The ERICA Switch Algorithm for ABR Traffic Management in ATM Networks[1]

Shivkumar Kalyanaraman[2], Raj Jain, Sonia Fahmy[3], Rohit Goyal[4] and Bobby Vandalore
The Ohio State University, Department of CIS
Columbus, OH 43210-1277
Email: *shivkuma*@ecse.rpi.edu, {*jain, fahmy, goyal, vandalor*}@cis.ohio-state.edu

Abstract

This paper describes the "Explicit Rate Indication for Congestion Avoidance (ERICA)" scheme for rate-based feedback from ATM switches. In ERICA, the switches monitor their load on each link, determine a load factor, the available capacity, and the number of currently active virtual channels. This information is used to advise the sources about the rates at which they should transmit. The algorithm is designed to achieve high link utilization with low delays and fast transient response. It is also fair and robust to measurement errors caused by the variations in ABR demand and capacity. We present performance analysis of the scheme using both analytical arguments and simulation results. The scheme is being considered for implementation by several ATM switch manufacturers.

1 Introduction

The key new feature of ATM that distinguishes it from other networking technologies is that it provides very sophisticated traffic management. ATM networks use connection admission control, traffic shaping, policing, selective discard, packet discard, and explicit feedback to manage the traffic. Traffic management is particularly important at high speeds since even a short-term overload at these speeds can lead to significant queues and data loss.

ATM networks provide several services. Of these, the available bit rate (ABR) service is ideal for data. In this service, switches use an option to provide an explicit rate feedback to the sources and sources control their traffic accordingly. The ATM Forum Traffic Management Specification [1] contains detailed rules for the ABR source and destination end systems. The rules for switches are also specified. The switch behavior, however, is only coarsely specified so that various vendors can implement their own switch rate allocation algorithms and distinguish their products. The switch rules simply ensure that switches from different vendors will interoperate, though the operation may not be optimal. Several switch algorithms have been developed [2]-[10]. This paper describes one of the earliest of such switch algorithms.

The explicit rate indication for congestion avoidance (ERICA) algorithm was presented at the ATM Forum in February 1995. Since then, its performance has been independently studied in many papers [5, 6, 8]. Also, we have incorporated several modifications into the algorithm [10, 11]. This paper provides a consolidated description and a performance analysis of the algorithm.

The paper is organized as follows. The next two sections examine the ABR service, and describe the switch model and design goals. Section 4 describes the algorithm and examines the tradeoffs involved in selecting the algorithm metrics and parameters. Section 6 presents representative simulations to show the performance of the scheme under strenuous conditions. We also present limited analytical arguments of the convergence of the algorithm in appendix A.

2 The ABR Control Mechanism

According to the ATM Forum specifications, ATM networks currently offer five service categories: constant bit rate (CBR), real-time variable bit rate (rt-VBR), non-real time variable bit rate (nrt-VBR), available bit rate (ABR), and unspecified bit rate (UBR). Of these, ABR and UBR are designed for data traffic exhibiting bursty unpredictable behavior.

The UBR service is simple and does not give sources any guarantees. The network elements try to improve throughput and reduce loss using intelligent buffer allocation [17], cell drop [18] and scheduling. The ABR service provides better service for data traffic by periodically advising sources about the rates at which they should be transmitting. The switches monitor their load, and divide the available bandwidth fairly among active flows. This allows competing sources to get a fair share of the bandwidth while also allowing the link to be fully utilized. The feedback from the switches to the sources is indicated in resource management (RM) cells, which are periodically generated by the sources and turned around by the destinations (refer to figure 1).

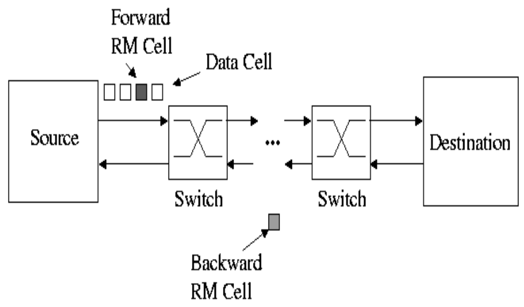


Figure 1: RM cell path

The RM cells contain the source's current cell rate (CCR) and several fields that can be used by the switches to provide feedback to the source. These fields are: Explicit Rate (ER), Congestion Indication (CI) flag, and No Increase (NI)

---

[1]This work was supported by NSF Contract #9628438

[2]Shivkumar Kalyanaraman is now in Department of ECSE, Rensselaer Polytechnic Institute, Troy, NY 12180-3590

[3]Sonia Fahmy is now with Department of Computer Science, Purdue University, Indiana.

[4]Rohit Goyal is at Nexabit Networks (Lucent), Massachusetts.

flag. The ER field indicates the rate that the network can support at the particular instant in time. When starting at the source, the ER field is usually set to the peak cell rate, and the CI and NI flags are clear. On the path, each switch reduces the ER field to the maximum rate it can support, and sets CI or NI if necessary [13].

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, the CI and NI flags, and the ER field of the RM cell [15].

3 Switch Model

Our switch model is shown in figure 2. Each service category has a separate first-in, first-out (FIFO) output queue which feeds to the output link under the control of a scheduling mechanism. The ERICA algorithm works at every ABR queuing point (assumed to be at the output for this paper). For simplicity, we assume that there are at most two classes (VBR and ABR) and ABR has the lower priority, i.e., it gets the left-over capacity after VBR cells are transmitted. We do not consider the case of ABR virtual connections (VCs) guaranteeing non-zero minimum cell rates in this paper. Techniques for adapting a switch scheme to guarantee such rates are suggested in [19]. Other issues not addressed in this paper include the effect of more complex queuing strategies like per-VC queuing, network segmentation using the virtual source/virtual destination (VS/VD) option [1], and multipoint ABR connections. Some of these issues are addressed in [10, 21].

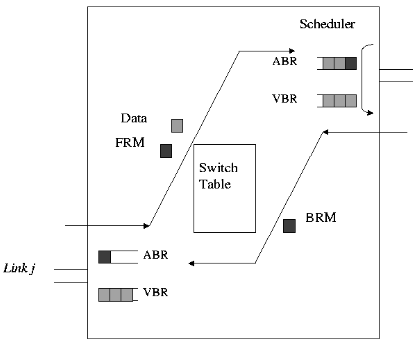


Figure 2: Switch Model

In ERICA, the time is divided into consecutive equal-sized slots called *"switch averaging intervals."* As shown in figure 2, the measured load in the forward direction in each slot is used to provide feedback in the reverse direction in the next slot. The feedback may be computed at the end of each slot or when a backward RM cell (BRM) is received.

One of the design features of ERICA is that switches give *at most one* feedback value per flow during any averaging interval. This precludes the switch from giving multiple conflicting feedback indications in a single averaging interval using the same control values.

ERICA gives feedback only in the explicit rate field in the RM cell. It is possible to additionally throttle or moderate the sources by setting the CI and NI bits in the RM cell using policies suggested by several other schemes [5, 19].

3.1 Design Goals

In designing ERICA, our main goals were to maximize link utilization, minimize queuing delays, achieve fair allocation, reduce transient response time, and achieve stable and robust operation. Each of these goals is explained below.

Utilization: Our first goal was to maximize the link utilization $\rho(t)$. This is done by allocating as much of the available capacity to active ABR flows as possible. The entire link capacity that is not used by the higher priority VBR and CBR service categories is potentially available to ABR.

ERICA tracks ABR utilization using a metric called "load factor" (z, refer to section 4.1). Specifically, z is proportional to the ratio of the ABR input rate to the ABR capacity. Using z, the ERICA utilization goal is a *steady-state* operating point in the neighborhood of $z = 1$. In certain configurations, this goal cannot be realized for all bottlenecks, in which case, the maximization of z (where $z \leq 1$) is desired for those bottlenecks.

Queuing Delay: High link utilization can result in large queues and long delays. The instantaneous utilization $\rho(t)$ is unity when a large queue backlog exists at the bottleneck. Our efficiency goal is to maximize the link utilization while keeping the steady-state queuing delay under a target maximum value. The concept of maximizing throughput (utilization) and minimizing delay is known as "congestion avoidance" [14]. In ERICA, the definition of congestion avoidance has been broadened to include the goal of reducing the delay. When the queuing delay is more than the target, allocations are reduced so that queue size decreases.

Fairness: A commonly used criterion for describing fairness is the *max-min allocation* [7]. Among the allocation vectors (with elements in a descending order) that result in link utilizations less than or equal to 100%, the lexicographically largest vector is the max-min allocation. In simple terms, the goal of max-min allocation is to give equal shares of a resource to all contending sources. However, if some sources cannot use their shares (since they are constrained at some other bottleneck), they are given the maximum share that they can use, and the left-over is available for equal allocation to other contending sources. This process can be shown to lead to max-min allocation, provided the source demands are fixed and non-infinite. Several schemes including ERICA [3, 4, 19] use this method and search for a "maximum equal share" value to allocate to all contending sources. In ERICA, fairness is sought only after efficiency has been achieved, that is, the load factor z is in the neighborhood of unity. If the load is too high or too low, rates for all sources are decreased or increased so that efficiency is rapidly achieved.

In addition to the above "steady state" goals, ERICA aims to achieve the following goals:

Stability and Transient Performance: A *stable* system is one which can re-establish its steady state after perturbations. The *transient performance* of the scheme determines how quickly the steady state is re-established.

Robustness: In cases where the system has *no steady-state* (e.g., due to persistent variation in capacity and demand), the scheme should be robust. This means that its essential performance metrics should not degrade to unacceptable levels.

We emphasize that ERICA is an engineering solution which incorporates these design goals. We provide limited analytical arguments and simulations to support our performance claims.

4 The ERICA Algorithm

The ERICA algorithm periodically monitors the load on each link and determines the ABR capacity, the load factor (z), and the number of active virtual connections (N) during each "averaging interval."

The complete pseudo-code including all features of ERICA, is given in [10]. Below we present the key steps in ERICA as a pseudo-code. The variable MaxAllocPrevious (or MaxAllocPrev., abbreviated) represents the maximum allocation given during the previous averaging interval to any source transmitting to this output link. Similarly, MaxAllocCurrent (or MaxAllocCurrent., abbreviated) is used to determine the maximum allocation given to any source so far in the current averaging interval.

Initialization:

MaxAllocPrevious ← MaxAllocCurrent ← FairShare

End of Averaging Interval:

$$\text{Total ABR Cap.} \leftarrow \text{Link Cap.} - \text{VBR Cap.} \quad (1)$$

$$\text{Target ABR Cap.} \leftarrow \mathit{Fraction} \times \text{Tot. ABR Cap.} \quad (2)$$

$$z \leftarrow \frac{\text{ABR Input Rate}}{\text{Target ABR Cap.}} \quad (3)$$

$$\text{FairShare} \leftarrow \frac{\text{Target ABR Capacity}}{\text{Number of Active VCs}} \quad (4)$$

$$\text{MaxAllocPrevious} \leftarrow \text{MaxAllocCurrent} \quad (5)$$

$$\text{MaxAllocCurrent} \leftarrow \text{FairShare} \quad (6)$$

When an FRM is received:

CCR[VC] ← CCR_in_RM_Cell

When a BRM is received:

$$\text{VCShare} \leftarrow \frac{CCR[VC]}{z} \quad (7)$$

IF ($z > 1 + \delta$)

THEN ER ← Max (FairShare, VCShare) (8)

ELSE ER ← Max (MaxAllocPrev., VCShare) (9)

MaxAllocCur. ← Max (MaxAllocCur.,ER) (10)

IF (ER > FairShare AND CCR[VC] < FairShare)

THEN ER ← FairShare (11)

$\text{ER}_{RM\_Cell} \leftarrow$ Min (ER_{RM_Cell}, ER, Target ABR Cap.) (12)

This pseudo-code achieves the goals of efficiency, fairness, and bounded delay as explained next.

4.1 Efficiency: Using the Load Factor Metric

The key metric used in ERICA is the load factor (z) which is the ratio of the measured input rate at the port to the target ABR capacity, as given by equation (3):

$$z \leftarrow \frac{\text{ABR Input Rate}}{\text{Target ABR Capacity}}$$

The target ABR capacity is a fraction of the total ABR capacity (equation (2)), where the fraction may be determined based upon queuing delays (refer to section 4.3). The load factor is a compact and accurate congestion indicator, and is arguably better for rate-based schemes than using queue length alone [9].

The load factor is used in ERICA with the goal of driving the system towards an *efficient operating point, defined as the neighborhood of* $z = 1$. The simplest way to achieve efficiency is to reduce each VC's activity by a factor of z. In other words, each VC's allocation ("VCShare" in the pseudo-code above) is set to the VC current cell rate (CCR) divided by the load factor z, or $\frac{CCR[VC]}{z}$. Here, CCR is the estimate of the source current rate. CCR may be read from the forward RM cells of the VC, or measured by the switch. Either way, the CCR value is stored in a table and used for this calculation. The analytical arguments given in the appendix show that this technique does drive the system to the neighborhood of $z = 1$.

Though VCShare can be used to achieve efficiency, it may not be a fair allocation. A mechanism is required to equalize the rate allocations, while ensuring that the bottleneck load factor remains in the neighborhood of unity. This is the topic of the following section.

4.2 Max-Min Fairness: Equalizing Allocations

One way to equalize allocations is to calculate the *maximum* of the VCShare values and assign this maximum value to all sources. This can result in sharp load changes (and extended periods of overload). For example, consider the case when the allocation of N sources are $(C - n\epsilon, \epsilon, ..., \epsilon)$, where ϵ and $n\epsilon$ are negligibly small. The load factor is close to unity (assuming no prior queue buildups). The maximum of these allocations is $C - n\epsilon$, which we call "MaxAllocPrevious" referring to the fact that the maximum is calculated in the previous cycle and used in the next cycle. In the next cycle, the allocations based upon this maximum value would be: $(C - n\epsilon, C - n\epsilon, ..., C - n\epsilon)$, leading to a load factor of almost n. To avoid this, ERICA uses a two step process. A variable, "FairShare," is computed as the ratio of the target ABR capacity and the number of active connections. If a VC is initially sending at a rate below FairShare, it is not allowed to increase its rate to MaxAllocPrevious in the current cycle. In particular, all sources sending below the FairShare are allowed to rise to FairShare, and those sending above FairShare are allowed to rise to MaxAllocPrevious. Therefore a VC sending at a rate below FairShare would take two cycles (steps) to reach the maximum possible allocation.

These features and mechanisms (VCShare, rate equalization, at least FairShare, at most FairShare if rate is low) are incorporated into the ERICA algorithm as presented in equations (7) through (11). The parameter δ is used for the equalization of allocations (equation (9)) and defines the "neighborhood of unity."

4.3 Queue Control

In section 4.1, we noted that the Target ABR Capacity is a fraction of the Total ABR Capacity. This fraction can be a function of the current queue length, f(Q), i.e.,
Target ABR Capacity = f(Q) $\times$ Total ABR Capacity.

The function f(Q), called the "queue control function," allows only a specified fraction of the available capacity to be allocated to the sources. Such a function should not artificially restrict the system utilization to a value smaller than 100%, and it should compensate for errors in measurement (which manifest as queues). Further, given a fixed number of persistant sources of traffic, it should allow the system to achieve a queuing delay target. A control-theoretic definition of these steady-state and robust stability conditions is given in Ozbay et al [24] and an analysis of queue management techniques for ABR while maintaining max-min fairness is presented by Ma and Ramakrishnan [23]. A simple queue control function such as a constant function used in earlier versions of ERICA and the OSU scheme (called "Target Utilization (U)') does not meet these requirements.

The alternative is for f(Q) to vary depending upon the queuing delay. A number of such functions can be designed

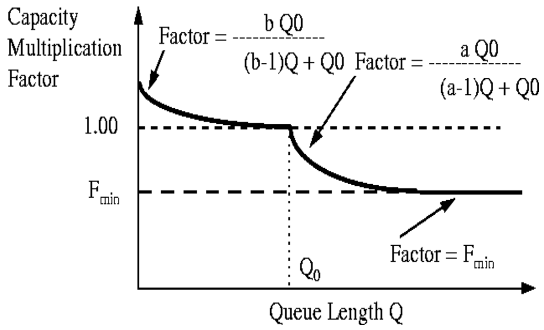


Figure 3: The queue control function in ERICA

[29, 23]. One of the functions that worked particularly well is the following (also refer to figure 3):

$$f(Q) = Max(QDLF, \frac{a \times Q0}{(a-1) \times q + Q0}) \text{ for } q > Q0$$

and

$$\text{f(Q)} = \frac{b \times Q0}{(b-1) \times Q + Q0} \text{ for } 0 \le Q \le Q0$$

Here, f(Q) is a truncated rectangular hyperbola assuming values between 1 and $QDLF$ (queue drain limit factor) in the range $Q0$ to infinity, and values between b and 1 in the range 0 to $Q0$. Both curves intersect at $Q0$, where the value is 1. To emphasize the control of queuing delay, ERICA uses a parameter $T0$ (target queuing delay) which is converted into the target queue length $Q0$ parameter before performing the calculation given above. The parameter settings are discussed in section 5.

5 Parameter Choices and Robustness Issues

The robustness and performance of ERICA are significantly dependent upon how measurements are performed and parameters are chosen. ERICA parameters may be classified into: (1) parameters related to measurement and averaging, (2) queue control parameters, and (3) max-min fairness parameter δ.

5.1 Measurement and Averaging Related Parameters

As previously mentioned, the essential metrics used in ERICA, i.e., load factor (z) and number of active connections (N, for FairShare calculation) are measured during consecutive switch averaging intervals. Variation in demand, available capacity, and number of currently active connections, could lead to errors in the estimation of these metrics, which, in turn, would lead to errors in feedback. ERICA is particularly sensitive to under-estimation of N, because FairShare (which is calculated using N) is the *minimum* allocation given to sources. ERICA is also sensitive to oscillations in estimation of the load factor z during alternating periods of demand activity and inactivity, and in the presence of higher priority VBR traffic. Therefore, the choice of the switch averaging interval is critical to the performance of ERICA.

To determine a reliable averaging interval, observe that the activity of any source is determined within a round trip time (RTT). Moreover, the maximum time for feedback from any switch to reach a source, and the resultant activity to be experienced at the switch (called the "feedback delay") is the maximum RTT (max RTT) plus the maximum inter-RM-cell-time (max inter-RM-cell-time). Allowing time for transient loads between averaging intervals to subside, a reliable value for the switch averaging interval is at least 2 $\times$ *(max RTT + max inter-RM-cell-time).*

Choosing averaging intervals smaller than *max RTT* poses the risk of errors in z and N (due to temporary inactivity of sources), and choosing intervals smaller than *max inter-RM-cell-time* poses the risk of not giving feedback to some sources in every measurement interval. In fact, intervals smaller than the maximum inter-cell-time would guarantee that N is underestimated.

One solution to the problem of estimation errors with small intervals is to use separate averaging intervals for N and z to allow reliable estimation of each, and give feedback in every z-interval which would be the smaller of the two. ERICA employs an alternate method. The method is to use a single *base averaging interval*, and optionally use exponential averaging techniques to improve reliability and reduce variance in the measurements. The base averaging interval is chosen statically in the range [5 ms, 20 ms] for a OC-3 bottleneck link (used in our simulations) and may be scaled by the ratio of OC-3 (155 Mbps) speed to the given bottleneck link speed for slower or faster links.

Exponential averaging can be applied for the load factor z using the formula: z = (exponential average of input rate)/(f(Q) $\times$ (exponential average of available capacity)), where the exponential average of input rate or available capacity (denoted as x) is calculated as: $x = \alpha x + (1-\alpha)x$. Our simulations indicate that an α value of 0.8 is sufficient given a base averaging interval choice in the range [5 ms, 20 ms]. This α value gives significant weight to the latest sample of input rate or available capacity. Our simulations use OC-3 bottleneck links, and encompass LAN/WAN/satellite configurations with significant variation in demand and available capacity [10].

Averaging the number of active VCs, N, is performed in a different manner. The problem is that when not even one cell of an "active" VC is seen in the base averaging interval, it would be counted as inactive. This error would result in an increase in FairShare which is the *minimum* allocation given to VCs, and could lead to instability (manifested as unbounded queues). This problem can be simply addressed by using a separate interval for measuring N and set this interval to Max { RTT, 1/(minimum rate allocation) } of any VC. Since this is not possible, we approximate it though this procedure. We first define the "activity level" of a VC as a real number between 0 and 1. The activity level is initialized to 1 whenever any cell from the VC is seen and decayed by a multiplicative parameter, *DecayFactor*, in each successive interval in which a VC is inactive. At the end of each interval, the sum of all activity levels would give the value of N (which is now a real number). Setting *DecayFactor* to a value sufficiently close to unity would ensure that the error in estimation due to the exponential decay would be small. We have observed that a value of $DecayFactor$ in the range [0.9, 0.95] is sufficient given our base averaging interval choice in the range [5 ms, 20 ms].

5.2 Queue Control Parameters

Recall that the queue control function f(Q) used in ERICA (section 4.3) is one of several possible functions [29], and has four parameters: $T0$, $QDLF$, a and b. The parameter $T0$, which specifies the target queuing delay, is affected by several other system parameters such as the available buffer size, the bottleneck link speed, and the maximum round trip time (or the base averaging interval length). $T0$ also affects the decrease function component of f(Q) in conjunction with the parameters a and b. The decrease function affects how quickly excess queues are drained. The combination of these factors make the choice of $T0$ important.

A heuristic used in ERICA ensures that the maximum oscillation of queues in the steady-state will be no larger than $Q0$. As described in appendix A, in steady-state, the maximum deviation of the load factor is determined by the parameter δ. Specifically, assuming that queuing deviations are corrected in one averaging interval, we have the relationship: $T0 \geq \delta \times BaseAveragingInterval$. Given that our choice of δ is 0.1 (refer to next section) and the base averaging interval lies between [5 ms, 20 ms], then $T0$ lies between [0.5 ms, 2 ms].

The parameter $QDLF$ (queue drain limit factor) limits the amount of available capacity that can be allocated as *drain capacity* to clear excess queues, and determines the effectiveness of the queue control policy in reacting to transient queues. When the aggregate input rate is equal to the available capacity (i.e., a balanced load), $QDLF$ also determines the minimum value of the load factor, z. The range of z determines the range of possible feedback values or the maximum possible oscillations in feedback (a stability concern). We have found that a $QDLF$ choice of 0.5 balances these conflicting concerns for a wide range of configurations and loads.

The parameters a and b, in conjunction with $T0$, determine the slope of the rectangular hyperbolas. The steeper the slope, the more sensitive the scheme is to small variations in queue length. Further larger difference in the slopes of the two hyperbolas accentuates the effect of the discontinuity of f(Q) at $Q0$ leading to larger oscillations around Q0 in the steady state (if one exists and is reached). Since a and b affect these slopes, the choice must be made considering these issues as well.

To be consistent with the steady-state queue fluctuation heuristic for choosing $T0$, the ideal choice for b is 1 which eliminates the b-hyperbola. In practice, a value between [1, 1.05] can be chosen where a larger value of b allows the steady-state queuing delay to be closer to the target, at the risk of incurring steady-state oscillations. For the parameter a, we have found that a value in the range [1.10, 1.25] sufficiently differentiates the ERICA queue control function from simple step or linear functions. Larger values of a make the function closer to a step function with the possibility of larger queue oscillations, and smaller values make the function closer to a linear function with a small slope, which limits the speed of response to transient queues.

5.3 The Max-min Fairness Parameter δ

The max-min fairness parameter, δ, defines the steady-state operating region towards which ERICA attempts to drive the system. Specifically, in ERICA, we consider the system be max-min fair when the load factor z is in the range $[1, 1+\delta]$ and all allocations are equal. We use this weaker definition of max-min fairness because converging to $z = 1$ exactly is not guaranteed in ERICA. Further, when $z > 1+\delta$, we consider the system allocations to be "infeasible" (i.e. we estimate average load to be larger than average capacity which is unsustainable) [23, 7], and therefore not max-min fair. When $z < 1+\delta$, the allocations cannot be max-min fair by definition [7].

Observe that in the steady-state, the minimum drain capacity is determined by the relation:

$$0 \leq \delta \times \text{Target Cap.} \leq \text{Available Cap.} - \text{Target Cap.}$$

Rearranging the terms and applying the relationship that target capacity is at least $QDLF \times$ Available Capacity, we have:

$$\delta \in (0, (1/QDLF) - 1]$$

For $QDLF$ of 0.5, this gives us a range of (0, 0.5] for δ. The upper bound is a weak one since a δ value of 0.5 would result in minimal drain capacity and possibly large transient queues (due to the equalization of rates to the maximum allocation). The value of δ chosen in ERICA is 0.1 which allows sufficient drain capacity, and leaves a non-trivial zone for rate-equalization to improve convergence towards max-min fairness.

6 Performance Evaluation of ERICA

In this section, we present simulations to verify the performance of ERICA under strenuous conditions not considered in the analytical arguments in appendix A. Our simulations use all the features of ERICA algorithm.

The parameter set used in our simulations is shown in Table 1:

6.1 Max-Min Fairness

We use the popular Generic Fairness Configuration-2 (GFC-2) to test the utilization, queue lengths and fairness of the scheme. The configuration (illustrated in Figure 4) has multiple bottlenecks and connections with different round-trip times. This configuration was selected by the ATM Forum traffic management working group as the test configuration to compare various schemes.

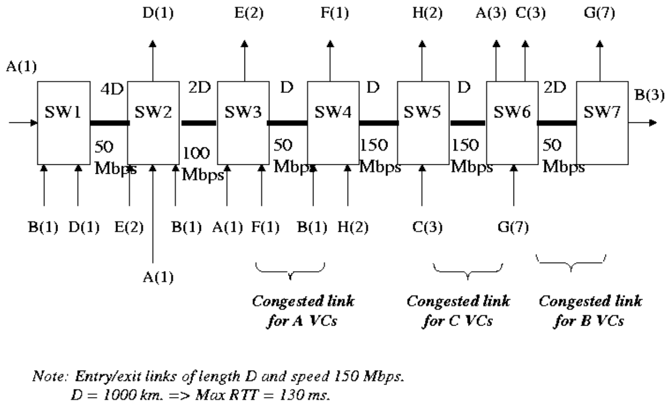


Figure 4: The Generic Fairness Configuration-2

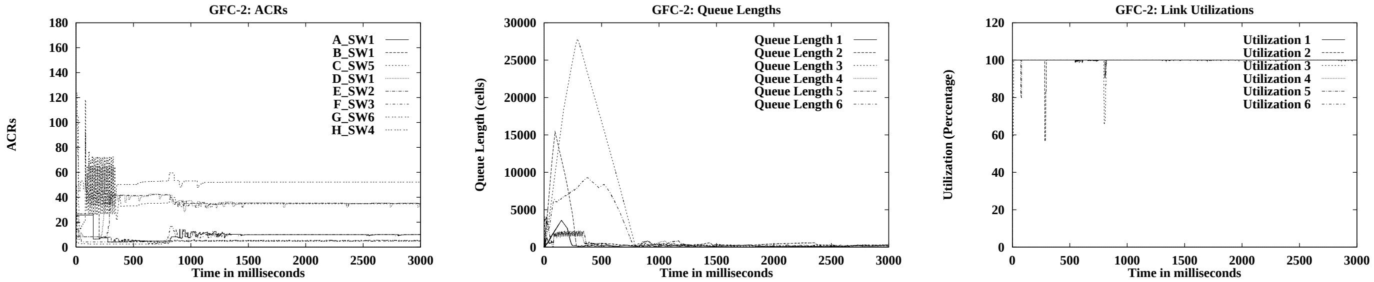
The following are the expected rate allocations as per the max-min fairness criterion. Note that the link bandwidth is adjusted by 48/53 to get an expected application throughput.

VC	Fair share calculation			Fairshare	
A	1/4 of 40	=	$10 \times 48/53 =$	9.1	Mbps
B	1/10 of 50	=	$5 \times 48/53 =$	4.5	Mbps
C	1/3 of 105	=	$35 \times 48/53 =$	31.7	Mbps
D	35	=	$35 \times 48/53 =$	31.7	Mbps
E	1/2 of 70	=	$35 \times 48/53 =$	31.7	Mbps
F	10	=	$10 \times 48/53 =$	9.1	Mbps
G	1/10 of 50	=	$5 \times 48/53 =$	4.5	Mbps
H	1/2 of 105	=	$52.5 \times 48/53 =$	47.6	Mbps

Simulation results are shown in Figure 5. Figure 5(a) shows the allowed cell rates (ACRs) of the sources. Figure 5(b) shows the queue lengths at the ports connected to the next switch for the first six switches. Figure 5(c) shows the link utilizations of the links connecting two switches. The optimal allocations are achieved in under 400 ms (under 4 round trips), and the queues are drained within 800 ms (under 7 round trips). During the transient period, the link utilizations are close to 100% and the queue lengths are controlled to reasonable values (maximum queue is $<$ 30000 cells, i.e.,

Table 1: Parameter values

Parameter	Value	Suggested Range	Purpose
δ	0.1	(0, 0.5]	Max-min fairness
$T0$	1.5 ms	[0.5 ms, 2 ms]	Queue control
a	1.15	[1.10, 1.25]	Queue control
b	1	[1, 1.05]	Queue control
$QDLF$	0.5	0.5	Queue control
Averaging Interval	5 ms	[5 ms, 20 ms]	Measurement of metrics
$DecayFactor$	0.9	[0.9, 0.95]	Long-term averaging of number of active VCs
α	0.8	[0.7, 0.9]	Long-term averaging of input rate and capacity



(a) Allowed Cell Rate (ACR)

(b) Bottleneck Queue Lengths

(c) Bottleneck Link Utilizations

Figure 5: Simulation results w

< 270 ms or 2 round trip times at 50 Mbps bottleneck rate). The steady-state utilizations are close to 100% and the queue lengths are kept close to zero. The minimal oscillations in the steady-state are due to the small variation in queuing delays. The initial rate assignment for each source in this simulation was picked randomly. For reasonable confidence, we repeated this experiment with other random values which gave similar results.

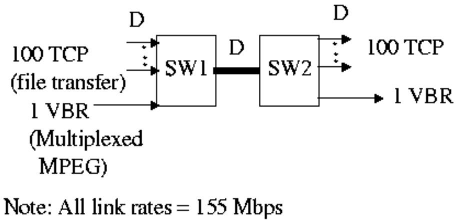


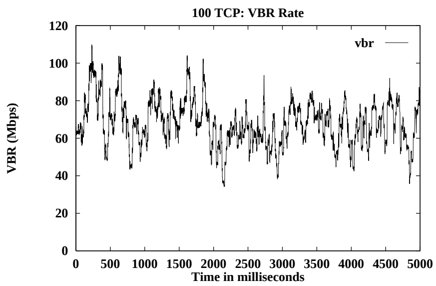
Figure 6: TCP + VBR Configuration

6.2 Robustness

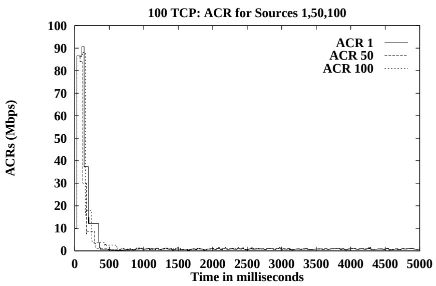
For testing the robustness of the scheme, we need a configuration which attacks the weaknesses of the scheme, namely, its dependency upon measurements. Variation in load and capacity could lead to measurement and feedback errors resulting in unbounded queues or low average utilization. The TCP and VBR configuration (refer to figure 6) is designed to test this case.

The configuration simulates capacity variation by using a higher priority VBR virtual circuit which carries traffic multiplexed from fifteen long-range dependent sources [10]. The traffic generated by this VC (and as a result, the ABR capacity) is highly variable as shown in figure 7(a). The configuration simulates load variation by using TCP sources carrying infinite *ftp* traffic. The load variation is caused by the startup dynamics of TCP. The TCP slow start protocol begins with small window sizes, and the amount of data it sends is limited by the window size (window-limited), rather than a network-assigned rate. As a result, the load offered by an individual TCP connection is bursty, i.e., it consists of active and idle periods. As the TCP window size grows, the active periods become longer. Assuming no packet losses, the TCP source eventually appears to be the same as a persistent source and its load is controlled by network-assigned rates (rate-limited). The queues build up when both demand variation and capacity variation exist in the system. We use 100 sources and synchronize them such that the load phases (idle and active periods) of multiple sources coincide to create a worst case scenario.

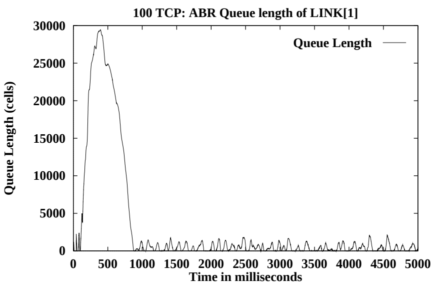
Figures 7(b), (c) and (d) show ATM level metrics (ACRs of sources 1, 50 and 100; queue length at output port of switch 1; link utilization of bottleneck link), while figures 7(e) and (f) show the TCP level metrics (congestion windows and sender sequence numbers of sources 1, 50 and 100).



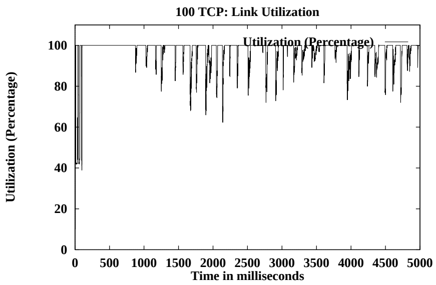
(a) VBR Rate



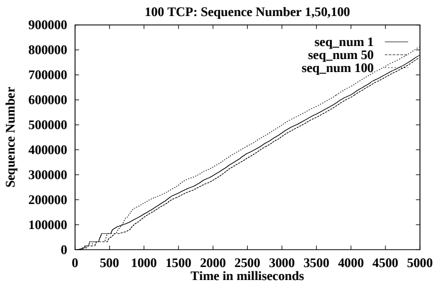
(b) Allowed Cell Rate (ACR)



(c) Bottleneck Queue Length



(d) Bottleneck Link Utilization



(e) TCP Send Sequence Numbers

Figure 7: Simulation results with the 100 TCP and VBR configuration

The graphs show that ERICA successfully controls the TCP sources once they become rate-limited. As a result, the buffer requirement at the bottleneck is not a linear function of the number of sources. Though the system does not have a steady-state (VBR traffic is always variable), ERICA controls the maximum and average queues and keeps utilization high (consistent with the priorities assigned in section 3.1). The congestion window and sender sequence number graphs show that the allocations to contending sources are fair despite the variation in load and capacity.

7 Concluding Remarks

In this paper, we have described the design and evaluation of the ERICA switch algorithm for ATM ABR congestion control. We presented a simple switch model and explained design goals. The key design goals are max-min fair steady-state operation with controlled queuing delays, stability, and robustness to variation in ABR load and capacity. We then presented the ERICA algorithm, showing how the goals and simplicity requirements determine every step in the algorithm.

The scheme requires that the switches periodically monitor their load on each link and determine a load factor, the available capacity, the queue length, and the number of currently active virtual connections. This information is used to calculate a fair and efficient allocation of the available bandwidth to all contending sources. The measurement aspects which determine the robustness of the algorithm are treated in depth.

A limited analysis of the convergence properties is given in appendix A. In addition, we present simulation results illustrating that the scheme meets the desired goals, including good steady-state behavior (high utilization, controlled queuing delay, max-min fairness), rapid convergence from network transients, and robustness to load and capacity variations.

In conclusion, we note that the promise of explicit rate control is higher fidelity control in terms of a number of objectives (fairness, utilization, queueing delays). But the addition of provable robustness as a goal, especially with the uncertainty in a large number of parameter dimensions (like time-delays, load, capacity, number of active sources), and extension to multiple bottleneck cases with independent controllers makes it a non-trivial control problem. ERICA represents an engineering tradeoff.

References

[1] "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

[2] L. Kalampoukas, A. Varma and K. K. Ramakrishnan, "An efficient rate allocation algorithm for ATM networks providing max-min fairness," In *Proceedings of the 6th IFIP International Conference on High Performance Networking*, September 1995.

[3] K. Siu and T. Tzeng, "Intelligent congestion control for ABR service in ATM networks," *Computer Communication Review*, Volume 24, No. 5, pp. 81-106, October 1995.

[4] L. Roberts, "Enhanced PRCA (Proportional Rate-Control Algorithm)," *AF-TM 94-0735R1*, August 1994.

[5] Y. Afek, Y. Mansour and Z. Ostfeld, "Phantom: A simple and effective flow control scheme," In *Proceedings of the ACM SIGCOMM*, August 1996.

[6] D. Tsang and W. Wong, "A new rate-based switch algorithm for ABR traffic to achieve max-min fairness with analytical approximation and delay adjustment," In *Proceedings of IEEE INFOCOM*, pp. 1174-1181, March 1996.

[7] A. Charny, D. D. Clark and R. Jain, "Congestion Control with Explicit Rate Indication," In *Proceedings of ICC'95*, June 1995. http://www.cis.ohio-state.edu/~jain/papers/charny.htm

[8] A. Arulambalam, X. Chen and N. Ansari, "Allocating Fair Rates for Available Bit Rate Service in ATM Networks," *IEEE Communications Magazine*, 34(11):92–100, November 1996.

[9] R. Jain, S. Kalyanaraman and R. Viswanathan, "The OSU Scheme for Congestion Avoidance in ATM Networks: Lessons Learnt and Extensions," Performance Evaluation Journal, Vol. 31/1-2, December 1997. http://www.cis.ohio-state.edu/~jain/papers/perf-sub.htm

[10] S. Kalyanaraman, "Traffic Management for the Available Bit Rate (ABR) Service in Asynchronous Transfer Mode (ATM) networks" *Ph.D. Dissertation*, Dept. of Computer and Information Science, The Ohio State University, August 1997. http://www.cis.ohio-state.edu/~jain/theses/shiv.htm

[11] R. Jain, S. Kalyanaraman, R. Viswanathan, R. Goyal and S. Fahmy, "ERICA: Explicit Rate Indication for Congestion Avoidance in ATM Networks," *U.S. Patent No. 5,805,577*, Issued September 8, 1998. http://www.cis.ohio-state.edu/~jain/patents/patente.htm

[12] A. Arora and M. Gouda, "Closure and convergence: A foundation of fault-tolerant computing" *IEEE Transactions on Software Engineering,* Vol. 19, No. 10, pages 1015 – 1027, 1993.

[13] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," *Computer Networks and ISDN Systems*, October 1996. http://www.cis.ohio-state.edu/~jain/papers/cnis.htm

[14] R. Jain and K. K. Ramakrishnan. "Congestion avoidance in computer networks with a connectionless network layer: Concepts, goals, and methodology," In *Proceedings of IEEE Computer Networking Symposium*, pp 134-143, April 1988. http://www.cis.ohio-state.edu/~jain/papers/cr1.htm

[15] R. Jain, S. Kalyanaraman, S. Fahmy and R. Goyal, "Source Behavior for ATM ABR Traffic Management: An Explanation," *IEEE Communications Magazine*, November 1996. http://www.cis.ohio-state.edu/~jain/papers/src_rule.htm

[16] S. Kalyanaraman, R. Jain, S. Fahmy, R. Goyal and J. Jiang, "Performance of TCP over ABR on ATM backbone and with various VBR traffic patterns," In *Proceedings of ICC'97*, June 1997. http://www.cis.ohio-state.edu/~jain/papers/tcp_vbr.htm

[17] R. Goyal, R. Jain, S. Kalyanaraman, S. Fahmy, B. Vandalore and S. Kota, "TCP Selective Acknowledgments and UBR Drop Policies to Improve ATM-UBR Performance over Terrestrial and Satellite Networks", In *Proceedings of IC-CCN'97*, September 1997, http://www.cis.ohio-state.edu/~jain/papers/ic3n97.htm

[18] A. Romanov and S. Floyd, "Dynamics of TCP Traffic over ATM Networks," *IEEE Journal on Selected Areas of Communications*, Vol 13, No. 4, May 1996.

[19] F. M. Chiussi, Y. Xia and V. P. Kumar. "Dynamic max rate control algorithm for available bit rate service in ATM networks," In *Proceedings of the IEEE GLOBECOM*, volume 3, pages 2108–2117, November 1996.

[20] R. Goyal, X. Cai, R. Jain, S. Fahmy and B. Vandalore, "Per-VC Rate Allocation Techniques for ATM-ABR Virtual Source Virtual Destination Networks," In *Proceedings of GLOBECOM*, November 1998. http://www.cis.ohio-state.edu/~jain/papers/globecom98.htm

[21] S. Fahmy, R. Jain, R. Goyal, B. Vandalore, S. Kalyanaraman, S. Kota and P. Samudra, "Feedback Consolidation Algorithms for ABR Point-to-Multipoint Connections in ATM Networks," *Proceedings of IEEE INFOCOM*, volume 3, pp. 1004–1013, April 1998. http://www.cis.ohio-state.edu/~jain/papers/cnsldt.htm

[22] N. Plotkin and J. Sydir, "Behavior of Multiple ABR Flow Control Algorithms Operating Concurrently within an ATM Network," In *Proceedings of IEEE INFOCOM*, April 1997.

[23] Q. Ma and K.K. Ramakrishnan, "Queue Management for Explicit Rate Based Congestion Control," In *Proceedings of ACM SIGMETRICS'97*, pp. 39-51, Seattle, Washington, June, 1997.

[24] H. Özbay, S. Kalyanaraman and A. İftar, "On Rate-Based Congestion Control in High Speed Networks: Design of an $\mathcal{H}^\infty$ Based Flow Controller for a Single Bottleneck," In *Proceedings of American Control Conference, 1998*. Through http://www.ecse.rpi.edu/Homepages/shivkuma/

[25] E. Altman, T. Basar and R. Srikant, "Multi-user rate-based flow control with action delays: a team-theoretic approach," In *Proceedings of 36th Conference on Decision and Control*, December 1997, pp. 2387–2392.

[26] L. Benmohamed and S. M. Meerkov, "Feedback control of congestion in packet switching networks: the case of a single congested node," *IEEE/ACM Trans. on Networking*, Vol. 1 (1993), pp. 693–707.

[27] C. Rohrs and R. Berry, "A linear control approach to explicit rate feedback in ATM networks," In *Proceedings of INFOCOM '97*, April 1997.

[28] R. Jain, "The Art of Computer Systems Performance Analysis," *John Wiley & Sons*, 1991. http://www.cis.ohio-state.edu/~jain/books/perfbook.htm

[29] B. Vandalore, R. Jain, R. Goyal and S. Fahmy, "Design and Analysis of Queue Control Functions for Explicit Rate Switch Schemes," In *Proceedings of IC3N*, October 1998. http://www.cis.ohio-state.edu/~jain/papers/qctrl_bv.htm

[30] S. Kalyanaraman, R. Jain, S. Fahmy, R. Goyal and B. Vandalore, "Analytical arguments for the convergence of the ERICA switch algorithm to max-min fairness: Single and multiple bottleneck cases" *Technical Report*, Ohio State University, 1998.

A Analytical Argument of Convergence to Max-Min Fairness

In this appendix, we give a limited analytical argument for the convergence of a single bottleneck node implementing ERICA towards max-min fair rate allocations. Some model assumptions in this argument are unrealistic, but they simplify the analysis considerably. This section should be used only to get further insights into the engineering decisions taken in the ERICA algorithm. We have not published an extension of this argument for a multiple-bottleneck system, but are exploring it under a new effort to cast such non-linear rate-based algorithms in a control-theoretic framework [24]. The general proof of convergence, stability and robustness (under assumptions of multiple bottlenecks, queuing delay targets and asynchrony) of rate-based algorithms is currently an open problem studied by several researchers (see [24, 25, 26, 27]).

Our modeling assumptions are:

- There is only one bottleneck node.
- The "core" ERICA algorithm (defined in section 4.2) is used with two exceptions:
  1. We ignore the effect of the queue control function.
  2. We ignore the moderation step (equation (11)):
     IF (ER $>$ FairShare AND CCR $<$ FairShare)
     THEN ER = FairShare
- Sources are persistent (always have data to send), though some (not all) might be source-bottlenecked at low rates.
- Round-trip times of various sources are different (the case of equal round trip times is a special case of this).
- The effect of any transient queuing between intervals is ignored. (unrealistic).
- Switch averaging interval is at least the twice the sum of **a)** the largest RTT of any VC though the bottleneck, and **b)** the maximum time required to see at least one RM cell of every active source (maximum inter-RM cell time). This means that measurements are reliable, and transient loads caused by asynchrony can be ignored. We call such an averaging interval a *"cycle"*.
- Load factor $(z) > 0$, ER $<$ Link Rate, and for any source CCR $<$ Link Rate. The last condition is satisfied since ATM signaling ensures that the "peak cell rate (PCR)" parameter is never larger than any link rate along the path
- Source-bottleneck behavior (if any) does not change during the convergence period.

Notation:

- Rate of source i in cycle j (CCR) is $R(i, j)$
- MaxAllocPrevious in cycle j is $Max_i\ R(i, j)$
- The ER for source i in cycle j is the same as the rate of source i in cycle $j+1$, i.e., $R(i, j+1)$
- z_j = overload factor measured in jth cycle (and used in $(j+1)$ th cycle).
- C: Target ABR capacity of the bottleneck.
- B: Sum of the rates of bottlenecked sources, also equal to $b \times C$, $b \leq 1$
- N: Number of active sources

Definition: A source is said to be *satisfied* at a given rate if it is bottlenecked elsewhere and cannot utilize higher rate allocations.

To prove: that for the system described above, the ERICA algorithm causes it to converge towards max min operation in at most O(log N) number of cycles.

Proof:

The proof methodology used here was proposed in reference [12]. We first prove a set of *safety (closure)* properties which show that the system remains within a closed state space, S. The closed state space, S is:
S: $0 < z < N$

Then we prove a set of *convergence properties* which show that the system reaches and remains in a target state space, T. The target convergence state space, T is:
T: $(1 \leq z \leq 1+\delta)$ AND Allocations are Max-Min fair,

where the term "Max-Min fair" implies that contending sources are allocated the highest possible *equal* rates, satisfying the condition on z.

Closure Properties:

Lemma 0: Given that the maximum rate (C) of any VC is at most the target link rate (condition imposed during ATM signaling and connection setup), the overload factor lies between 0 and N, where N is the number of VCs set up (assumed active).

Trivial based upon the assumptions. □

Convergence Properties:

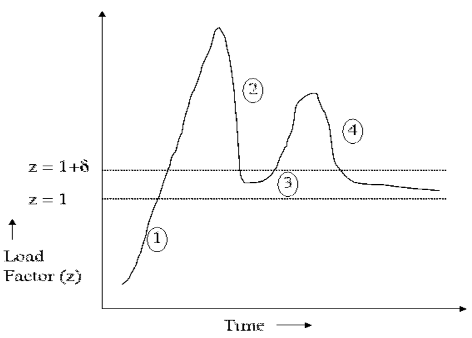


Figure 8: Movement of single bottleneck controlled by ERICA towards max-min fairness

Figure 8 shows how ERICA converges to max-min fairness under these model assumptions. It should be noted that this convergence property is what motivated the design of the components of the algorithm, based upon z and FairShare.

Specifically, the load factor (z) can fall into one of three zones $(0, 1), [1, 1+\delta], (1+\delta, N]$. The goal is to reach the second zone, while ensuring that the rate allocations are equal, i.e. the state space, T. The convergence unfolds in five stages as follows (of which stages 1 to 4 shown in the figure):

Stage 0: Irrespective of the initial rates, each source is given a chance to reach FairShare (C/N) in one cycle (Lemma 1).

Stage 1: Assuming stage 0 is the initialization of the algorithm, if the bottleneck is in the zone: $z \in (0, 1)$, within O(log N) cycles the system reaches a state where $z \geq 1$ (Lemma 2).

Stage 2: Once the system is in a state where $z \geq 1$, then the switch remains in such a state $z \geq 1$, and converges within O(log N) cycles to the state where $z \in [1, 1+\delta]$ within O(log N) steps. (Lemma 3, parts A and B)

Stage 3: When the system is the state where $z \in [1, 1+\delta]$, the contending sources get an equal rate allocation. (Lemma 3, part C).

Stage 4: The system may now stay in the state $z \in [1, 1+\delta]$, in which case max-min fairness is achieved (Lemma 3, Part C, Note 1). Alternatively it may move to a state $z \in [1+\delta, N]$, from where in O(log N) additional cycles it reaches the state $z \in [1, 1+\delta]$, but now since rate allocations are equal, rate allocations are unchanged and max-min fairness is achieved (Lemma 3, Part C, Note 1 and Theorem 1)

The details of the proof are presented below:

Lemma 1: ERICA takes one cycle to satisfy sources bottlenecked at rates below equal FairShare (C/N).

Proof: In every cycle, ERICA allocates at least FairShare = "fs" = C/N to every source. If there exist sources which are bottlenecked such that they cannot utilize rate allocations above fs, the system *satisfies* such sources in one cycle. This first cycle is called *"initialization cycle"* in what follows. □

Note 1: During convergence, there is at most one initialization cycle for any configuration.

Note 2: After the VCs below fs are satisfied, the unused capacity (if any) will be reflected in the value of the overload factor, z (which is the ratio of the total load and the target capacity).

Note 3: The following lemmas assume that the initialization cycle is completed, and that there is at least one "greedy" or "unconstrained" source going through each bottleneck which can utilize any bandwidth allocated to it.

Lemma 2: If a switch is underloaded, i.e., $z < 1$, then in O(log(N)) cycles, either the system converges to the target state space, T, or the load factor increases to reach a value greater than unity.

Proof:

During underload ($z < 1$), ERICA uses the following formula to allocate rates:
ER = Max(MaxAllocPrevious, CCR/z).

Recall that ER = $R(i, j)$, MaxAllocPrevious = $Max_i\ R(i, j-1)$, and $CCR/z = R(i, j-1)/z_{j-1}$

Hence, the ERICA formula can be rephrased as:

$$R(i,j) = Max(\ Max_i\ R(i,j-1),\ R(i,j-1)/z_{j-1}\ ) \quad (13)$$

Note that MaxAllocPrevious ($Max_i\ R(i,j-1)$) is at least C/N (equal to the maximum of the allocations in the previous cycle) and CCR/z is greater than CCR. As a result, the allocation of *every unsatisfied source increases.*

If all sources are greedy and and initially equal, the new load factor is unity, with all sources equal. *In this case the target T is achieved in a single cycle.*

In the case that source rate allocations are unequal and/or some sources are satisfied, the behavior of the system is different. Satisfied sources stay constant and the overload factor increases in the next cycle. If all sources are greedy, they get a rate of C/N in the first cycle. As a result, the new load factor is at least load/capacity = $(N \times (C/N))/C = 1$. *In this case, the load factor becomes greater than unity in a single cycle.*

We now show that even if the above special conditions do not hold, the load factor becomes greater than unity in O(log N) cycles. Assume that some sources at bottlenecked at rates below C/N, and the sum of their rates is B. The remaining sources get at least the maximum allocation of the previous cycle, i.e. $Max_i\ R(i,j-1)$. Starting from an initial load factor of z_0, the system increases its load factor in every cycle. Assume that, in the $(j-1)$ st cycle the overload factor, z_{j-1} is less than $1/(1+\epsilon)$, for small ϵ. Now,

$$z_j = \frac{B + \sum_i R(i,j)}{C} \geq \frac{B + \sum_i \frac{R(i,j-1)}{z_{j-1}}}{C} \qquad \text{from (13)}$$

$$\geq \frac{B + \sum_i \frac{R(i,j-2)}{z_{j-2} \times z_{j-1}}}{C} \geq \frac{B + \sum_i \frac{R(i,0)}{z_0 \times z_1 \times \ldots \times z_{j-1}}}{C}$$

$$\geq \frac{B + \sum_i \frac{R(i,0)}{(1/1+\epsilon)^j}}{C}$$

For z_j to become greater than 1, it is sufficient that:

$$B + \sum_i \frac{R(i,0)}{(1/1+\epsilon)^j} > C, \quad \text{i.e.,}$$

$$j < \left| \log_{1+\epsilon} \sum_i \frac{R(i,0)}{C-B} \right|$$

Since B and $R(i,0)$ are constants, and C is upper bounded by the link capacity $j = O(logN)$ in the worst case.

Note 1: z_j can also become greater than 1 when:

$B + \sum_i Max_i\ R(i,j-1) = B + (N - Nb) \times Max_i\ R(i,j-1) > C$, where Nb is the number of bottlenecked sources. Here, we have taken the $Max_i\ R(i,j-1)$ term in the ERICA step given in equation (13) instead of the $R(i,j-1)/z_{j-1}$ term which is used in the above proof. This new inequality reduces to:
$Max_i\ R(i,j-1) > (C-B)/(N-Nb)$

Observe that the right hand side of the above inequality is the target max-min rate allocation, which means that z_j becomes greater than unity in *one cycle* when *any one* of the rates $R(i,j-1)$ is greater than the final max-min allocation. Note that this assumes that the moderation step (see list of assumptions) has been ignored. □

Lemma 3: If a switch is overloaded, i.e., $z \geq 1$, then the switch remains overloaded i.e. $z \geq 1$, and converges within O(log N) cycles to the desired operating region, T.

Proof: We split the proof into three parts:
Part A: We first prove that the system remains in the region $z \geq 1$.

With the system starting at $z_{j-1} \geq 1$, we show that the minimum value of the new load factor after a cycle, z_j, is greater than or equal to unity.

The ERICA code segment used for this proof is:

IF $(z \leq 1+\delta) ER = Max$(MaxAllocPrevious,$CCR/z$)
ELSE $ER = Max(C/N, CCR/z)$

We argue that the ER value obtained by the assignment statement $ER = Max$(MaxAllocPrevious,CCR/z) does not reduce the load factor below its current value. Recall that MaxAllocPrevious $= Max_i\ R(i,j-1)$ and $CCR/z = R(i,j-1)/z$. Now, since $z \geq 1$, MaxAllocPrevious $\geq CCR/z$. As a result, this term is not going to reduce z. Therefore, we simply deal with the second assignment statement in the ERICA code segment above, i.e., $ER = Max(C/N, CCR/z)$.

Split the set of sources into two categories:

1. Sources bottlenecked at rates equal to or below C/N, which have a total rate of $b \times C, b \geq 0$.

2. Sources above C/N, with a total rate of $d \times C$.

The current load factor is $z_{j-1} = ((b+d) \times C)/C = b+d > 1$. If all sources were to divide their rates by z_{j-1}, the new load factor z_j would be unity. In our case only sources above C/N reduce their rates. The new load factor is $b + (d/z_{j-1})$. To complete the proof of part A, note that:

$z_j = b + (d/z_{j-1}) \geq (b+d)/z_{j-1} = 1$ □

Part B: In the worst case, the system first reaches the region $1 \leq z < 1+\delta$ in O(log N) cycles.

If the system is already in region $1 \leq z < 1+\delta$, the proof is trivial.

Else, let the initial load factor be z_0 and the current load factor be z_j. Let $B = b \times C$ be the sum of bottlenecked rates at or below C/N. The remaining rates $R(i,j) \geq C/N$, and $z_k > 1+\delta$, $\forall k < j$. A technique similar to the one shown in lemma 2 can be used to prove that $j = O(logN)$, i.e., the system reaches the operating region $1 \leq z < 1+\delta$ in O(log N) cycles. □

Part C: The contending sources get an equal rate allocation in the region $1 \leq z < 1 + \delta$

The ERICA allocation in this region (in the $j+1$ th cycle) is: Max(MaxAllocPrevious, CCR/z) i.e. $Max(\ Max_i\ R(i,j),\ R(i,j)/z_j\ )$

Since $z_j \geq 1$, $R(i,j)/z_j < Max_i\ R(i,j)$, and the ERICA allocation is simply $Max_i\ R(i,j)$ for all sources. In other words, the rate allocations to all sources in this region are equal.

Note 1: Observe that if $R(i,j)$ s were already equal, the load factor would be unchanged in subsequent cycles, i.e., the system would remain at $1 \leq z < 1 + \delta$, and rates of contending sources $R(i,j)$ are equalized, leading to max-min fair allocations. That is, the system has reached the state, T, and stays in this state until new input changes occur.

If the rates $R(i,j)$ are not equal before this *"equalization cycle"*, the new load factor can be greater than $1 + \delta$. As proved in part B, the system requires at most O(log N) cycles to converge to the state where $1 \leq z < 1 + \delta$. However note that at every cycle of this aforementioned convergence process, all rate allocations remain equal since they are scaled by the same factor (z). This implies that the system has reached a state where $1 \leq z < 1+\delta$ AND all rate allocations of unconstrained sources are equal. But this state is the same as the target state space, T.

Theorem 1: From an arbitrary initial state, the ERICA algorithm brings the system to the target operating region T within O(log N) cycles.

An arbitrary initial state can be characterized by a value of the load factor z between 0 and N (closure, lemma 0). If $z < 1$, we have shown in lemma 2 that the system reaches a state where $z \geq 1$ within O(log N) cycles. Once $z \geq 1$, we have shown that the load factor does not reduce below unity (lemma 3, part A). Further, the system moves to the region $1 \leq z < 1 + \delta$ within O(log N) cycles (lemma 3, part B) and the rates are equalized in a single in this region (lemma 3, part C). The system may now remain stable in the region $1 \leq z < 1+\delta$, with equal rates (i.e. max-min fair allocations), or move out of the region and converge back and remain in this region in O(log N) cycles with the rates being equal at every cycle during this convergence process (lemma 3, part C, note 1). This final region of stability is in fact the target state space, T, i.e., $1 \leq z < 1 + \delta$, and allocations are max-min fair. The maximum number of cycles to converge to T from an arbitrary initial state is O(log N). $\square$