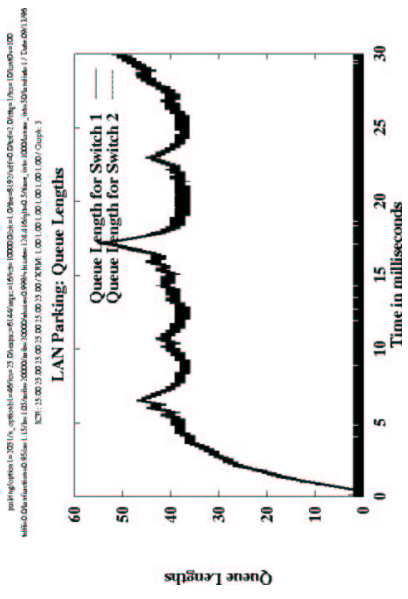


(d) Cells received

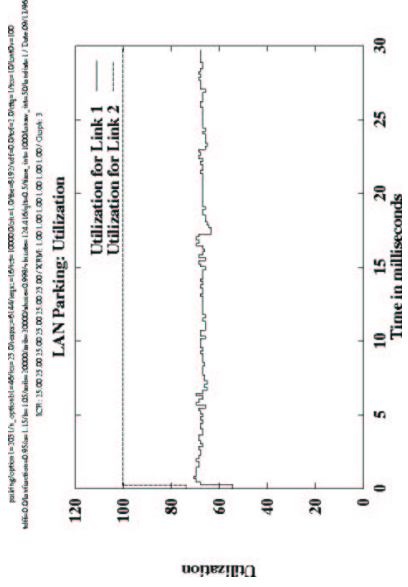
Figure 11: Results for a parking lot configuration in a LAN (ERICA)



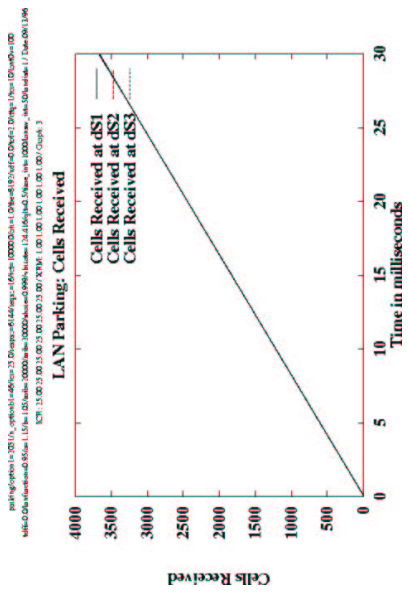
(a) Allowed cell rate



(b) Queue length



(c) Link utilization



(d) Cells received

Figure 12: Results for a parking lot configuration in a LAN (ERICA+)

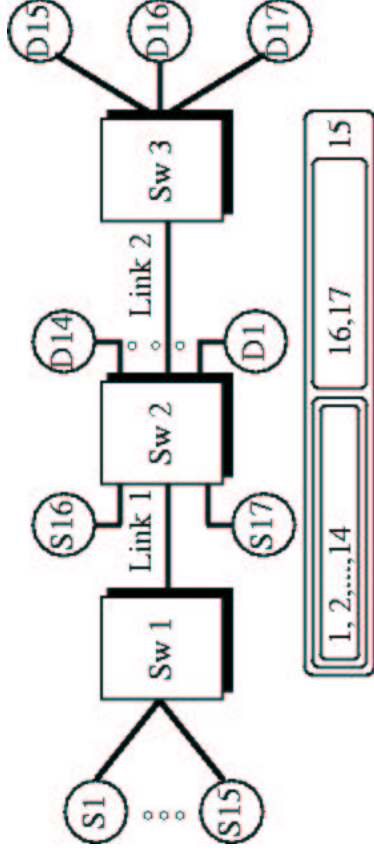
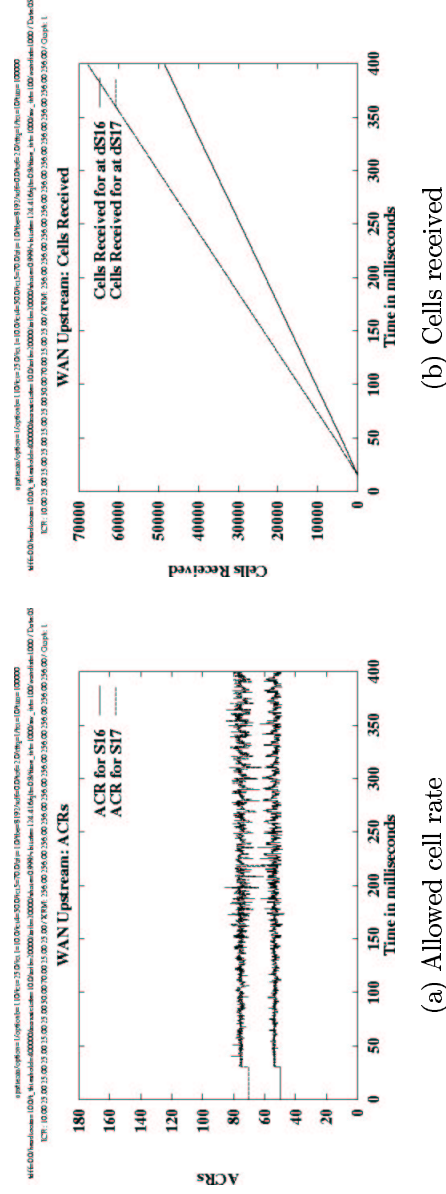


Figure 13: Upstream Configuration



(a) Allowed cell rate

(b) Cells received

Figure 14: Results for an upstream configuration in a WAN (ERICA without the max-min fairness enhancement)

throughout. Besides illustrating the transient response of the system, this configuration also illustrates the effect of the “fairshare first” algorithm discussed in section 4.3. That algorithm (see section 4.3) prevents a low rate VC to rise above *FairShare*. This VC takes an extra round trip compared to the basic ERICA because it first comes to *FairShare* before rising further. The switch can use the extra round trip to give feedback to all the sources, measure a new load factor and reduce overloading sources. The modification reduces the maximum queues in transient situations.

Figures 16 and 17 illustrate the effect of the “fairshare first” modification on a transient configuration in a LAN. Figure 16 shows the transient performance of ERICA without the “fairshare first” modification, and figure 17 illustrates how ERICA with the “fairshare first” modification avoids transient overloads. It is clear that ERICA exhibits good transient response characteristics to changing load, and the modification mitigates sudden overloads, constraining the queue length when the second source starts transmission. The figure also shows that the steady state performance of the scheme is excellent, as there are minimal

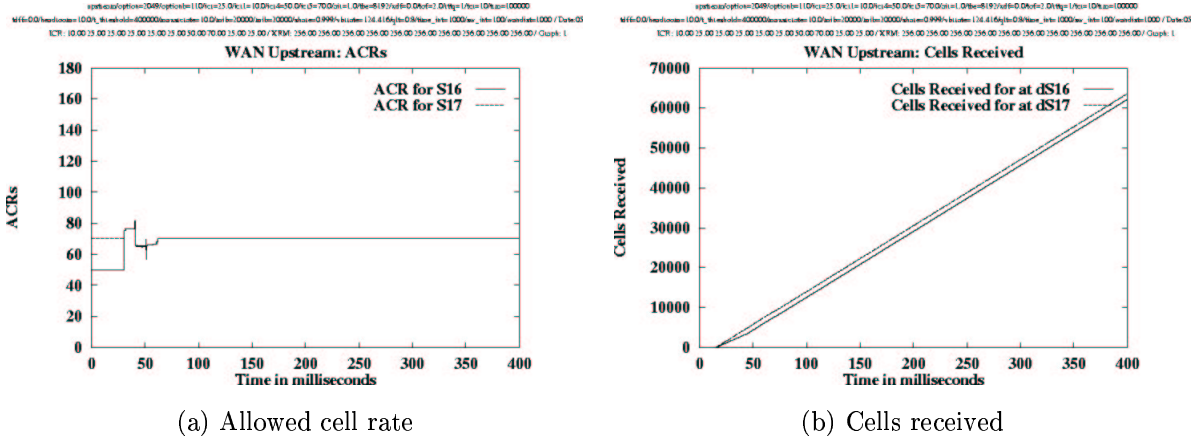


Figure 15: Results for an upstream configuration in a WAN (ERICA with the max-min fairness enhancement)

oscillations in the rates of the sources.

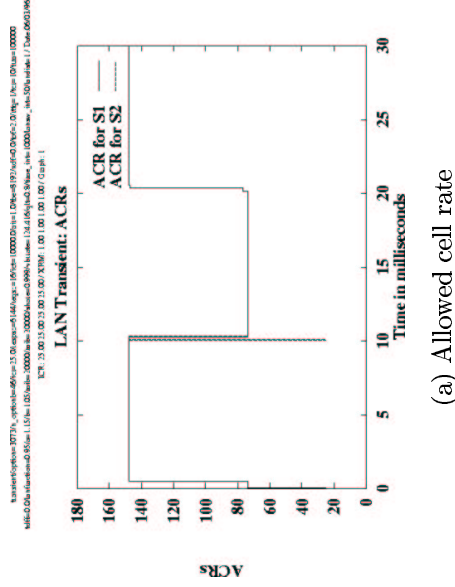
## 5.6 Adaptation to the Presence of Multiple Traffic Classes (Variable Capacity)

Constant Bit Rate (CBR) and Variable Bit Rate (VBR) service classes have a higher priority than the ABR service. In cases of VBR traffic, the ABR capacity becomes a variable quantity.

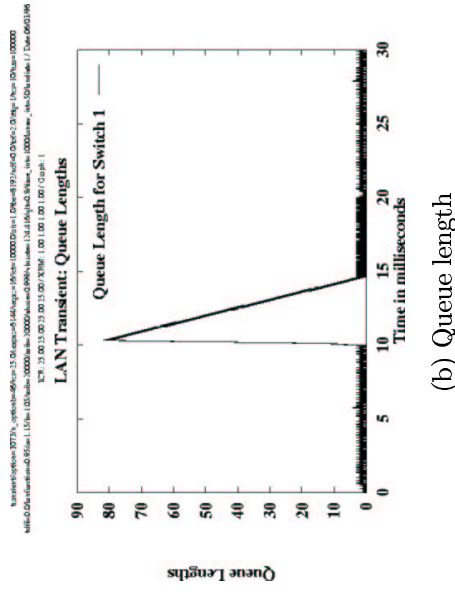
The two source configuration in a wide area network is used to demonstrate the behavior of ERICA in the presence of VBR sources. A deterministic VBR source is used whose peak rate is 124.42 Mbps (80% of the link capacity). Figure 18 illustrates the behavior of ERICA on a WAN where the VBR source was active for alternating periods of 1 ms with 1 ms inactive periods in between (high frequency VBR), while figure 19 shows the performance with VBR on/off periods of 20 ms (low frequency VBR).

From the figures, it is clear that ERICA rapidly detects the change in the available ABR capacity and gives the appropriate feedback to the sources. When the VBR source is active, the ABR sources rapidly reduce their rates (figures 18(a) and 19(a)). The utilization is generally high; the utilization drops reflect the time taken for the feedback to reach the sources: the feedback delay (figures 18(c) and 19(c)). The spikes in the queue lengths seen in figures 18(b) and 19(b) also reflect the feedback delay, but the queues are rapidly drained. Observe that the number of cells received in both cases (figures 18(d) and 19(d)) is approximately equal, which shows that the performance is approximately the same.

Figure 20 illustrates how ERICA+ adapts to high frequency VBR in the background, and figure 21 shows its performance with low frequency VBR. ERICA+ adapts rapidly to the changing background traffic, recomputing the available bandwidth and the rate allocations. The link utilization is higher than that with ERICA, and the queue lengths are constrained.

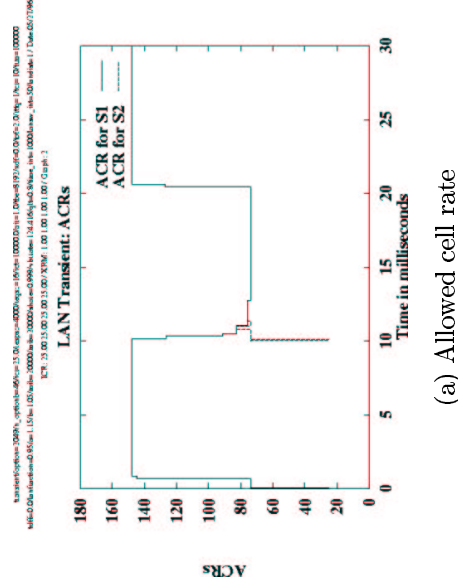


(a) Allowed cell rate

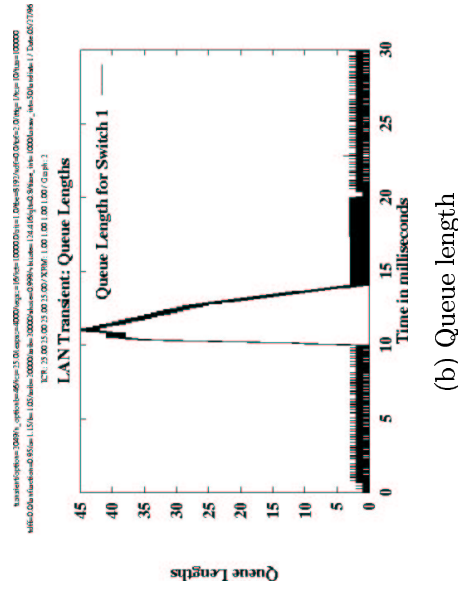


(b) Queue length

Figure 16: Results for a transient source configuration in a LAN (ERICA without the “fair-share first” enhancement)

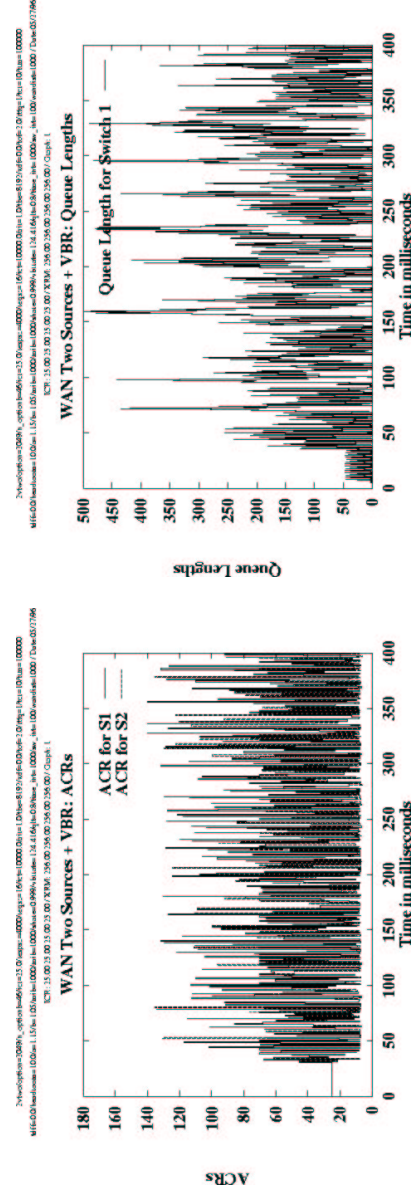


(a) Allowed cell rate



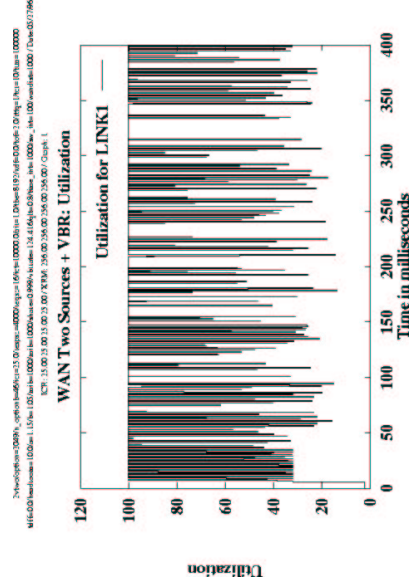
(b) Queue length

Figure 17: Results for a transient source configuration in a LAN (ERICA with the “fairshare first” enhancement)

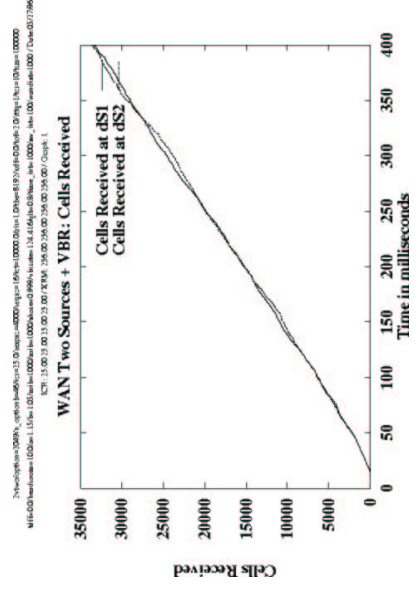


(a) Allowed cell rate

(b) Queue length

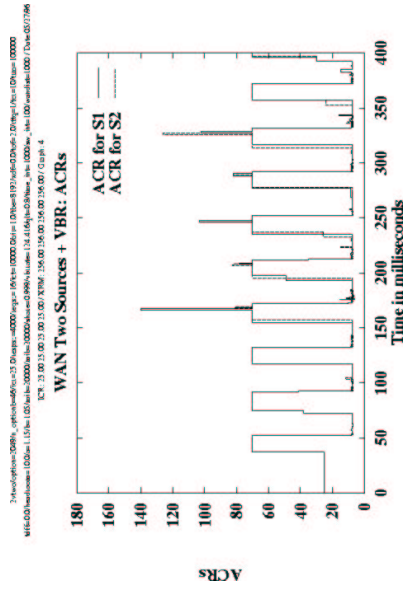


(c) Link utilization

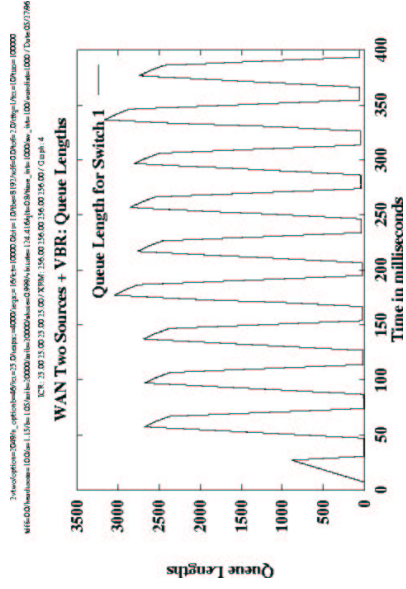


(d) Cells received

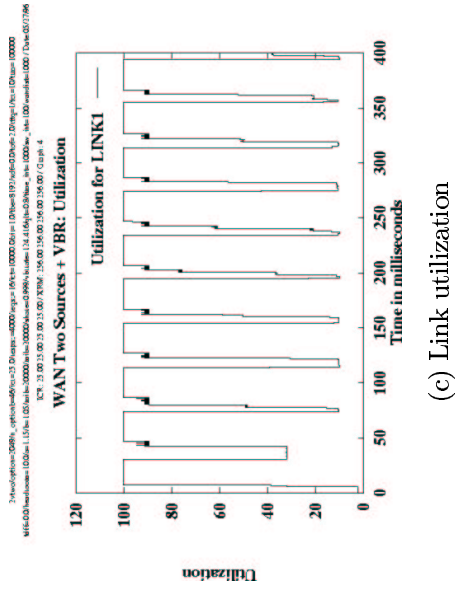
Figure 18: Results for a two source configuration with (1 ms on/1 ms off) VBR background in a WAN (ERICA)



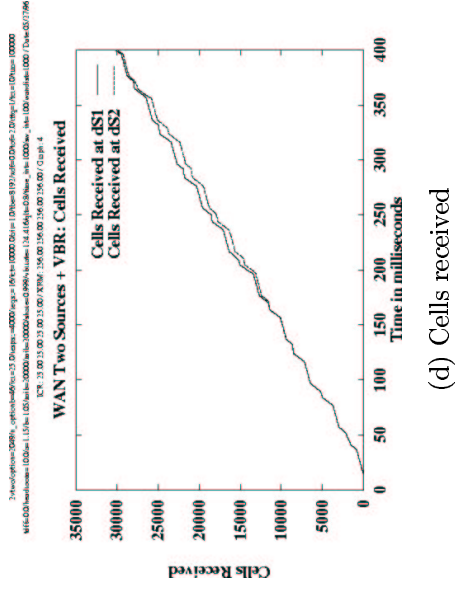
(a) Allowed cell rate



(b) Queue length



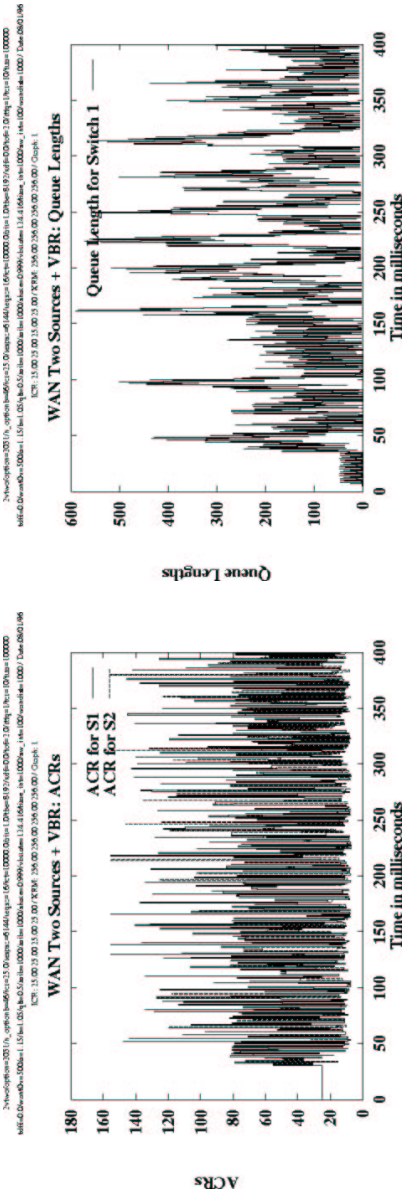
(c) Link utilization



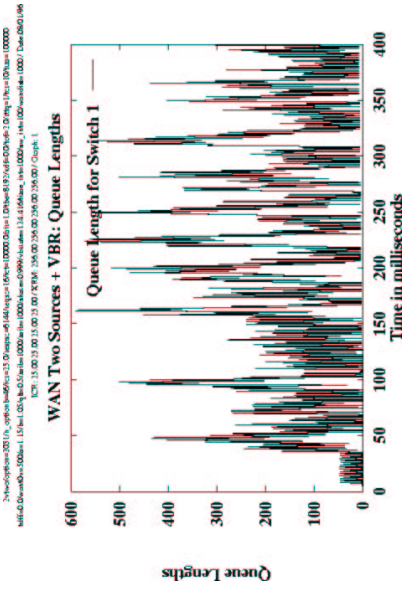
(d) Cells received

Figure 19: Results for a two source configuration with (20 ms on/20 ms off) VBR background in a WAN (ERICA)

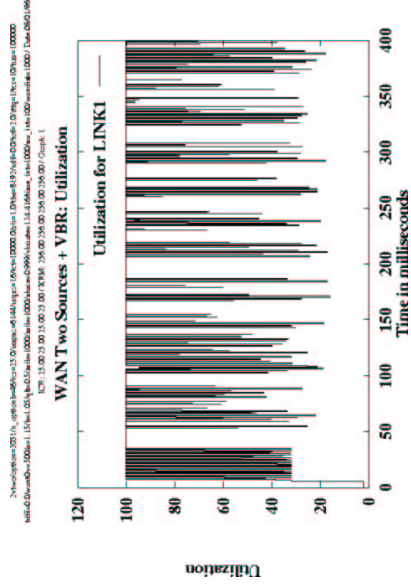




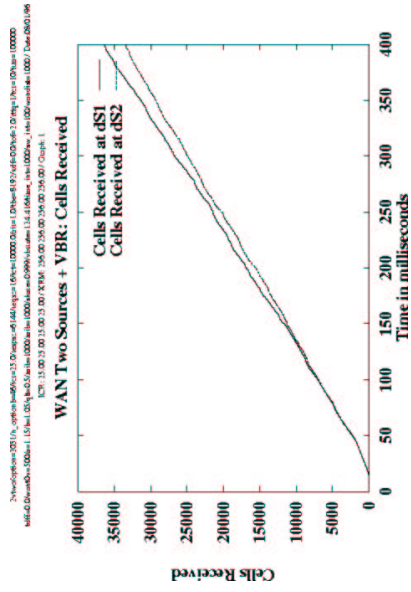
(a) Allowed cell rate



(b) Queue length



(c) Link utilization



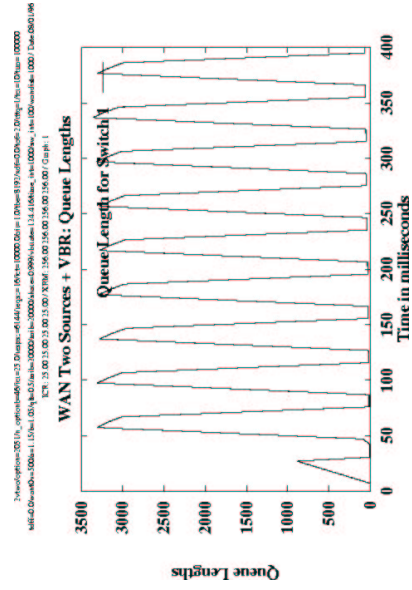
(d) Cells received

Figure 20: Results for a two source configuration with (1 ms on/1 ms off) VBR background in a WAN (ERICA+)

The target queue goal is never reached due to the high variance, but the utilization goal is partially reached.

## 5.7 Adaptation to Various Source Traffic Models

In all the previous experiments, the ABR sources are assumed to be persistent sources, which means that they always have data to send, and can utilize their full capacity at all times. It is essential to examine the performance of the scheme with bursty sources which alternate between active periods when they utilize their full capacity, and idle periods when they do not send any data. In addition, situations where sources are bottlenecked are also of particular interest. The scheme should be able to rapidly react to the overload that can arise if the bottlenecked sources suddenly start utilizing their full capacity.



(a) Allowed cell rate

(b) Queue length

(c) Link utilization

(d) Cells received

Figure 21: Results for a two source configuration with (20 ms on/20 ms off) VBR background in a WAN (ERICA+)

### 5.7.1 Bursty Traffic

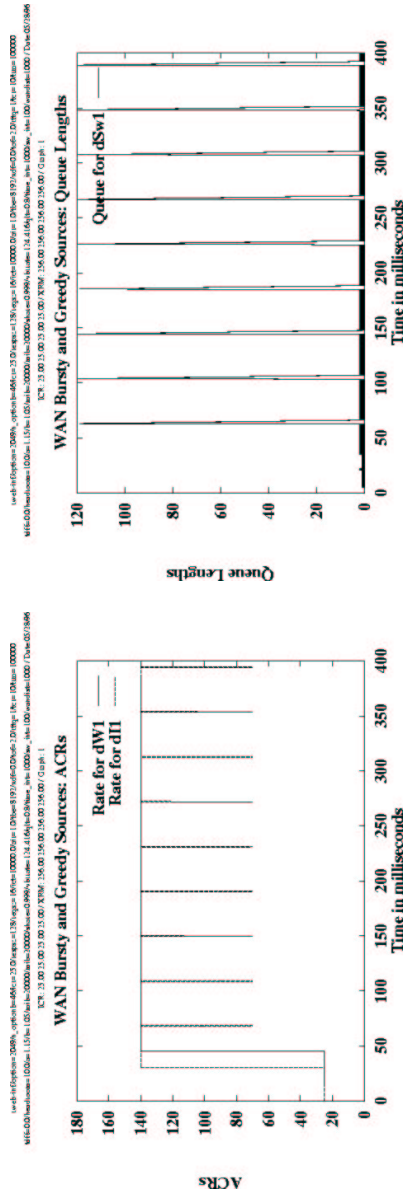
Figures 22 through 24 illustrate the performance of ERICA in a wide area network two source configuration where one of sources is a persistent (greedy or infinite) source, while the other connection is a request-response type connection. The request-response connection consists of a source sending a request (of size 16 cells), and the destination responding with a burst of data. Three different burst sizes are used in our simulations: small bursts are 128 cells, medium bursts are 1024 cells and large bursts are 6144 cells. Upon the receipt of the response at the source, and after a certain period of time, the source sends another request for data, and the cycle is repeated [12]. The figures show the performance of the reverse (response) connection where a burst of data is sent in response to every request.

Figure 22 illustrates the performance of ERICA with small response burst sizes, figure 23 shows the effect of medium burst sizes and figure 24 illustrates the effect of large burst sizes. As seen in the figures, ERICA can adapt to small and medium bursts of data, and the queue lengths are constrained. However, with a target utilization of 90%, ERICA does not have enough capacity to drain large bursts of data from the switch queues before the next burst is received. This problem can be solved by using smaller values for the target utilization parameter.

Figure 25 shows that bidirectional counting of the number of active sources (as discussed in section 4.4) limits the queue sizes for large bursts. This is because it counts the bursty source as active if its RM cells are traveling in the reverse direction, even though it might not be sending any data in the forward direction during its idle periods. This situation is called the “out-of-phase” effect, and is also a common problem with TCP sources. The problem affects the load measurement, as well as the measurement of the number of active VCs. As seen in figure 25(b), the queue lengths are constrained, and the problem seen in figure 24(b) has been solved, even for a target utilization of 90%.

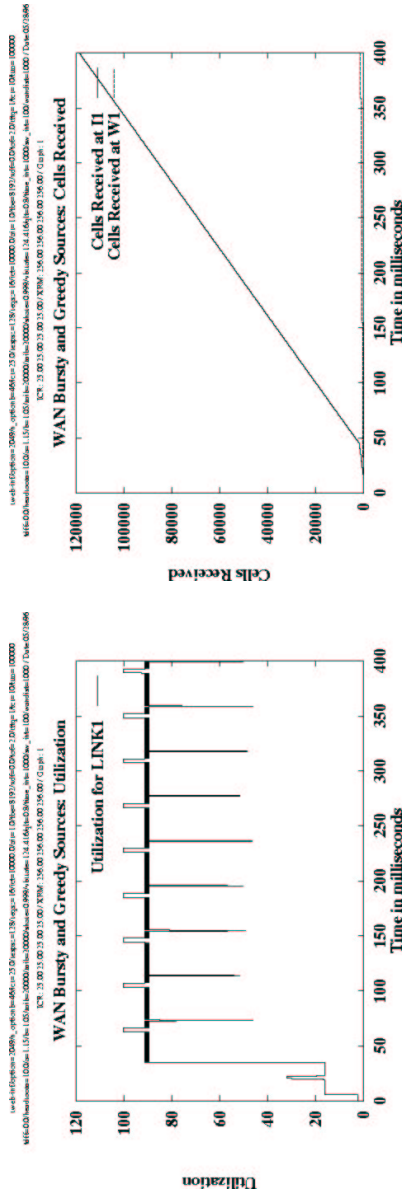
Another method to limit the queue sizes in this case is by averaging the number of active sources as discussed in section 4.4. As previously explained, we should account for the presence of a source, even though it might be currently idle. The effect of averaging the value of the number of active sources is illustrated in figure 26. The bidirectional counting option is not used in this case. Figure 26(b) shows that the queue length is constrained.

Figures 27, 28 and 29 illustrate the performance of ERICA+ for the same configuration without the averaging the number of sources option. Figure 27 illustrates the adaptability of ERICA+ to small burst sizes. Here, the target queue delay is not reached during the simulation time since the burst sizes are small. Figure 28 shows the effect of medium burst sizes and figure 29 illustrates the effect of large burst sizes. It is clear that ERICA+ can adapt to bursty traffic better than ERICA, because it accounts for the time to drain the queues when estimating the available capacity. Even with large burst sizes, the queues built up when the bursty source is active can drain before the next burst arrives at the switch. The bidirectional counting and the averaging of number of sources options are not necessary in this case.



(a) Allowed cell rate

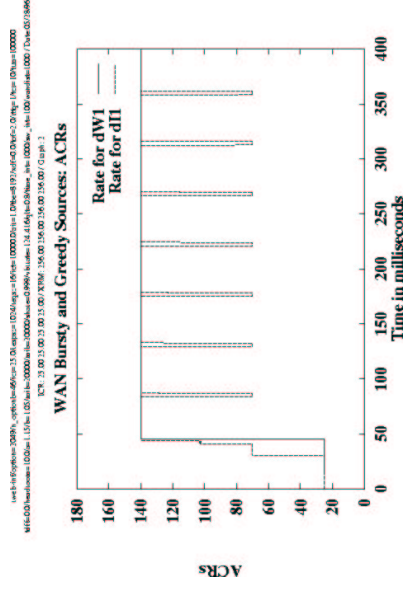
(b) Queue length



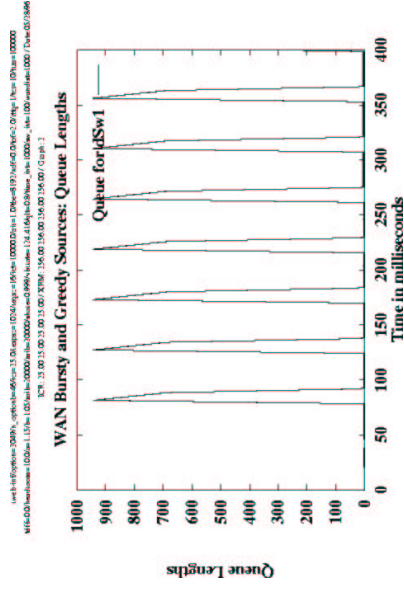
(c) Link utilization

(d) Cells received

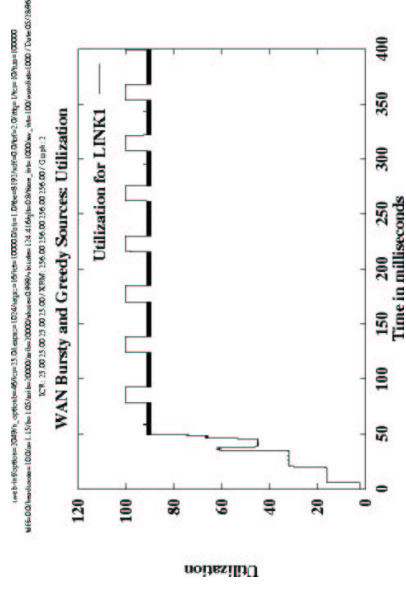
Figure 22: Results for one persistent source and one bursty source (small bursts) in a WAN (ERICA)



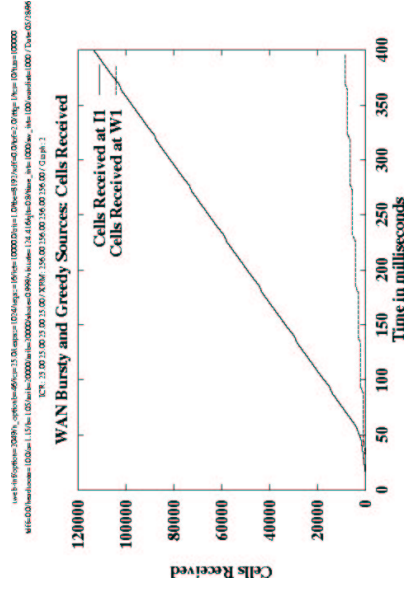
(a) Allowed cell rate



(b) Queue length

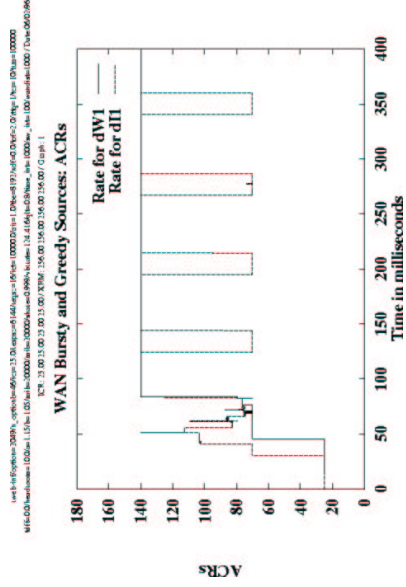


(c) Link utilization

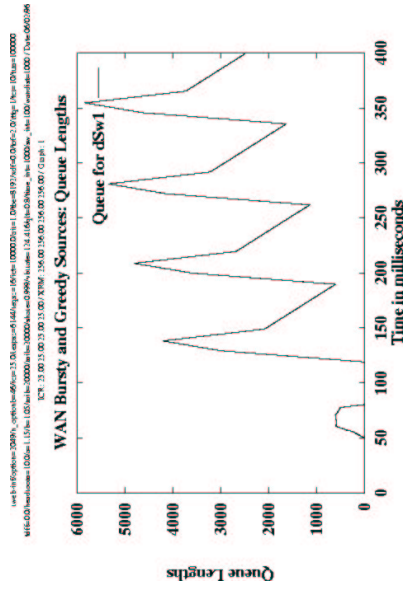


(d) Cells received

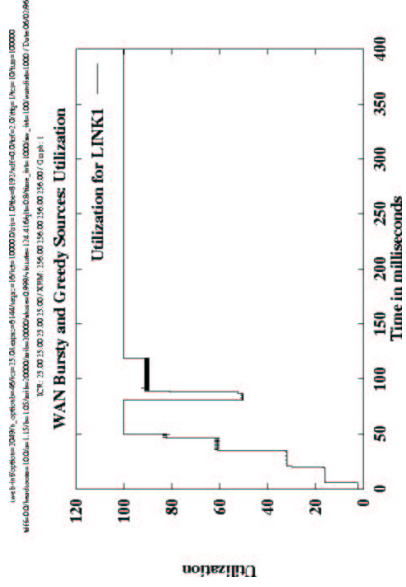
Figure 23: Results for one persistent source and one bursty source (medium bursts) in a WAN (ERICA)



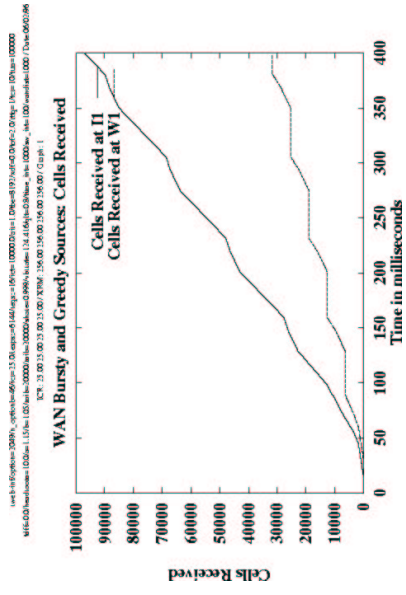
(a) Allowed cell rate



(b) Queue length



(c) Link utilization



(d) Cells received

Figure 24: Results for one persistent source and one bursty source (large bursts) in a WAN (ERICA)

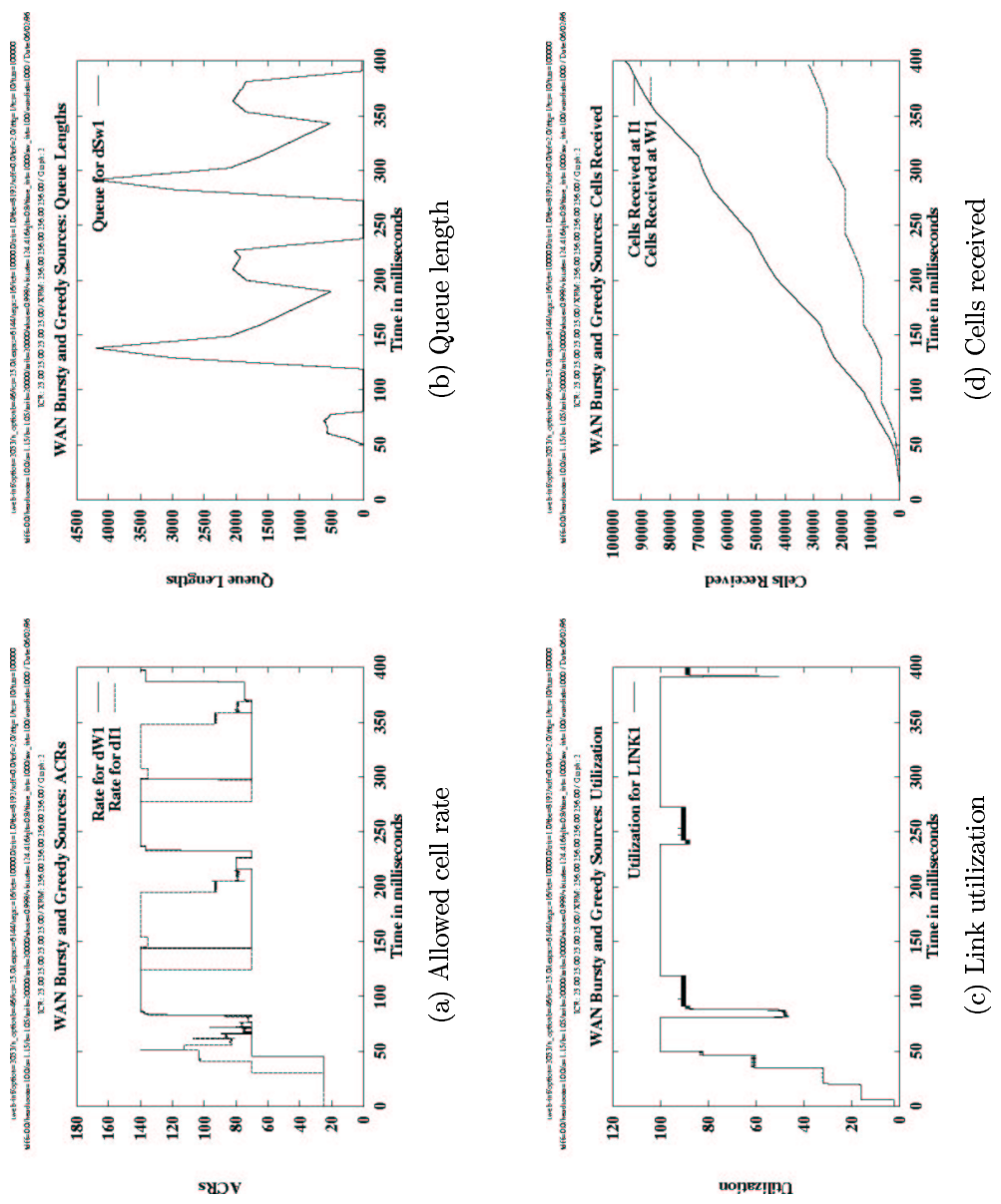
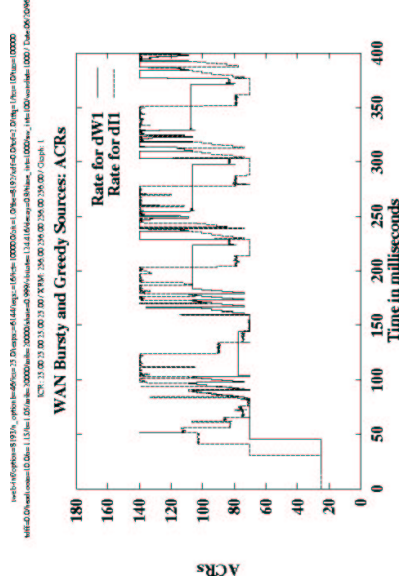
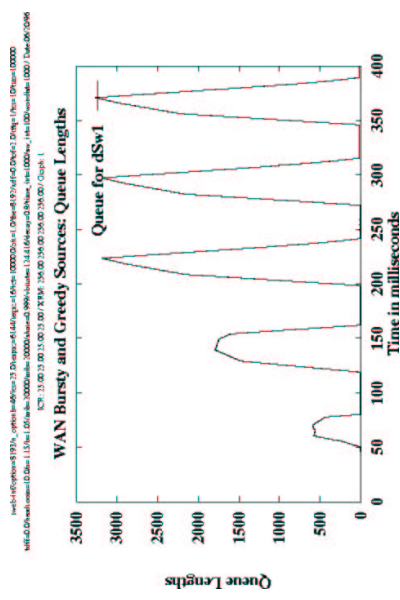


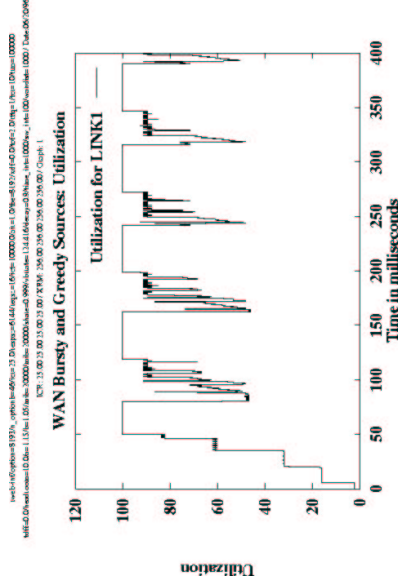
Figure 25: Results for one persistent source and one bursty source (large bursts) in a WAN (ERICA with bidirectional counting)



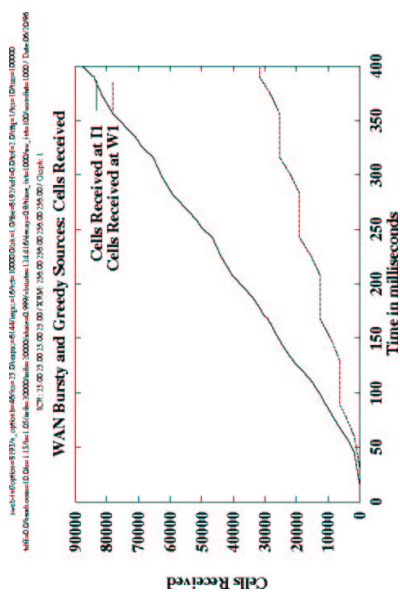
(a) Allowed cell rate



(b) Queue length



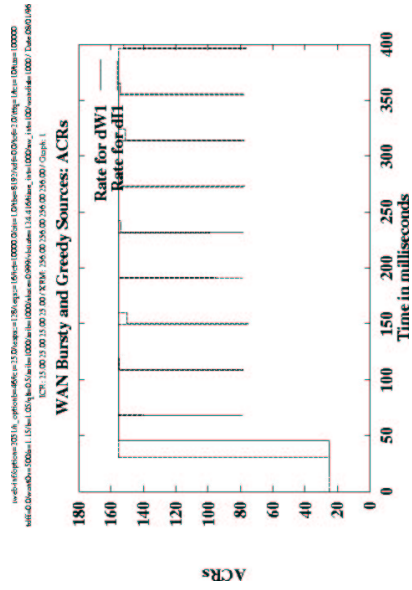
(c) Link utilization



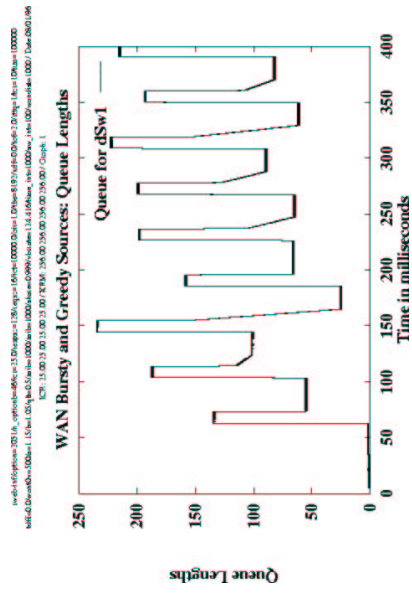
(d) Cells received

Figure 26: Results for one persistent source and one bursty source (large bursts) in a WAN (ERICA with averaging of number of sources)

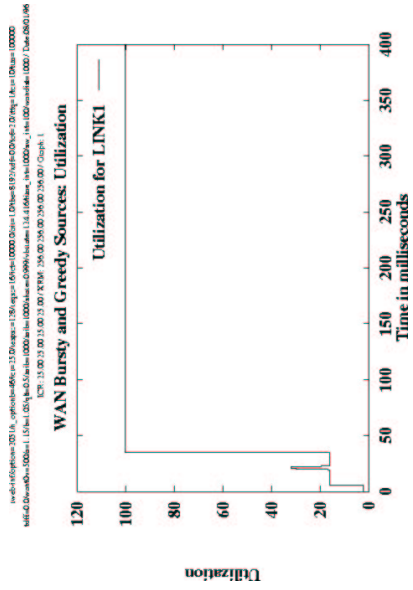




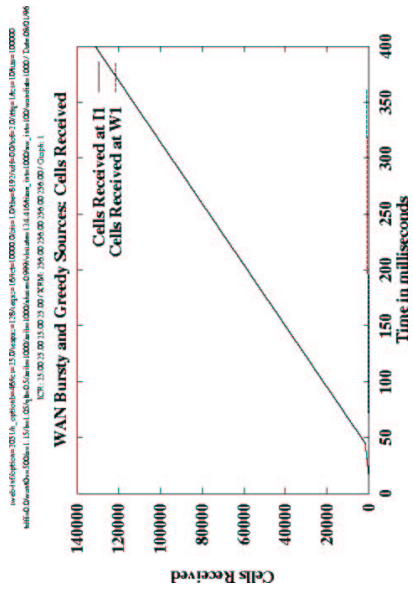
(a) Allowed cell rate



(b) Queue length

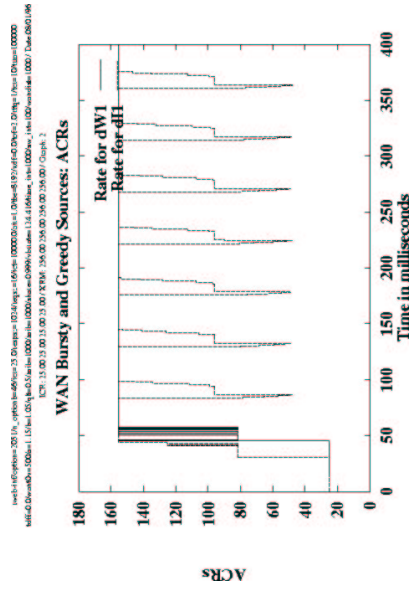


(c) Link utilization

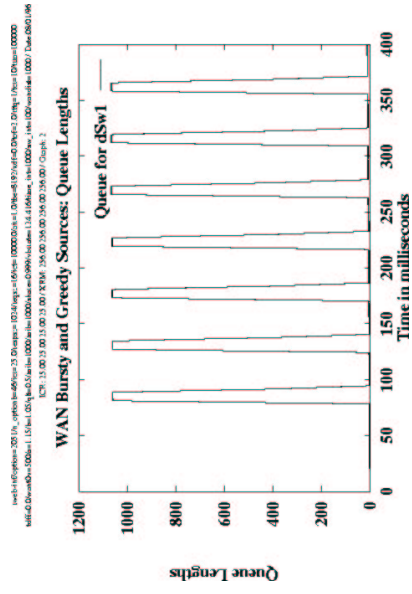


(d) Cells received

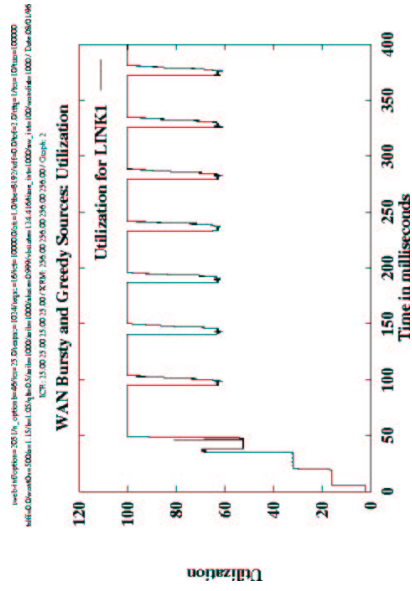
Figure 27: Results for one persistent source and one bursty source (small bursts) in a WAN (ERICA+)



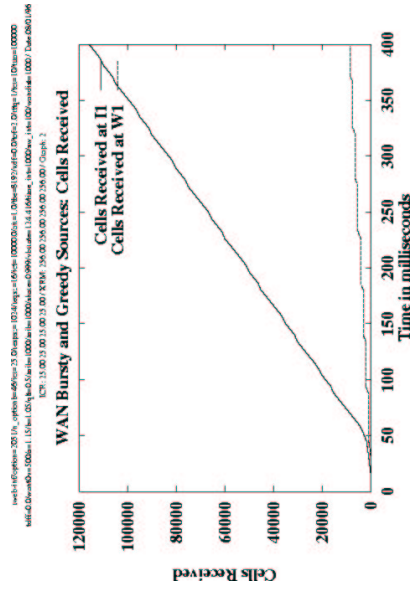
(a) Allowed cell rate



(b) Queue length

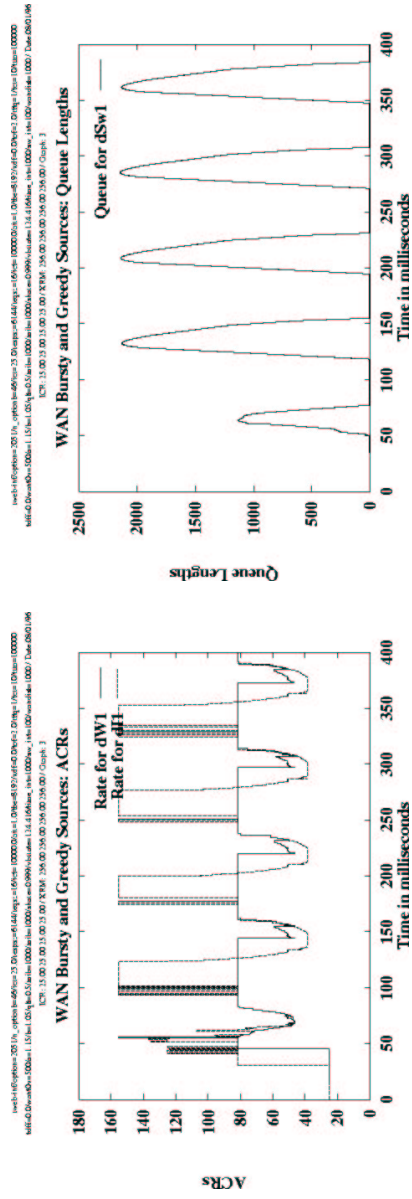


(c) Link utilization



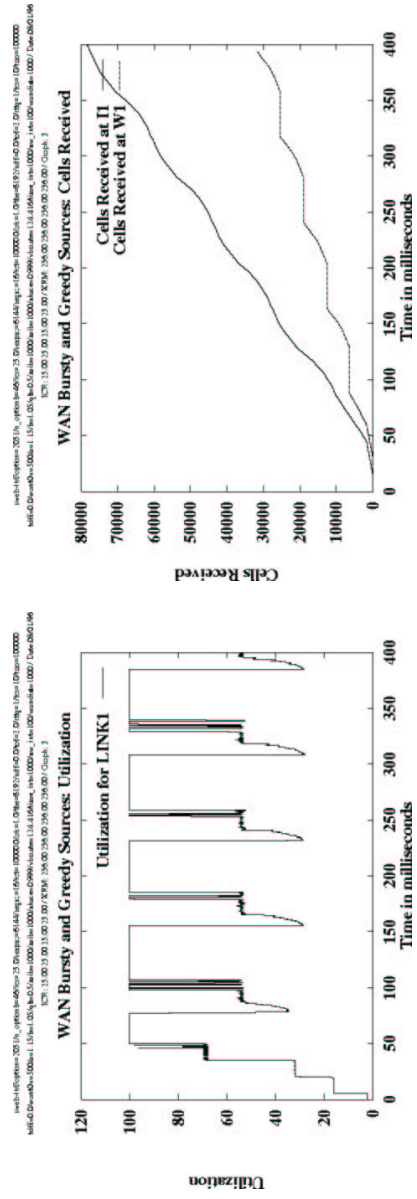
(d) Cells received

Figure 28: Results for one persistent source and one bursty source (medium bursts) in a WAN (ERICA+)



(a) Allowed cell rate

(b) Queue length



(c) Link utilization

(d) Cells received

Figure 29: Results for one persistent source and one bursty source (large bursts) in a WAN (ERICA+)

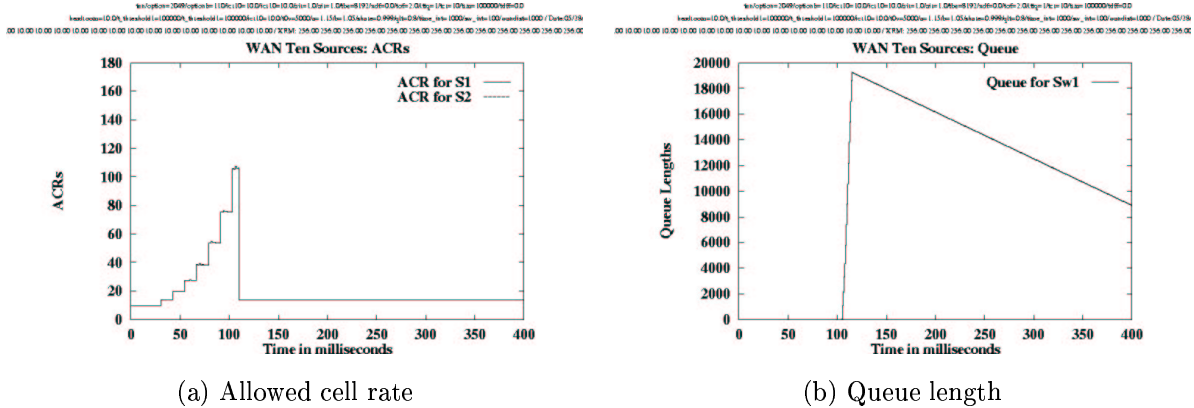


Figure 30: Results for ten ACR retaining sources in a WAN (ERICA)

### 5.7.2 ACR Retention

ACR retention is the problem which occurs when sources are not able to fully use their rate allocations. For example, the input to the ATM end-system can be steady, but have a rate lower than its ABR allocation (allowed cell rate). Another example is an end-system which supports multiple VCs (to possibly different destinations) on a single outgoing link. A VC may not be able to use its ACR allocation because the outgoing link is running at capacity. In such situations, the switches reallocate the unused capacity to the other sources which are unconstrained. However, if the ACR retaining sources suddenly use their allocations, a potential overload situation exists [11].

Figure 30 illustrates the performance of ERICA when there are ten VCs sharing a link. This larger number of connections has been selected to demonstrate the scalability of ERICA to more VCs, as well as to aggravate the problem of ACR retention. Initially, the ten sources are retaining their ACRs, and each cannot send at a rate of more than 10 Mbps. After 100 ms, all the sources suddenly start sending at their full allocations. ERICA rapidly detects the overload and gives the appropriate feedback asking sources to decrease their rates. All the ten sources stabilize at their optimal rates after that.

Figure 31 shows how the per-VC CCR measurement option can mitigate the overload situation arising when all the ACR retaining sources start transmission at their full capacities. The per-VC CCR measurement results in more conservative initial allocations, and hence smaller queues in this case.

## 6 Conclusions

We have discussed the requirements of ABR switch algorithms, and examined how each of these requirements can be tested. As a case study, we presented a comprehensive performance evaluation of the ERICA switch algorithm and demonstrated the effect of several features

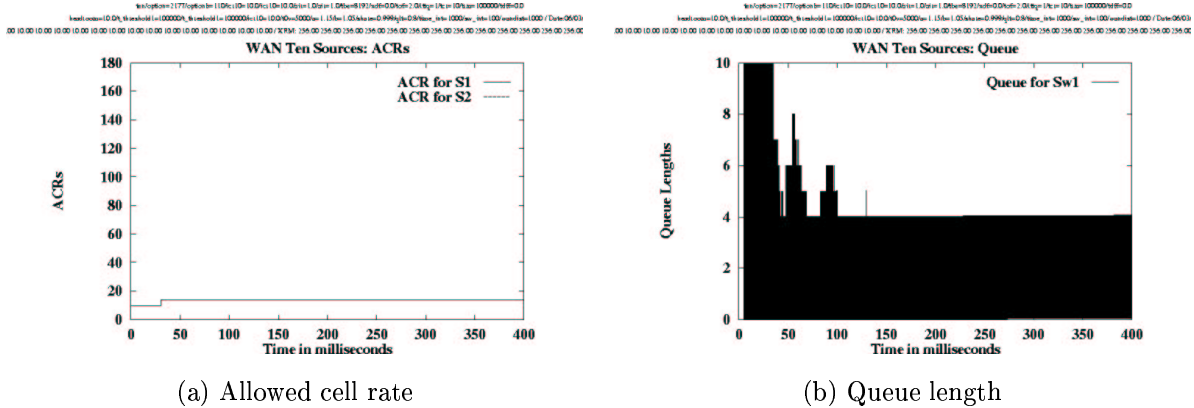


Figure 31: Results for ten ACR retaining sources in a WAN (ERICA with per-VC CCR measurement)

and options of the algorithm. We have examined the efficiency and delay requirements, the fairness of the scheme, its transient and steady state performance, its scalability, and its adaptation to variable capacity and various source traffic models. Simulation results have illustrated that the algorithm gives optimal allocations, and rapidly adapts to load and capacity changes. The performance of the algorithm was examined for various configurations, source models and background traffic patterns.

## References

- [1] Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, and Ram Viswanathan, “ERICA Switch Algorithm: A Complete Description,” *AF-TM 96-1172*<sup>1</sup>, August 1996<sup>2</sup>.
- [2] “The ATM Forum Traffic Management Specification Version 4.0,” ATM Forum Traffic Management AF-TM-0056.000, April 1996. Available as <ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.000.ps>
- [3] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, and F. Lu, “ERICA+: Extensions to the ERICA Switch Algorithm,” *AF-TM 95-1346*, October 1995.
- [4] R. Jain, S. Kalyanaraman, R. Viswanathan, and R. Goyal, “A Sample Switch Algorithm,” *AF-TM 95-0178R1*, February 1995.
- [5] Raj Jain, Shiv Kalyanaraman, and Rohit Goyal, “Simulation Results for ERICA Switch Algorithm with VBR+ABR traffic,” *AF-TM 95-467*, April 1995.

<sup>1</sup>Throughout this section, AF-TM refers to ATM Forum Traffic Management sub-working group contributions.

<sup>2</sup>All our papers and ATM Forum contributions are available through <http://www.cis.ohio-state.edu/~jain>

- [6] Raj Jain, Shiv Kalyanaraman, Ram Viswanathan, and Rohit Goyal, "Simulation Results for the Sample Switch Algorithm," *AF-TM 95-0179*, February 1995.
- [7] Bo-Kyoung Kim, Byung G. Kim and Ilyoung Chong, "Dynamic Averaging Interval Algorithm for ERICA ABR Control Scheme," *AF-TM 96-0062*, February 1996.
- [8] Raj Jain, Shiv Kalyanaraman, Sonia Fahmy and Rohit Goyal, "Source Behavior for ATM ABR Traffic Management: An Explanation," To appear in *IEEE Communications Magazine*, November 1996.
- [9] J. Jaffe, "Bottleneck Flow Control," *IEEE Transactions on Communications*, Vol. COM-29, No. 7, pp. 954-962.
- [10] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," invited submission to *Computer Networks and ISDN Systems*, 1995, also *AF-TM 95-0177*, February 1995.
- [11] Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, Fang Lu and Saragur Srinidhi, "A Fix for Source End System Rule 5", *AF-TM 95-1660*, December 1995.
- [12] Raj Jain, Shiv Kalyanaraman, Fang Lu, and Sonia Fahmy, "Bursty ABR Sources," *AF-TM 95-1345*, October 1995.
- [13] R. Jain, "The Art of Computer Systems Performance Analysis," *John Wiley & Sons*, New York, 1991.
- [14] R. Jain and S. Routhier, "Packet Trains - Measurement and a new model for computer network traffic," *IEEE Journal of Selected Areas in Communications*, Vol. SAC-4, No. 6, September 1986, pp. 986-995. Reprinted in Amit Bhargava, Ed., "Integrated Broadband Networks" Artech House, Norwood, MA, 1990.
- [15] R. Jain, K. K. Ramakrishnan, and D. M. Chiu, "Congestion Avoidance in Computer Networks with a Connectionless Network Layer," *Digital Equipment Corporation, Technical Report*, DEC-TR-506, August 1987, 17 pp. Also in C. Partridge, Ed., *Innovations in Internetworking*, Artech House, Norwood, MA, 1988, pp. 140-156.
- [16] R. Jain, S. Kalyanaraman, Rohit Goyal, R. Viswanathan, and Sonia Fahmy, "ERICA: Explicit Rate Indication for Congestion Avoidance in ATM Networks," *U.S. Patent Application* filed July 19, 1996.