A Survey of the Use-It-Or-Lose-It Policies for the ABR Service in ATM Networks

Shiv Kalyanaraman, Raj Jain, Rohit Goyal, Sonia Fahmy
Department of Computer and Information Science
The Ohio State University
2015 Neil Ave., Columbus, OH 43210-1277
E-mail: {*shivkuma, jain, goyal, fahmy*}@cis.ohio-state.edu

Abstract

The Available Bit Rate (ABR) service has been developed to support data applications over Asynchronous Transfer Mode (ATM). The ABR service uses a closed-loop rate-based traffic management framework where the network divides available bandwidth among contending sources. The ATM Forum then worked on incorporating open-loop control capabilities to make the ABR service robust to temporary network failures and source inactivity periods. One of the problems addressed was whether rate allocations of sources should be taken away if sources do not use them. The proposed solutions, popularly known as the Use-It-or-Lose-It (UILI) policies, have had significant impact on the ABR service capabilities. In this paper we survey the design, development, the final shape of these policies and their impact on the ABR service.

Keywords: Asynchronous Transfer Mode (ATM), Available Bit Rate (ABR), traffic management, congestion control.

1 Introduction

ATM networks provide five classes of service [23]: constant bit rate (CBR), real-time variable bit rate (VBR-rt), non real-time variable bit rate (VBR-nrt), available bit rate (ABR), and unspecified bit rate (UBR). The VBR and CBR classes are higher priority classes used to transport real-time or high quality audio and video data. Hence, a switch first allocates link bandwidth to these classes. The remaining bandwidth, if any, is given to ABR and UBR traffic. Data traffic uses either the ABR or the UBR service.

The ABR service class has built-in traffic management mechanisms which rapidly allocate the ABR bandwidth among active sources. Applications using this service class expect guarantees regarding cell loss, but can control their data rate dynamically as demanded by the network [1]. The ABR service is an also an efficient way of carrying data traffic over an ATM backbone network.

The UBR service class, on the other hand, does not have any standard traffic management support and does not give any quality of service (QoS) guarantees to applications. UBR is useful for data applications which have no requirements regarding cell loss, but can use the network bandwidth whenever it is available. E-mail traffic is an example UBR application. In this paper, we concentrate on the ABR service and its traffic management mechanisms.

The ATM Forum Traffic Management group adapted a rate-based end-to-end framework to control ABR traffic. In this framework, sources periodically transmit Resource Management (RM) cells. Switches indicate single-bit or multiple-bit (explicit rate) feedback using these RM cells asking sources to control their rates. The destinations simply return RM cells back to the source.

The ABR traffic management model is a closed-loop model, i.e., the sources normally change their rates in response to feedback from the network. Another form of control, called open-loop control, is when sources can independently change their rates without feedback from the network. Open-loop control is useful when the network delays are large compared to application traffic chunks, or when network feedback does not reach sources. It is also useful to control applications which are *bursty*, i.e., alternate between active periods (have data to send) and idle periods (have no data to send). Further, some applications might not always be able to sustain a data rate as high as the network allocated rate. Open-loop control can be used along with the default closed-loop mechanism to improve performance under such conditions.

In this paper, we discuss some of the efforts of the ATM Forum to provide open-loop control support in the ABR service. Specifically, we discuss the issue of whether rate allocations of sources should be taken away if they do not use them. The proposed solutions, popularly known as the Use-It-or-Lose-It (UILI) policies, have had significant impact on the ABR service capabilities. We survey the design, development, the final shape of these policies and their impact on the ABR service.

The paper is organized as follows. Section 2 gives an overview of the rate-based model and introduces the terminology used. Section 3 discusses the problems of supporting bursty traffic in this model and introduces the idea of a Use-It-Or-Lose-It (UILI) policy. Section 4 discusses the specific issues in the design of UILI policies. We then historically survey the UILI policies. Section 5 surveys the early UILI proposals. Section 6 explains the problems with these early approaches. Section 7 surveys the new UILI approaches. These approaches were debated in the ATM Forum and a choice was made. Section 7.6 summarizes the implications of the UILI policy choice on the ABR service.

2 ABR Rate-based Traffic Management Model

This section gives a brief overview of the ATM Forum's rate-based traffic management model and introduces the terminology used in the rest of the paper.

The main components of the framework are the source end system (SES), switch, and the destination end system (Figure **??**). The ABR Source End System (SES) sends data at the Available Cell Rate (ACR) which is less than a negotiated Peak Cell Rate (PCR). Immediately after establishing a connection, ACR is set to an Initial Cell Rate (ICR), which is also negotiated with the network. The source sends an RM cell after transmitting Nrm-1 cells (default Nrm value is 32). The RM cells are turned around by the Destination End System (DES). The RM cells collect the network feedback and return to the SES, which responds to the feedback.

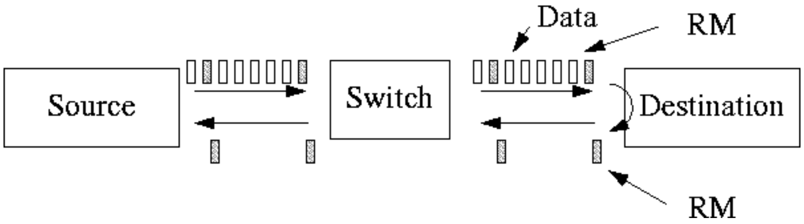


Figure 1: ABR Traffic Management Model: Source, Switch, Destination and Resource Management Cells

The important RM cell fields are: the Current Cell Rate (CCR) field which is initialized with the current ACR, an Explicit Rate (ER) field, a Congestion Indication (CI) and a No Increase (NI) bit. The ER, CI and NI fields are used by the network to give its feedback. When there are multiple switches along the path, the feedback given by the most congested link is the one that reaches the source.

Data cells also have an Explicit Forward Congestion Indication (EFCI) bit in their headers, which may be set by the network when it experiences congestion. The DES saves the EFCI state of every data cell. If the EFCI state is set when it turns around an RM cell, it uses the CI bit to give (a single bit) feedback to the source. When the source receives the RM cell from the network, it adjusts its ACR using the ER, CI, NI values, and SES parameters.

3 Bursty Traffic and Rate-Based Control

ABR Traffic management has to perform for the following workload conditions:

- The ABR capacity may be highly variable due to the presence of higher priority CBR or VBR classes. Further, high link transmission speeds (155Mbps and above) imply that the bandwidth needs to be allocated quickly. Older data networks did not have to deal with capacity being variable since they did not support multiple service classes. Further, transmission speeds were lower in such networks and a longer delay.

  ABR switch algorithms like ERICA and ERICA+ [22] have been developed which quickly allocate the available capacity among the active sources.

- The ABR sources' demand may be constant or highly variable. Applications like file transfers generate a constant demand, whereas, client-server and web traffic have variable demand, because such traffic has active periods, where data is sent using the available rate, and idle periods, where the available rate goes wasted because the source does not send data.

  Traffic sources which have active and idle periods are called "bursty traffic" sources, whereas traffic sources which can use their rate allocation at any point of time are called “infinite traffic” sources. There exists an intermediate kind of traffic sources which does not have idle periods like bursty traffic sources, but may not use the entire rate allocated to it, like infinite traffic sources. Such traffic sources are termed “source bottlenecked.”

The ABR Source End System (SES) transmits data and RM cells during the active periods. For bursty traffic, during any idle period no data cells are sent and hence no RM cells are sent. However, some RM cells sent during the previous active period are present in the network. Feedback from the switches to the sources continues as long as no RM cell from this source remains in the network. If the source is still idle after all its RM cells are exhausted, it loses contact with the network. Its ACR has a value set according to the latest feedback (the last RM cell) received from the network.

When the source transmits data after such an idle period, there are two problems. First, the source is using a stale value of ACR to transmit its data (network conditions may have changed). Second, this source will not receive feedback from the network for one round trip time, which is the time taken by a cell to traverse the path from the source to the destination and back. These conditions may lead to a potential overload situation for the network lasting upto one round trip time.

A solution to this problem is to detect an idle source and reduce its rate allocation before it can overload the network. But, this has an important side effect - if the rates are reduced after every idle period and the active periods are short, the aggregate throughput experienced by the source is low. This tradeoff was discovered and studied carefully in the ATM Forum. The solutions proposed are popularly known as the Use-It-or-Lose-It (UILI) policies, referring to the fact that the source's ACR is reduced (lost) if it is not used (for example, the source is idle). In the following sections we trace the design and development of these policies.

4 Issues in Use-It-Or-Lose-It

The ABR specification of December 1994 allowed RM cells to be sent during active and idle periods as follows:

“The Source shall send an RM cell for at least every Nrm cells transmitted, and also at least every 100ms if data is being transmitted. If more than 100ms have transpired since the last RM cell then an RM cell should precede any data cells and ACR sould be set to the minimum of ACR and ICR.”

This statement implied that RM cells would normally be sent whenever data is sent. If the source did not send data for a while, it also did not send RM cells for the same period. However, if the source did not send an RM cell for 100ms, it was considered an “idle” source. An idle source becoming active again had to precede its data cells with an RM cell. This would allow the RM cell to probe the network and bring the feedback faster than waiting for the normal turn of the RM cell.

Barnhart [2] observed that the specification had the following problems:

- **Idle Period Definition**: As mentioned above, as source is considered idle if it does not send an RM cell for 100ms . The value of 100ms as idle period is too long to provide cell loss protection for OC3 rates.

- **ACR Retention**: This is the problem when a source with high ACR becomes idle, i.e, it does not send data cells or RM cells into the network. The network conditions may change during this idle period and there may be no opportunities (RM cells) to inform the source of such changes in network conditions.When such sources suddenly use their ACR, the network experiences overload.

- **ACR Promotion**: This is the problem where a source sends cells at rate S smaller than ACR. A source may intentionally or unintentionally behave this way, or it may be bottlenecked at the Network Interface Card (NIC, which is the SES) at rate S. When there is no congestion, the network redistributes the unused bandwidth. Under such circumstances, the source may get an allocation higher than its current ACR value. Moreover, it may get an allocation which is higher than what it would have gotten if its source rate, S, was equal to ACR (because S being smaller than ACR is partially responsible for the network underload). Since the source is still bottlenecked at a rate S, the ratio ACR/S increases. When such sources suddenly use their ACR, the network experiences overload.

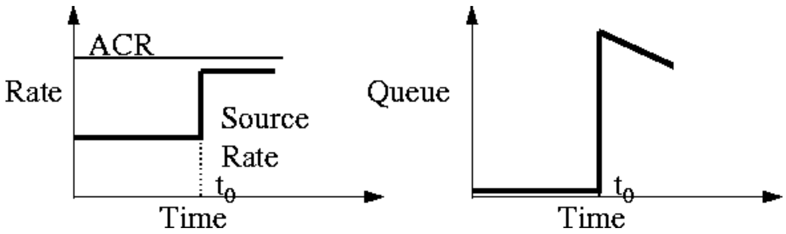


Figure 2: Effect of ACR Retention/Promotion

Note that both ACR retention and ACR promotion imply that the source rate is smaller than the ACR allocation. This results in the network receiving an unexpected burst of cells when the source rate increases to ACR. We henceforth use the words ACR retention and ACR promotion (preferring the former) interchangeably to mean that the source rate is smaller than ACR.

Figure 2 shows the typical behavior of source rates and network queues due to ACR retention. Before time t_0, the source rate is much smaller than its ACR allocation. The ACR allocation remains constant. At time t_0, the source rate rises to ACR and the network queues correspondingly rise. Jain et al [14] note that the maximum queue size is $(ACR - SourceRate) * Feedbackdelay * (\#sources - 1)$. Each of the three factors (ACR-SourceRate, Feedback delay, number of sources) can be high and increase the maximum queue size.

In the above expression, *feedback delay* is the time lag from the point of the network giving feedback to the point when the effect of the feedback is felt in the network. For a source sending data after a long idle period, the feedback delay equals the round trip time. For source which are active, the feedback delay is approximately the sum of the propogation and transmission delay from the switch to source, and that from the source to the switch.

The solution to these problems was the so called "Use-It-Or-Lose-It" (UILI) policy. An UILI policy specifies a method to detect ACR retention in sources and a method to reduce their allocations under when ACR retention is detected.

4.1 Features of Use-It-or-Lose-It Proposals

UILI proposals detect ACR retention by measuring the source's actual rate and comparing the measured source rate with the allocated ACR. If the source rate is lower than the ACR by a factor called the "headroom", the source is declared to be an ACR retention source and UILI reduction is applied. A switch which allocates rates proportional to ACR may allocate a rate proportional to the smaller measured source rate.

The proposals for UILI differ in:

Where the ACR retention is detected and the ACR reduction is done: These functions may be implemented at the SES (source-based) or at the switch (switch-based) or a combination of both methods.

How ACR retention is detected: the test for ACR retention involves a factor which can be additive or multiplicative. Implicit detection is also possible, for example, in switches which allocate rates proportional to the source rates.

How ACR reduction is done.

5 Early UILI proposals

The early UILI proposals were all source-based proposals. In a source-based proposal, the test for ACR retention done when the RM cell is being sent by the source. If ACR retention is detected, the ACR is immediately reduced using the reduction algorithm, and the new value of ACR is carried in the CCR field of the RM cell being sent. Some proposals also ignore the next feedback from the switch (if the feedback requests a rate increase) to prevent the ACR reduction to be overridden by a network feedback.

5.1 February 1995 Proposals

Based on Barhart's contribution [2], the ABR source specification in February 1995 read as follows:

"1. The source shall send an RM-cell for every Nrm cells transmitted, and also at least every 100ms if data is being transmitted. If more than 100ms have transpired since the last RM

cell was sent, then an RM-cell should precede any data cells."

...

*"4. Before sending a forward RM-cell if the time T which has transpired since the last time a backward RM cell was received or since the last decrease, is greater than TOF*Nrm cell intervals (1/ACR), then the source shall reduce its ACR by the number of cell intervals in the time T, (T*ACR) multiplied by ACR/RDF, down to ICR."*

The December 1994 rule (see section 4) was separated into two rules which addressed low rate sources and ACR retention sources respectively.

Low rate sources send RM cells infrequently and cannot receive rate increase feedbacks rapidly. Hence, it was decided to allow such sources to send an RM cell every 100ms.

The latter rule is targetted at the ACR retention problem. ACR retention is declared under one of the following two circumstances (assuming that the source is not a low rate source). First, the source does not hear from the network, but has sent out k RM cells. Here, k is the Time out Factor (TOF). Second, the source has not decreased its rate, but has sent out k RM cells.

The ACR reduction function used is:

$$ACRn = ACRo(1 - T * ACRo/RDF)$$

ACRn is the new ACR and ACRo is the old ACR. T is the time since the last RM cell was sent or a rate decrease was last done. RDF is the rate decrease factor also used for multiplicative decrease to support EFCI feeback. RDF is the rate decrease factor. This parameter is normally used to calculate the new rate given EFCI feedback. However, it was reused in the reduction formula to avoid choosing a new parameter. This formula decreases ACR as a linear function of time.

5.2 April 1995 Proposals

In April 1995, several flaws were discovered in the February proposal.

5.2.1 Test for ACR Retention

First, it was observed that the UILI detection method used the time measured since the last ACR decrease. This method would trigger UILI if the source does not decrease its ACR for a period of time. Such triggering would hamper sources from increasing their rates rapidly, leading to underutilization. Hence, this part of the condition was removed.

Second, the UILI detection depends on when the last BRM was received. A source may be continuously receiving BRM cells, but still not using its ACR allocation (an ACR retaining source). Hence, it was decided to make the test dependent on the time since the last RM was *sent*, and independent of when a BRM is received.

Roberts [4] noted that the purpose of the February 1995 test depending upon the last BRM received was not ACR retention, but to protect the network when no feedback (RM cell) returns to the source. He argued that this functionality deserved a separate SES rule, and associated parameters. For example, the switch could specify a parameter, (later) called Transient Buffer Exposure (TBE) which limits how many cells an idle source could send at a high rate after it became active. We discuss the effect of this rule and its parameters elsewhere [26].

5.2.2 ACR Reduction Function

The proposed reduction function was also found to be unsatisfactory. Barnhart [3] observed that the February 1995 reduction function had a very sharp slope and would reduce ACR too fast. In fact, an idle period of 1.4ms can force the same decrease as backward RM (BRM) cells with CI=1 (actual network congestion) can force in 14ms.

Hence a new decrease formula was proposed which reduces the ACR in a harmonic manner:

$$1/ACRn = 1/ACRo + T/RDF$$

The decrease curve for ACR approximates the decrease due to actual network congestion (BRM cells with CI=1). We henceforth call this latter curve as the "CI-decrease curve."

Further, it was also observed that the reduction statement says: "... multiplied by ACR/RDF, down to ICR."

For high ICR sources, it is possible that ACR may be lower than ICR and hence, the above statement does not make sense. This was clarified by using a second decrease formula when ACR was below ICR.

The next observation made was that ACR reduction is done when an RM cell is sent. This reduction may be immediately undone by the next BRM which requests a rate increase. Hence, the feedback of the immediately next BRM was decided to be ignored if such a reduction has been done.

5.2.3 The Updated UILI Policy

The UILI rule was hence reformulated in April 1995 as follows:

*5. "Before sending a forward RM-cell, if the time T which has transpired since the last time a forward RM cell was sent is greater than TOF*Nrm cell intervals (of 1/ACR), then:*
a. if ACR > ICR, then the source shall decrease its rate by adding at least T/RDF to its cell transmission interval (1/ACR), down to ICR.

*b. if ACR <= ICR then the source shall decrease its ACR by at least ACR*Nrm/RDF, down to MCR.*

c. the source shall not increase its rate upon the reception of the next RM cell."

5.3 May and June 1995 proposals

Kenny [5] argued that item 5b. of the April 95 SES specification be eliminated to avoid its side effects:

5.3.1 Idle Sources vs Low Rate Sources

First, it discriminates against sources which are source-bottlenecked, compared to sources which are idle.

An idle source differs from a source-bottlenecked source in that the former does not send RM cells or receive feedback, whereas the latter sends data and RM cells, but at a rate smaller than the allocated ACR. Jain et al [14] later observe that a source may be bottlenecked at the SES because there are multiple source sharing sharing the same outgoing link from the SES. The sum of the rates of the sources cannot exceed the ABR bandwidth of the outgoing link although the sum of the sources' ACR may exceed ABR bandwidth.

A source-bottlenecked source may trigger the UILI detection multiple times whenever it sends RM cells and its source rate is lower than ACR. For example, if the source rate is lower than ICR (possible for high ICRs), then the ACR may be reduced below ICR.

However, an idle source does not send RM cells and hence ACR Retention (UILI) is never detected till the next RM cell is sent. Hence, it is possible that a source can go idle for a long time when its ACR is greater than ICR. When it becomes active, the UILI triggered when the first RM cell will not reduce its rate below ICR. If the source subsequently sends data at ICR, no further reduction is seen. It was observed that the idle source was similar to a source bottlenecked source with a rate much smaller than ICR, but due to its bursty behaviour gets a higher ACR than the source-bottlenecked source.

5.3.2 Other Side Effects

Second, sending the very first RM cell of any source causes ACR to go below ICR which is against the design specification.

Another minor side effect is due to item c). After a long idle period, ACR retention is detected and ACR is reduced. Item c. is enabled after the reduction of ACR. Item c. disallows any ACR increase for a period of one interRM cell interval. The interRM cell interval is proportional to the rate of the source. Now, if the idle periods are long, the ACRs will be reduced to ICR, which may be very small. Since the feedback due to the first RM cell is ignored, the next RM cell might reach the source after a long time due to ICR being small. However, in practice the effect of item c. was mixed, i.e., its inclusion does reduce ACR Retention from 6:1 to 3:1 in many experiments [5].

5.3.3 A New ACR Reduction Formula

The ACR reduction formula was again modified based on the proposal by Kenny et al [6] for simpler implementation. Kenny et al observe that, while the April reduction formula has no multiply operations, it is difficult to implement, inconsistent with other reduction formulas, and is inflexibly tied to the RDF parameter. Finding 1/ACR requires reciprocation which is difficult. A second reciprocation operation is required to get ACR from 1/ACR. Further, 1/ACR represents an interval used by the cell scheduler. Schedulers operate based on integral cell slots. A floating point multiplication may be required to get the scheduling interval from ACR.

The formula suggested was :

$$ACRn = ACRo(1 - T * TDF)$$

where TDF is a new parameter called "Timeout Decrease Factor".

This represents a linear decrease with respect to the idle time. The slope of this line is determined by TDF. Further, TDF is independent of ACR. Hence, the formula no longer approximates the CI-decrease curve, as done by the February 1995 formula. However, since the slope of the curve can be varied depending upon TDF, and the straight line will intersect the CI-curve at different points depending upon the value of TDF. The intersection point being far away implies that the ACR decreases slowly as the idle time increases. Based on these considerations, TDF was recommended to be replaced by ICR/RDF, instead of using a new parameter. The group, however, decided to use linear curves with more aggressive slopes than ICR. Specifically, it decided to use multiples of PCR/RDF, where the multiplying factor is a parameter called TDFF (signalled version of TDF).

Further, as suggested in [8], the UILI rule applies as long as ACR is greater than ICR, and that the ACR should not be decreased below ICR.

5.3.4 The Updated UILI Policy

Incorporating the aforementioned changes, the specification in August read as follows:

*"5. Before sending a forward in-rate RM-cell, if the time T that has elapsed since the last in-rate forward RM-cell was sent is greater than TOF*Nrm cell intervals of (1/ACR), and if ACR > ICR, then:*

*a) ACR shall be reduced by at least ACR * T * TDF, unless that reduction would result in a rate below ICR, in which case ACR shall be set to ICR, and TDF is equal to TDFF/RDF times the smallest power of 2 greater or equal to PCR,* $TDFF = \{\ 0, 2^i, 2^j, 2^l\ \}$ *(2 bits), where the values of the integers* $i, j,$ *and* l *are to be determined in the specification.*

b) ACR shall not be increased upon reception of the next backward RM-cell."

The above UILI rule will also be interchangeably called "rule 5" henceforth, referring to the rule number in the ABR SES specification. The two parts are called "rule 5a" and "rule 5b" respectively.

6 Problems and Side Effects of Early Proposals

In August 1995, Anna Charny et al [7] pointed out certain undesirable side effects of the July 1995 UILI proposal. In particular, sources experience performance degradation in the transient phase when they increase from low ACR to high ACR. As a result the links are underutilized for a long period of time.

6.1 Worst Case Performance

The worst case occurs when the BRM cell is received just before an FRM is sent. The BRM carries the network feedback and asks the source to increase its rate (to a value greater than TOF*(old rate)). When the FRM is sent, the source triggers UILI (rule 5) and attempts to decrease ACR by ACR * T * TDF. Now, ACR is the new (high) value due to network feedback, and T is inversely proportional to the earlier rate; a smaller rate implies a larger decrease. Typically, ICR is larger than the value resulting from this formula. Hence the ACR drops to ICR. If ICR itself is small and the BRM is again received just before an FRM is sent, the cycle repeats. In effect a source starting from a low ICR may never send at a rate higher than ICR.

6.2 Bursty and RPC Traffic Performance

In fact, it was observed in simulations of Charny et al [7] that bursty traffic having low ICR experienced a long term performance degradation. The effective rate of transmission fluctuates between a low and a high value. The rule 5b which prevents the increase of the source rate even though the network may have bandwidth available. In such bursty traffic configurations, it was found that rule 5a without rule 5b yielded better performance than both the parts together. Further it was not possible to optionally turn off rule 5 completely because rule 5b would still exist. Hence, it was decided to introduce a PNI (Prohibit No Increase) bit, which when set, turns off rule 5b selectively or rule 5 completely (if TDF is also zero).

The performance degradation due to RPC (ping-pong) type traffic was independently observed by [8]. These authors pointed out that such applications may not want their rates to be decreased or reset to ICR after every idle period. They also suggested that UILI be performed by the switch and leave the source-based UILI optional.

These side effects of rule 5 are not seen when the source is in the steady state (with source rate approximately equal to ACR) or in the transient phase when the source is decreasing from a high ACR to a low ACR. The main problem seemed to be due to the fact that the decrease function was proportional to T resulting in large ACR decreases after an ACR increase, leading to ramp-up delays.

6.3 The Rescheduling Option

The problem of large T is alleviated by the rescheduling option originally introduced in the context of low rate sources [9]. This option allows data cells to be scheduled as soon as possible when the rate is increased, instead of waiting for the scheduling slot determined at the old (low) rate. The time variable T is reduced by at most one inter cell time at the old rate. This change makes a difference when the old rate (or ICR) is very small.

6.4 Parameters

Another problem which emerged was that the parameters RDF, ICR, and TBE were being used in multiple rules. Hence, choosing optimal values for these parameters became difficult due to their various side effects.

These problems were addressed in the new set of proposals in December 1995 when the issue was voted upon to arrive at a final decision.

7 December 1995 Proposals

There were three main proposals in December 1995: the time-based proposal [11, 12, 13], the count-based proposal [14], and the switch-based proposal [15]. The first time-based and the count-based proposals were later combined into one joint proposal. The ATM Forum voted between the switch-based proposal and the joint source-based proposal.

7.1 Outstanding UILI issues

The following were the outstanding issues in UILI in December. Essentially, a UILI proposal which works for both source-bottlenecked and bursty sources was desired.

- How to avoid UILI affecting the normal ACR increase (ramp up) of sources ?
- How long should the switch feedback be ignored after an ACR adjustment ?
- How to ensure good throughput and response time for bursty sources having small, medium and large active periods, when the idle periods are small, medium or large ?
- The floor of the August 1995 UILI ACR reduction function is ICR. If the source rate, S, is larger than ICR, the ACR may be reduced below the source rate down to ICR.

We want a reduction function which does not decrease the ACR below the source's rate, S.

- "Headroom" measures how much the ACR is greater than the source rate, S, when it is declared as not an ACR retaining source. Should the headroom be multiplicative (ACR $\leq$ TOF*S) or additive (ACR $\leq$ S + headroom) ? Is a separate headroom parameter necessary (to avoid depending on ICR) ?
- Can UILI be done effectively in the switch ?
- Under what circumstances is UILI unnecessary or harmful ?

7.2 Count-Based UILI Proposal

The count-based UILI proposal [14] solved a large subset of these problems and presented results of an extensive study on bursty traffic behavior.

7.2.1 Count vs Time

First, the count-based proposal removes the dependency of the ACR reduction function on the time factor, T, which is the time since the last FRM is sent. That is, the formula

$$ACR = ACR - ACR * T * TDF$$

is replaced by:

$$ACR = ACR - ACR * TDF$$

A fixed ACR decrease is achieved by triggering UILI n times. Time-based UILI decreases the ACR proportional to the time factor, T.

Note that the UILI test function still depends upon source rate and hence the time T.

7.2.2 Multiplicative vs Additive Headroom

The count-based proposal uses an additive headroom, as opposed to a multiplicative headroom used in the August 1995 proposal. The multiplicative headroom results in a large difference between ACR and source rate, SR, depending upon the value of SR. Recall that if the ACR of the source is within the headroom, UILI is not triggered. A large difference (ACR - SR) might result in large queues, when the source suddenly uses its ACR. The additive headroom allows only a constant difference (ACR - SR) regardless of the source rate, SR. The queue growth is bounded by a constant, $(ACR - SR) \times FeedbackDelay \times NumberofSources$. Hence, the additive headroom provides better network protection than the multiplicative

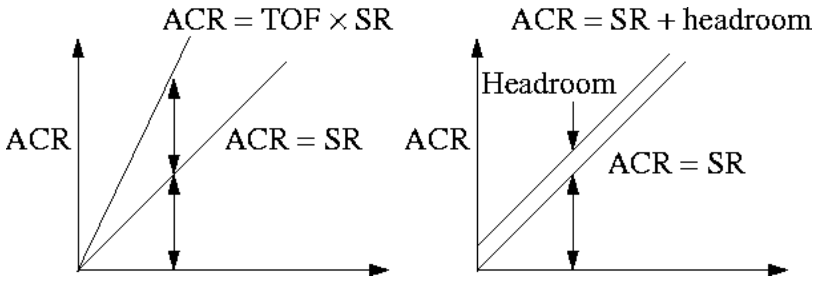


Figure 3: Multiplicative vs Additive Headroom

headroom. The difference between the multiplicative and additive headroom is shown in Figure 3

Further, it is easier to implement the additive headroom as opposed to the multiply option since fewer multiply operations are required.

The count-based UILI test uses the additive headroom technique. The use of additive headoom is a common feature with the time-based UILI proposal discussed in Section 7.3.

7.2.3 Floor of the ACR Reduction Function

The count-based proposal also observed that the floor of the August 1995 UILI ACR reduction function is ICR. If the source rate, S, is larger than ICR, the ACR may be reduced below the source rate down to ICR. The proposal uses a different floor function which ensures that the ACR does not decrease below source rate, SR, or the headroom. If the headroom equals the ICR, then the ACR is guaranteed not to decreased below ICR.

The floor of the reduction function used is: Source Rate + Headroom.

The use of new floor of the reduction function is a common feature with the time-based UILI proposal discussed in Section 7.3. The time-based UILI proposal also suggests an alternative floor for the reduction function.

7.2.4 Normal Rate Increase (Ramp Up)

The August 1995 proposal inhibited the ramp up or normal rate increase phase because it triggered UILI immediately (when the first FRM is sent) after the rate increase (due to a BRM received). Since the decrease function is proportional to time, and has the floor ICR, the amount of decrease could be large.

The count-based proposal decreases ACR by a step Δ = ACR*TDF. The next BRM cell brings the rate back to the ACR value before the decrease. For small values of TDF, UILI is no longer triggered. This is because the rate has increased only by the same step Δ which is smaller than the headroom.

However, for larger values of TDF, UILI may be triggered multiple times. The new floor function ensures that the source rate consistently increases by at least the "headroom" value and eventually UILI is no longer triggered.

The count-based proposal demonstrates a technique which avoids all oscillations due to normal rate increase: The UILI test is disabled exactly once after a normal rate increase. This is achieved using a bit called the PR5 ("Prohibit Rule 5") bit which is enabled whenever there is a normal rate increase. The bit is cleared otherwise.

This technique also has one important side effect: if a source which is not ACR retaining suddenly goes idle, and the network feedback requests a rate increase during the idle period (using the RM cells remaining in the network), the UILI test is disabled the when the source suddenly becomes active again and the first FRM is sent. We note that the first FRM cell opportunity is the only opportunity to reduce the source's ACR using UILI, because in when the next FRM is sent, the memory of the prior idle period is lost. Hence, in such cases UILI is never triggered. As mentioned, this technique is not necessary if TDF is small. Hence TDF can be set to small values and the technique not used.

7.2.5 Action on BRM

The count-based proposal observed that the ACR reduction function alone is not enough to ensure that that ACR retention is eliminated. The August 1995 proposal requires that if the immediately next BRM feedback, after an UILI ACR reduction, requests a rate increase, and the PNI bit is not set, the BRM feedback is ignored. However, subsequent feedbacks may undo the ACR reduction and the problem of ACR retention still persists.

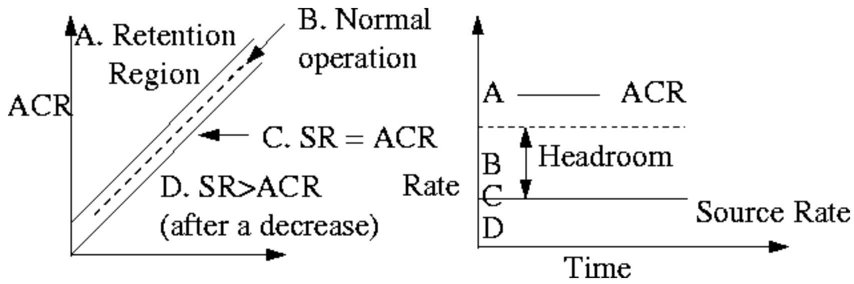


Figure 4: Regions of Operation

The count-based proposal ignores the BRM feedback as long as the source does not use its ACR allocation. The proposal uses the headroom area as a hysterisis zone in which network feedback to increase ACR is ignored. The proposal defines four regions of operation A, B, C, and D, as shown in Figure 4. Region A is called the ACR retention region. In this region, $ACR > SR + Headroom$, and UILI is triggered unless the PR5 bit (if used) is set. Region B is the headroom area. In this region, $ACR \leq SR + Headroom$, but $ACR > SR$. In this region BRM feedback requesting increase is ignored. Region C has the source rate equal to ACR. Region D has source rate greater than ACR. Region D is touched briefly when the ACR decreases, but the measured source rate is a mixture of the old and new ACRs. In regions C and D, the source obeys the feedback of the network to increase or decrease its ACR. In these regions, the source is not ACR retaining because its source rate is at least

Table 1: BRM Actions In The Different Regions Of Count-Based UILI

Region	Trigger Rule 5	Increase On BRM	Decrease On BRM
A	Yes unless PR5	No	Yes
B	No	No	Yes
C	No	Yes	Yes
D	No	Yes	Yes

equal to its current ACR allocation. The actions in various regions are shown in Table 1.

Note that there is no need for the PNI parameter, since UILI can be disabled by simply setting the parameter TDF to zero.

7.2.6 Pseudo Code For the Count-Based Proposal

In the pseudo code for the count-based proposal given below, the variable 'ACR_ok' indicates that the source has used its allocated ACR, and is allowed to increase its rate as directed by network feedback. 'PR5' when set conveys the fact that the network has just directed an increase. PR5 ('Prohibit Rule 5) is named similar to PNI (Prohibit No Increase). 'SR' is a temporary variable and is not stored between successive execution of the code. Further, the proposal requested a separate parameter 'headroom' instead of using ICR in the UILI formula.

We present two sets of the pseudo code, the first with the PR5 bit, and the second, without.

Pseudo Code Using PR5 bit:

- **At FRM Send event:**

```
SR = Nrm/T;
ACR_ok = ((ACR ≤ SR) OR (TDF == 0.0));

IF (PR5 == FALSE)
        IF (ACR > SR + headroom)
        ACR = Max(SR + headroom, ACR*(1.0 - TDF));
        ENDIF
ELSE
        PR5 = FALSE;
```

- **At BRM Receive event:**

```
IF (NI = 0 AND ACR_ok)
```

```
        IF (ACR < ER) PR5 = TRUE ELSE PR5 = FALSE;
        ACR = Min(ACR + AIR * PCR, PCR);
ENDIF

ACR = Min(ACR, ER);
ACR = Max(ACR, MCR);
```

- **Initialization**

```
ACR_ok = True;
PR5 = False;
```

Pseudo Code Without Using the PR5 bit:

- **At FRM Send event:**

```
SR = Nrm/T;
ACR_ok = ((ACR ≤ SR) OR (TDF == 0.0));

IF (ACR > SR + headroom)
        ACR = Max(SR + headroom, ACR*(1.0 - TDF));
ENDIF
```

- **At BRM Receive event:**

```
IF (NI = 0 AND ACR_ok)
        ACR = Min(ACR + AIR * PCR, PCR);
ENDIF

ACR = Min(ACR, ER);
ACR = Max(ACR, MCR);
```

- **Initialization**

```
ACR_ok = True;
```

Note that the comparison (ACR $\leq$ SR) may always yield false due to the fact that cells may be scheduled only at certain fixed slots. There is typically a minimum granularity Δ which dictates the cell scheduler at the Source End System. To account for this scheduler, the comparison may be replaced by (ACR $\leq$ SR + Δ).

7.2.7 Parameter Selection

The count-based proposal has two parameters: "headroom" and "TDF". The headroom parameter is recommended to be different from the ICR parameter due to the following reasons:

ICR is used just once during connection setup. It is equivalent to an explicit rate feedback in the first RM cell, but is available right after the connection is set up, as if a BRM was tagged to the "Connect-Confirm" message of the VC. Switches can negotiate ICR using a short term decision algorithm, just as they set ER in RM cells. ICR can be high or low dependent upon the current congestion level. Notice that ICR being set to high values does not disable UILI (as in the earlier proposals) because UILI is controlled by the headroom parameter.

The choice of the headroom parameter, on the other hand, can be viewed as a long term decision. This is because the value of the headroom parameter applies throughout the life of the VC, and could be several years (for PVCs). It controls how much the sources can lie about their rates at any time. It also determines how many cells the switch receives at once. Using these considerations, we would like to set headroom as low as possible. However, as discussed in the simulation results of bursty sources (Section 8.2), very small headroom is not desirable. A value of 10Mbps is recommended. This allows LANE traffic to go at full Ethernet speed. Smaller values can be used for WANs.

The proposal hence suggested a separation of the roles of ICR and headroom.

The parameter TDF determines the speed of convergence to the desired UILI goals (region B in Figure 4). Hence, it determines the duration for which the network is susceptible to load due to sources suddenly using their ACRs. Larger values of TDF give faster convergence. However, a low value is preferred for bursty sources as discussed in Section 8.2. TDF set to zero disables UILI. A value of 1/8 or 1/16 is recommended.

7.3 Time-Based UILI Proposal

The time-based UILI proposal has a ACR reduction function which depends upon the time, T, since the last FRM was sent. While this aspect is similar to the August 1995 UILI proposal, the other changes suggested are:

1. The time-based proposal also independently observes the problem with using ICR as the floor of the reduction function (as discussed in Section 7.2.3. The proposal suggests two possible floor values:

$$ACR_{max} = Max(ICR, TOF * SR)$$

or

$$ACR_{max} = ICR + SR$$

2. IF (ACR > ACR_{max}) $\quad\quad$ ACR_{new} = Max(ACR*(1 - T/Tc), ACR_{max});

   The recommended value for Tc is Max(ADDF*FRTT, TBE/PCR), where ADDF has a default value of 2. FRTT is the Fixed Round Trip Time measured at connection setup.

3. Rule 5b remains unchanged.

The first two key changes in the time-based UILI proposal since the August 1995 proposal are the use of a new floor (ACR_{max}) for the ACR reduction function, and the additive headroom option. The third important change is that 1/Tc replaces the parameter TDF, which is TDFF/RDF times the smallest power of 2 greater or equal to PCR. The parameter Tc is now dependent upon the FRTT, the Fixed Round Trip Time measured at connection setup.

This formula decreases ACR depending upon how long the idle period is, relative to the round-trip time. A detailed comparison of the count-based and the time-based alternatives is done in Section 8.

7.4 Joint Source-Based UILI Proposal

The time-based and count-based camps agreed on a consensus, which we refer to as the "joint source-based proposal." The proposal uses the count-based reduction function and a constant value for TDF. It uses the new floor of the reduction function and the additive headroom. However, ICR is used in the UILI function instead of the proposed "headroom" parameter. The hysterisis region (region B in Figure 4) suggested by the count-based proposal is not used. Rule 5b remains the same as the August 1995 proposal, and PR5 is not used since TDF is set to a small value (1/16), the count-based reduction formula is used.

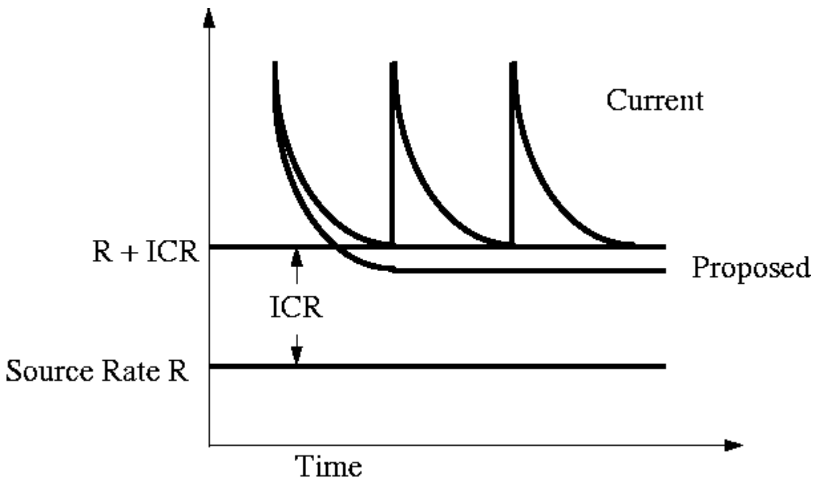


Figure 5: Joint Source-Based UILI Proposal vs Count-Based Proposal

The effect of removing the hysterisis region in the joint proposal is shown in Figure 5. In the joint proposal, the source will ignore one ER feedback after reducing the ACR to within the desired threshold. However, it may increase its rate based upon ER feedback henceforth. The source thus re-enters the danger zone of ACR retention. In the count-based proposal, a source which reaches the desired operating zone (ACR <= SR + ICR), it remains in this region until the source actually uses its ACR allocation.

The pseudo code for the joint proposal is as follows:

- **At FRM Send event:**

```
SR = Nrm/T;
IF (ACR > SR + ICR)
        ACR = Max(SR + ICR, ACR*15/16);
        ACR_ok = FALSE;
ELSE ACR_ok = TRUE;
```

- **At BRM Receive event:**

```
IF (NI = 0 AND (ACR_ok OR PNI))
        ACR = Min(ACR + AIR * PCR, PCR);
ENDIF

ACR = Min(ACR, ER);
ACR = Max(ACR, MCR);
```

7.5 Switch Based Proposal

Ramakrishnan et al [15] argued that the UILI function can be implemented in the switches on the following lines:

- Estimate rate of a connection and derive a smoothed average. This requires per-VC accounting at the switches.
- The switch maintains a local allocation for the VC based on, the max-min fair allocation and the rate the VC claims to go at, i.e., its CCR.
- Use an "aging" function at the switch which allocates a rate to the VC based on the the ratio of the CCR and the actual rate-estimate. Basically, this function widthdraws the allocations from ACR retaining sources.

  A suggested aging function was ($e^{\alpha u} - e^{\alpha\delta}$) where, u is the ratio of the expected rate and the actual rate, and, α and δ are parameters. The function has the property that the larger the difference between the CCR and the estimated actual rate, the greater the reduction factor.

Essentially, the switch allocates conservatively to sources which it knows are not using their allocations. The switch based policy however faces problems in handling sources which go idle. Idle sources do not send RM cells. The switch may take away the allocation of an idle source after a timeout, but there is no way to convey this information to the idle source, since there are no RM cells from the source.

Hence, the switch-based UILI proposal also suggests a simple timeout policy at the source which reduces the rate of the source to ICR after a timeout (parameter ATDF) of the order of 500ms. Idle sources which become active before the timeout expires may still overload the network. The UILI requirement is ignored for such sources.

7.6 ATM Forum decision

The ATM Forum debated considerably over the UILI issue in December 1995 before putting the issue to vote. The summary of the arguments were the following:

The use-it or lose-it can be implemented in switches or in NICs (sources) or both. The advantage of switch-only implementation is that NICs are simpler. The advantage of NIC implementation is that switches can be more aggressive in their bandwidth allocation without worrying about long-term implications of any one allocation. Without source based allocations, the switches have to have enough buffers to absorb the traffic that may result from overallocation of bandwidth. Also, switches control has delayed effect in the sense that if a switch finds its available bandwidth reduced (due to higher priority VBR traffic) and it sends a feedback (via returning RM cells or via backward explicit congestion notification (BECN)), there is a time delay before which a source comes to know that it should not use its previous allocation.

Finally, the ATM Forum decided to require only a switch-based approach. NICs can keep their allocation for a time period (parameter ATDF) of the order of 500 ms after which they are required to reduce their ACR to ICR. Those NIC vendors that feel 500 ms is too long can optionally implement additional measures that further reduce their ACR. The normative Annex F lists several such optional policies, which includes the joint-source based proposal.

8 Simulation Results

In this section, we study the tradeoffs in the UILI design through simulation results. Our first set of results in section 8.1 use a source-bottlenecked configuration. These results study the rate of convergence of the UILI approach to the desired objectives. The second set of results are for bursty sources whose active (bursty) periods are varied. We define a new closed loop bursty model and study the results based on the new metric, burst response time.

These results study the effect and tradeoffs of different UILI approaches on the performance of bursty sources.

8.1 Source Bottlenecked Configuration

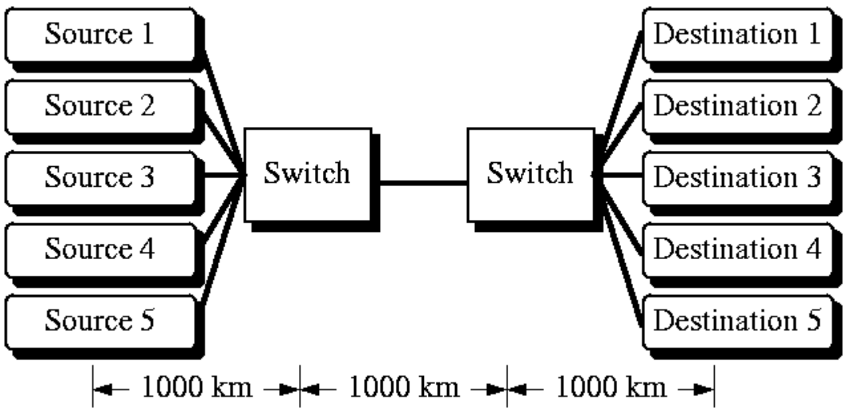


Figure 6: Five Sources Configuration

We present simulation results for the following five UILI alternatives:

1. No UILI
2. August 1995 UILI proposal
3. Baseline Rule 5, which is the August 1995 proposal, except that the time-based reduction formula is replaced by the count-based formula, an additive headroom is used in place of the multiplicative headroom, and a separate headroom parameter is used in place of ICR in the UILI reduction formula.
4. The count-based UILI proposal
5. The time-based UILI proposal

The configuration is a network consisting of five ABR sources (Figure 6) going through two switches to corresponding destinations. All simulation results use ERICA switch algorithm [22]. All links are 155 Mbps and 1000 km long. All VCs are bidirectional, that is, D1, D2, through D5 are also sending traffic to S1, S2 through S5.

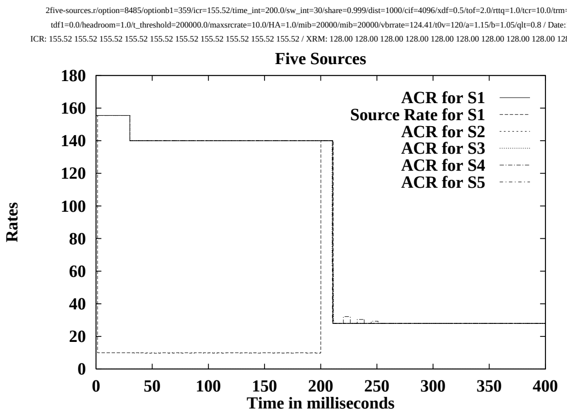
The following parameter values are used:

PCR = 155.52 Mbps
MCR = 0 Mpbs
ICR = 155.52 Mbps, 1 Mbps (Two values)
RIF (AIR) = 1
Nrm = 32
Mrm = 2
RDF = 1/512
Crm = MinTBE/Nrm, PCR*FRTT/Nrm
TOF = 2
Trm = 100 ms
FRTT = 30 ms
TBE = 4096 (Rule 6 is effectively disabled)
CDF (XDF) = 0.5
TDF = {0, 0.125}
{0 $\Rightarrow$ No rule 5, 0.125 for all versions of rule 5}

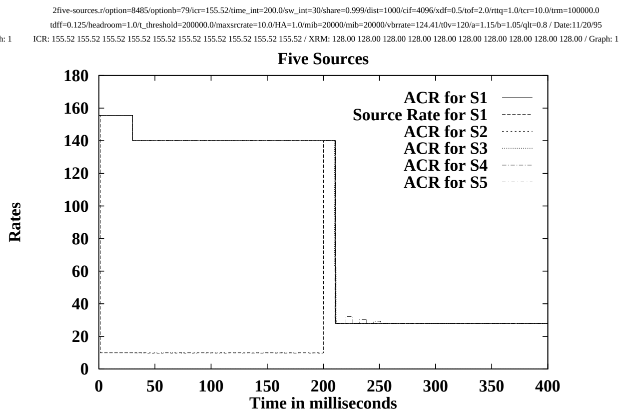
PNI = {0, 1}
{1 $\Rightarrow$ No rule 5b, 0 $\Rightarrow$ Rule 5b for August 1995 and Baseline UILI}

TCR = 10 c/s

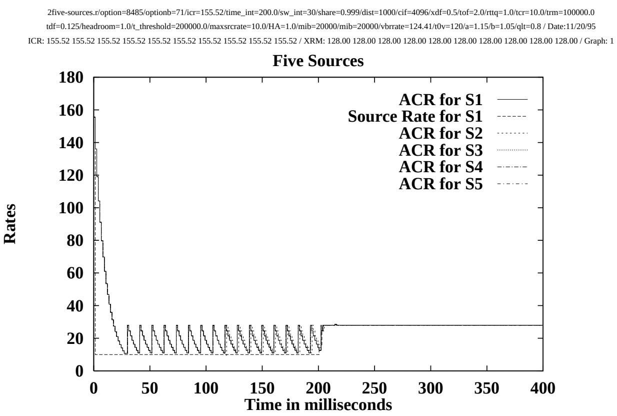
The simulation is run for 400 ms. For the first one-half (200 ms), the VCs are bottlenecked (probably because of other VCs from that source) at 10 Mbps. The system is underloaded and asks all sources to increase. The ER for all sources is high. At t=200 ms, the situation changes. All sources are able to use their allocated ER. In all these graphs we have set TBE to 4096 which ensures that SES Rule 6 is not triggered. This avoids the oscillations and reductions caused by rule 6 and allows us to study the effect of UILI separately.



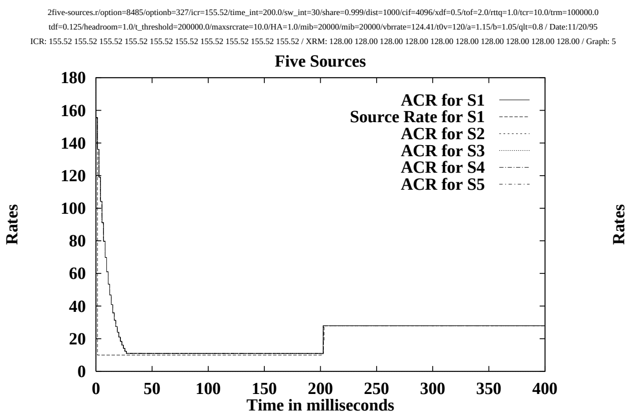
(a) No UILI



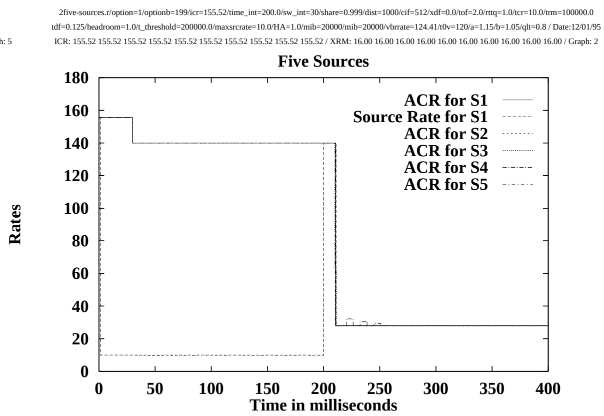
(b) Aug 1995 UILI



(c) Baseline UILI

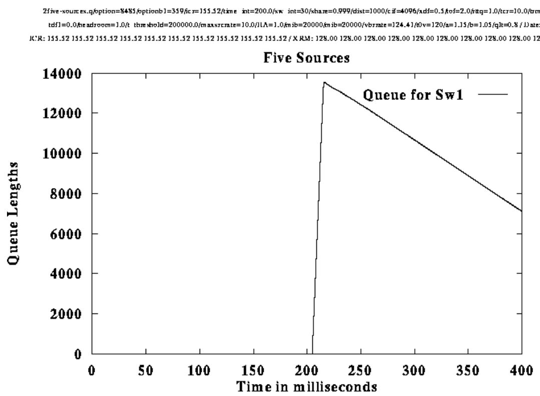


(d) Count-Based UILI

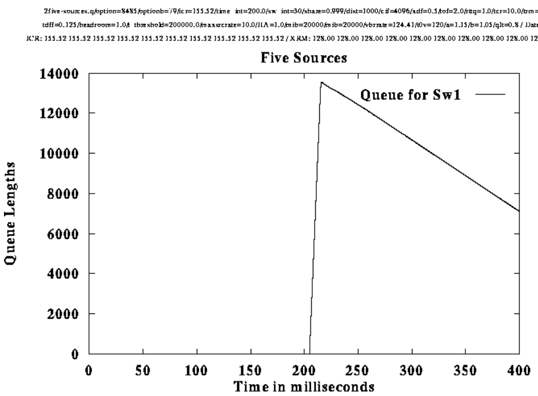


(e) Time-Based UILI

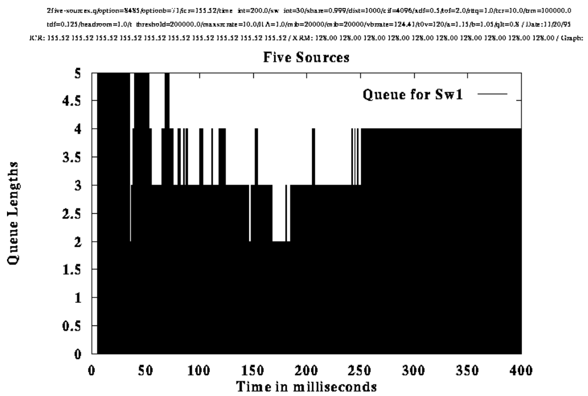
Figure 7: Five Source Configuration,Rates,High ICR=155.52Mbps,Headroom=1Mbps,MaxSrcRate=10Mb



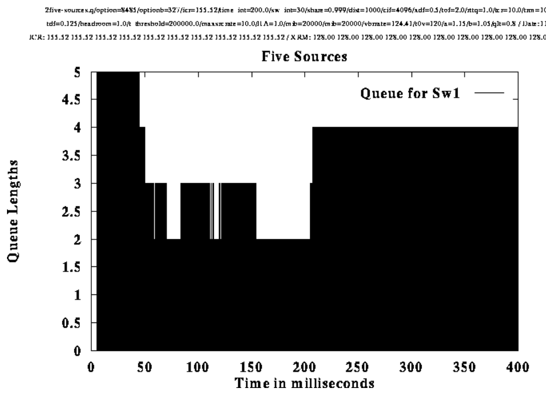
(a) No UILI



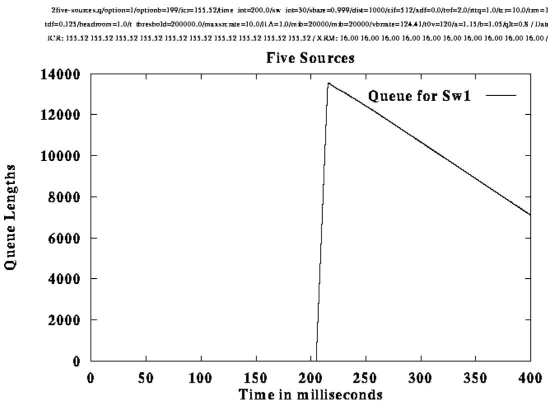
(b) Aug 1995 UILI



(c) Baseline UILI

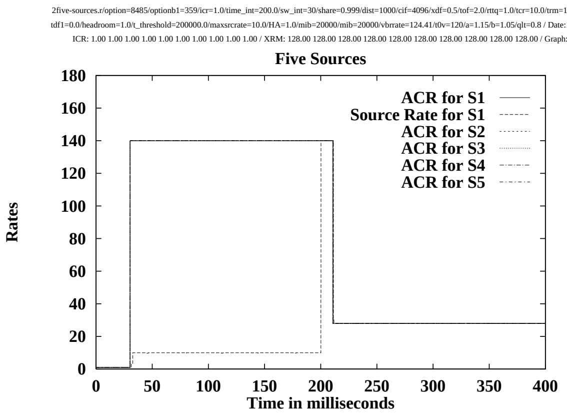


(d) Count-Based UILI

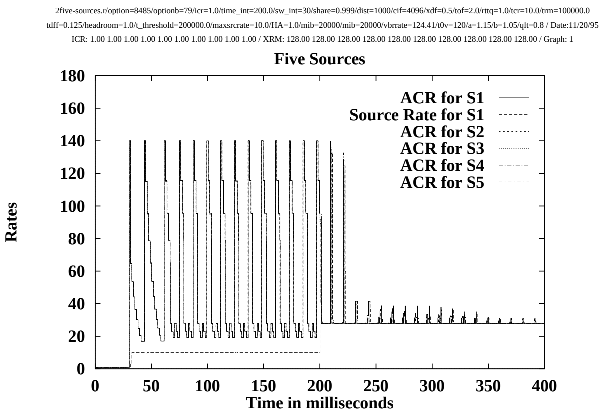


(e) Time-Based UILI

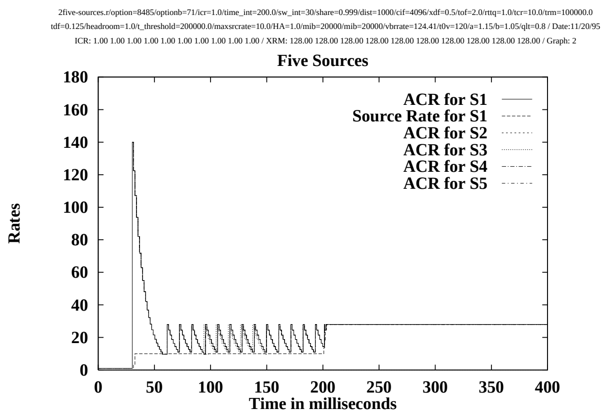
Figure 8: Five Source Configuration,Queues,High ICR=155.52Mbps,Headroom=1Mbps,MaxSrcRate=10M



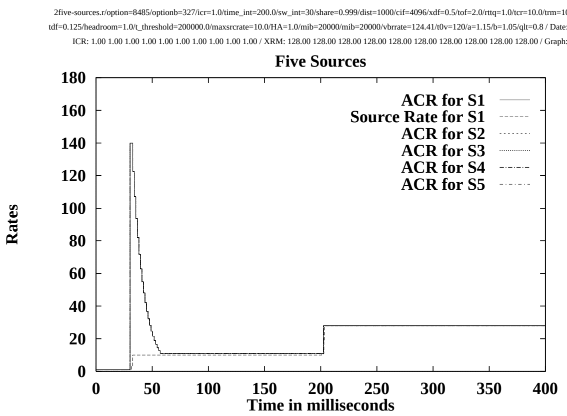
(a) No UILI



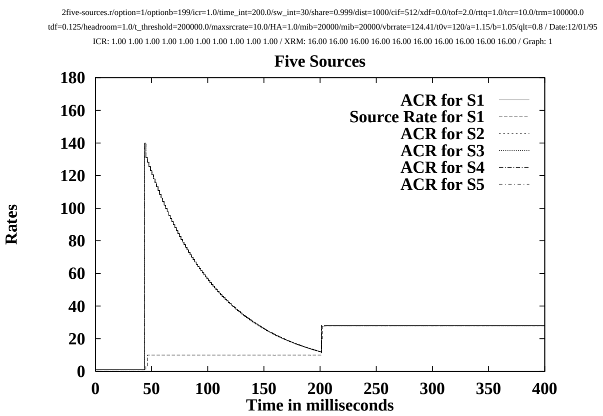
(b) Aug 1995 UILI



(c) Baseline UILI

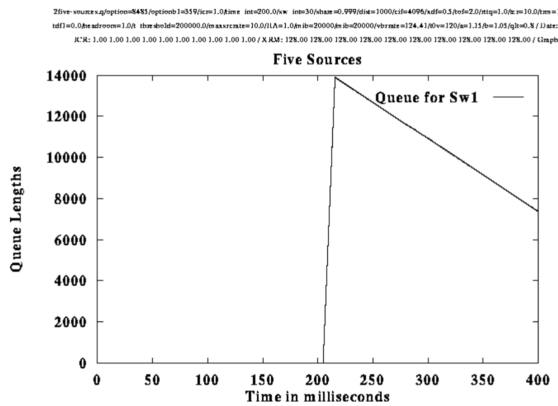


(d) Count-Based UILI

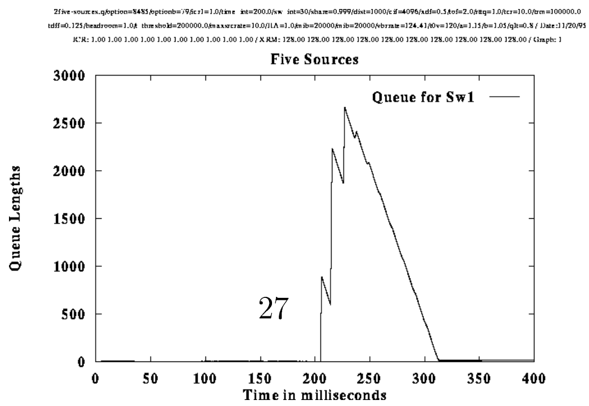


(e) Time-Based UILI

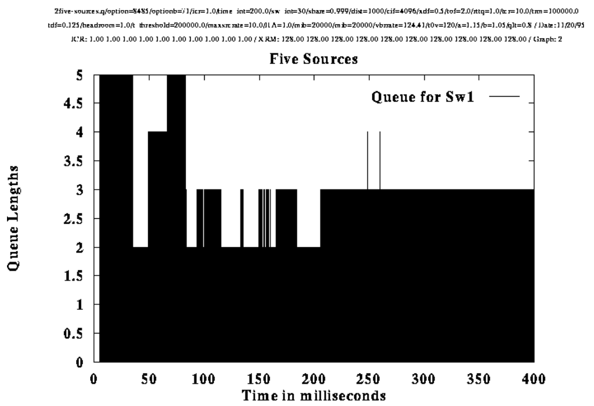
Figure 9: Five Source Configuration,Rates,Low ICR=1.0Mbps,Headroom=1Mbps,MaxSrcRate=10Mbps f



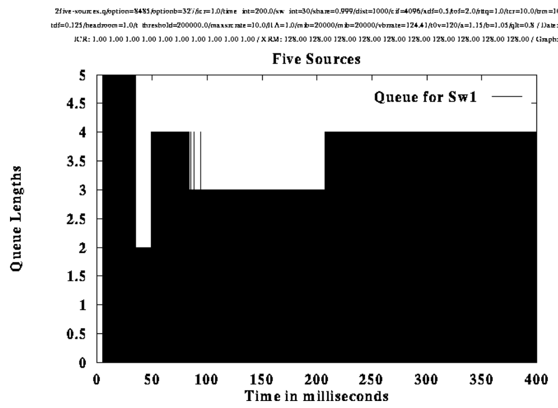
(a) No UILI



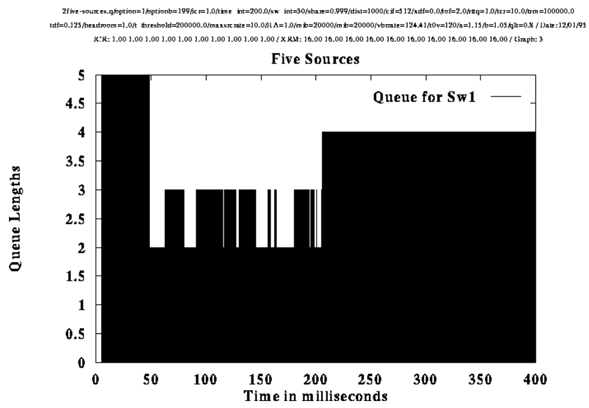
(b) Aug 1995 UILI



(c) Baseline UILI



(d) Count-Based UILI

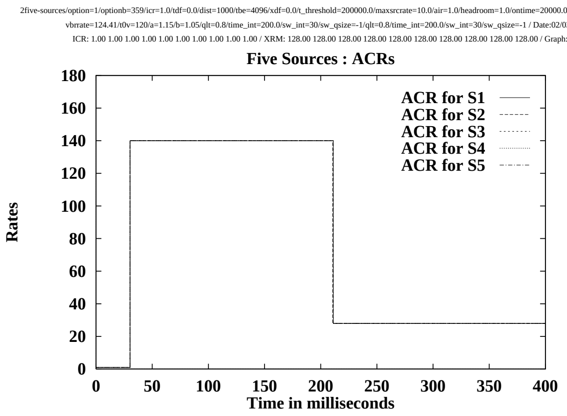


(e) Time-Based UILI

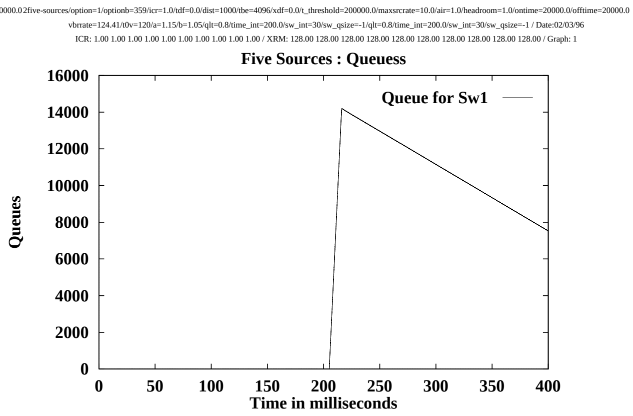
Figure 10: Five Source Configuration,Queues,Low ICR=1.0Mbps,Headroom=1Mbps,MaxSrcRate=10Mbp

Figures 7, 8, 9, 10 show ACR, actual source rates and queue lengths for the five UILI alternatives.

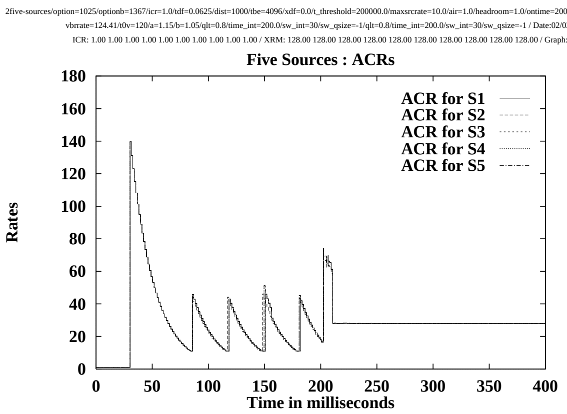
Figure 7 shows ACRs and Source rates for the five alternatives studied. There are six lines in each graph consisting of five ACR values and one source rates. Since all five sources are identical, the curves lie on the top of each other. Notice that with no UILI, the August 1995 UILI proposal, and the time-based proposal, the ACR can remain high (compared to source rate). UILI is not triggered because of the high ICR functioning as a high headroom. With the baseline UILI, which uses a separate headroom parameter, the ACR comes down to source rate plus headroom. Here headroom value of 1 Mbps was used. However, after ignoring one explicit feedback from the network (rule 5b), it jumps back to the rate sent by the network and thus oscillates between approximately 30 Mbps and 11 Mbps. The count-based UILI does not have these oscillations.



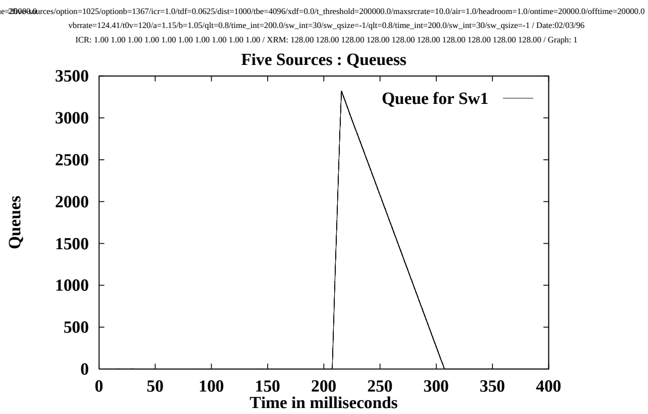
(a) August 1995 UILI: ACRs



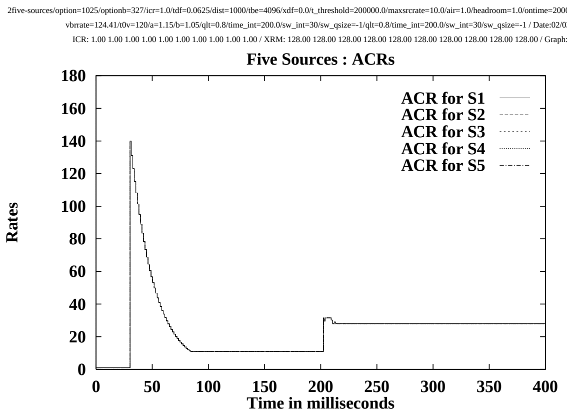
(b) August 1995 UILI: Queues



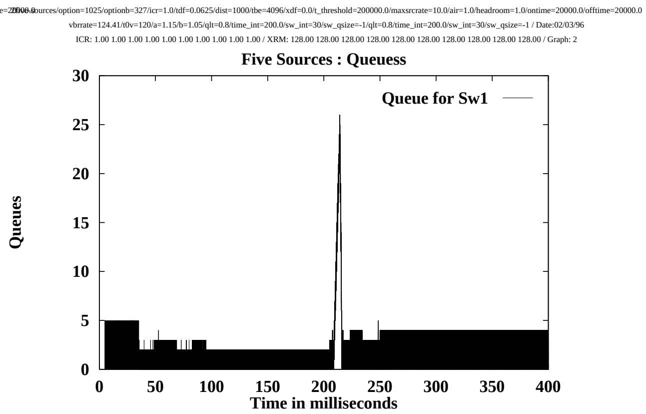
(c) Joint UILI : ACRs



(d) Joint UILI : Queues



(e) Count-Based UILI: ACRs



(f) Count-Based UILI: Queues

Figure 11: Five Source Configuration,Low ICR=1.0Mbps,MaxSrcRate=10Mbps for 200ms

These oscillations can cause queues if the switch algorihm overallocates and the overload the network. Figure 11 shows a comparison between the behavior of the August 1995 proposal, the joint proposal (count-based and time-based) and the count-based proposal using an old version of ERICA, which was not tuned to avoid oscillations. The simulation shows that, if the switch algorithm overallocates, and the UILI method oscillates out of the headroom zone, overload is possible.

Figure 8 shows queue lengths for the five alternatives depicted in Figure 7. Notice that time t=200 ms, the sources start using their high ACR and the queue length suddenly grows with no UILI, August 1995 UILI, and the time-based proposal. The baseline UILI and the count-based UILI do not have this problem. The ACR and source rates are close to each other and the queue never goes over 5 cells.

Figure 9 shows the same information (ACR and source rates) as Figure 7 for ICR of 1 Mbps. Notice that August 1995 UILI causes oscillations, due to the wrong floor of the ACR reduction function. Baseline UILI oscillates between the goal and the network feedback. The count-based UILI converges quickly to the goal and does not have oscillations after reaching the goal. The time-based UILI converges very slowly to the goal. Had the sources started using their ACR allocations earlier, it would have resulted in network queues.

Figure 10 shows the queue lengths for the low ICR cases. Notice that no UILI and August 1995 UILI again result in large queues (even though ICR is low). Baseline UILI, the count-based UILi and the time-based UILI limit the queue length to under 5 cells.

8.2 Bursty Sources

Bursty sources are characterized by having active periods where they send data and idle periods where they do not have data to send. The metric of importance for such sources is the *burst response time*, which is the time taken to transmit the burst. Another metric of interest is the *effective throughput* which measures the average transmission rate of the burst.

Figure 12 illustrates the use of the two metrics: burst response time and effective throughput. The figure has two parts, both of which show the arrival and departure of a burst at the end-system. The first part of the figure shows a burst takes a long time to be transmitted, i.e., it has a long burst response time and low effective throughput. The second part shows a burst which is transmitted quickly; the burst response time is short, and the effective throughput is higher. Note that the effective throughput is related to the size of the burst and the burst response time. In our simulations we simply count the number of equal size bursts transmitted per unit time and use this as the metric of performance.

The burst response time is inversely proportional to ACR allocated to the source. If the initial ACR is low and the network is not congested, the burst response time depends upon how fast a high rate can be allocated to the source, i.e., the transient response time of the switch scheme which is the time required by the scheme to bring the system into stable and efficient operation.

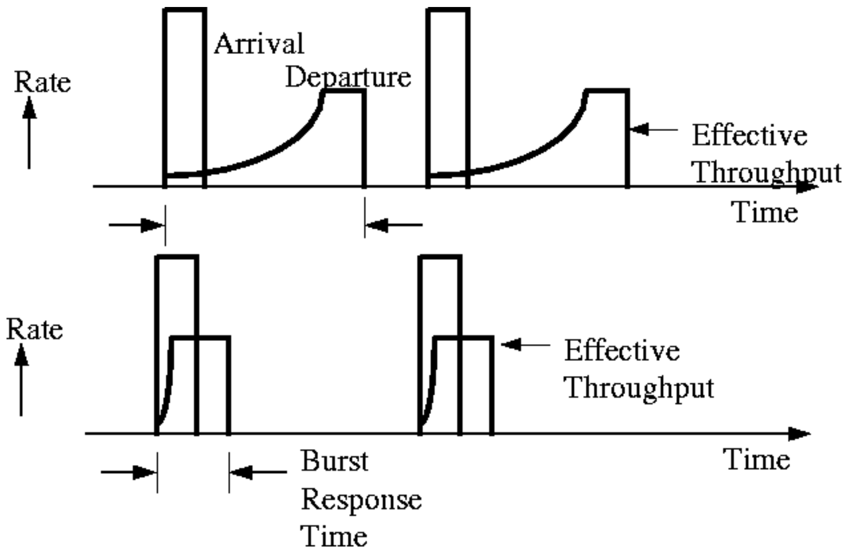


Figure 12: Burst Response Time vs Effective Throughput

If the initial ACR is high, sending a burst at a high rate might overload the network, and the network may ask the source to decrease its rate. In fact, the source may be considered as ACR retaining. The UILI mechanisms will be enabled and will reduce the source's ACR. This in turn increases the burst response time, leading to lower effective burst throughput. Hence, the goals of the UILI policies inherently conflict with the goals of bursty traffic performance. UILI policies which reduce ACR based upon the length of the idle time, without considering the length of the burst (which is not known a priori) can result in high burst response time and hence reduced performance.

We study the effect of the UILI policy when the active period is short (burst size is smaller than Nrm), medium (burst time smaller than round trip time (RTT), but burst size larger than Nrm) and large (burst time larger than RTT). The network queue is not a problem for short or medium bursts. It does become important when larger bursts active periods are used. The user-level metric is the burst response time.

We use the model described in the following section to generate short, medium and long bursts, and isolate the behavior of UILI for study.

8.2.1 Closed Loop Bursty Traffic Model

Jain et al [10] define a new "closed loop" bursty traffic model as shown in Figure 13. The model consists of cycles of request-response traffic. In each cycle a source application sends a set of requests and receives a set of responses from the destination application. The next cycle begins after all the responses of the previous cycle have been received and an inter cycle time has elapsed. There is a gap between successive requests. It is called inter-request time. The request contains a bunch of cells sent back to back by the application at PCR. The adapter controls the output rate to ACR, however.

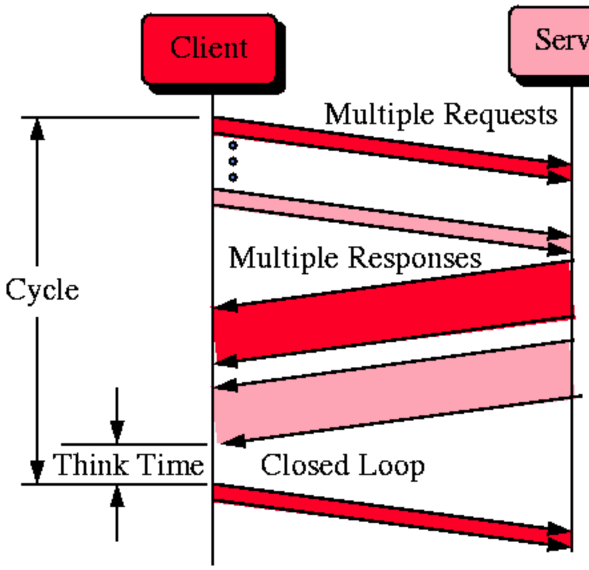


Figure 13: Closed Loop Bursty Traffic Model

The model as presented above may represent World Wide Web traffic, transaction oriented traffic, or request-response traffic. The model is "closed loop" in the sense that the rate at which cycles (and hence requests) are generated depends upon the responsivity of the network. If the network is congested, the response take longer time to come back and the sources do not generate new requests until the previous ones have been responded. This is more realistic than an "open loop" model where bursts are generated at a fixed rate regardless of the congestion in the network.

The packet train model [25] used previously to model bursty behaviour is an open loop model. This model has the source sending out cars (having back to back packets) with an inter-car time and trains having an inter-train time. But, the source does not wait for any response from the destination.

Note that the time between two sets of requests is a cycle time which is at least the sum of the round-trip time and the inter-cycle time. This is the idle time between two sets of requests. Hence the the idle time is greater than the round-trip time.

The idle time being greater than the round trip time implies that all the RM cells from the previous set of requests have returned to the source before the new set of requests are sent. Note that, if the sum of the response bursts are large (greater than Nrm*Nrm cells), more than Nrm RM cells of the response burst are turned around by the source (client). According to the SES rules, in such a case, an FRM is sent from the SES, which may return to the source with feedback. Since such RM cells are very infrequent, we can safely ignore them and assume that, when the new burst starts, there are no RM cells of the source in the network.

In our simulations, we use one request from the client and one response from the server. We use a small response burst size (16 cells), and vary the request burst size.

8.2.2 Single-Client Configuration and Parameter Values

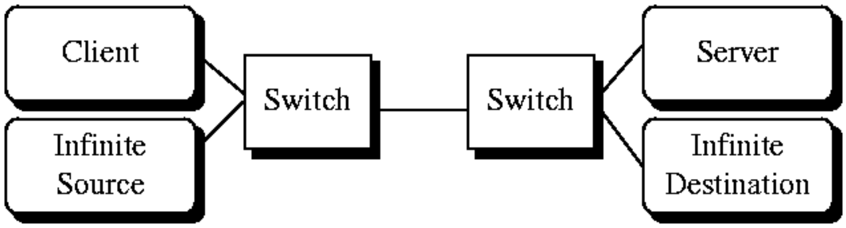


Figure 14: Client-Server Configuration With Infinite Source Background

The configuration we use is called the single-client configuration (Figure 14). It consists of a single client which communicates with the server, via a VC which traverses a bottleneck link. An infinite source is used in the background to ensured that the network is loaded, and any sudden bursts of traffic manifest as queues. All the links run at 155Mbps.

The response size is kept constant at 16 cells. The request size can be 16, 256 or 8192 for small, meduim or large bursts respectively. The inter-cycle time is chosen to be 1ms. All links are 500km long.

The other source parameters are chosen to maximize ACR and disable the effects of other source rules:
ICR = 10Mbps
TDF = 1/8
TCR = 10 cells/sec
TRM = 100ms
TBE = 512
CDF = 0 to disable SES Rule 6

For the time-based implementation, The switch algorithm used is ERICA [22]. It has two parameters, the target utilization which is set to 90%, and the averaging interval which is set to the minimum of a timeout value, 1ms and the time to receive 100 ABR cells.

8.2.3 Small Bursts

Small bursts are seen in LANE traffic. For example, the ethernet MTU, 1518 bytes is smaller than 32 (Nrm) cells. Since small bursts are smaller than Nrm cells, no RM cells are transmitted during some bursts. As a result, no source rules are triggered during these bursts. In other words, the entire burst is transmitted at one rate.

However, RM cells are transmitted in some bursts. UILI is triggered before the RM cell is transmitted and brings down the ACR to ICR. Source rate, SR, is nearly zero due to the short burst time and long idle time. Hence, ICR + SR is approximately equal to ICR.

Figure 15 shows the effect of UILI on the source rate of small bursts. The network feedback first arrives when the source is idle, asking it to increase its ACR. The source uses its ACR

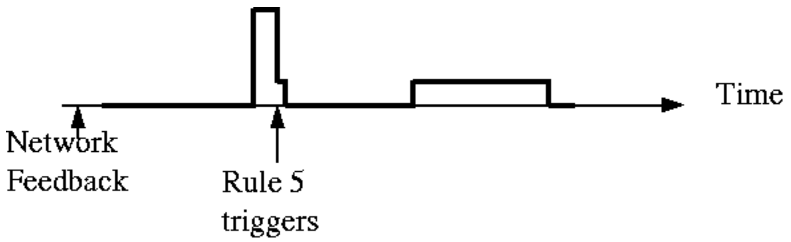
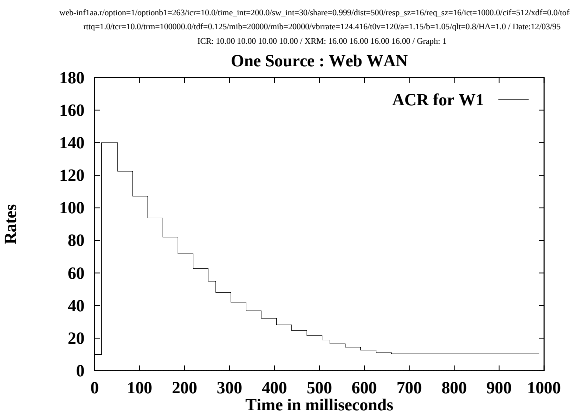


Figure 15: Effect of UILI on Small Bursts

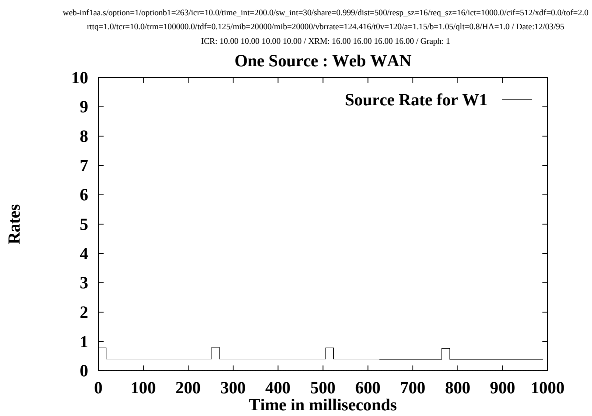
to almost send the full burst. The first RM cell sent reduces its source rate back to ICR. The source rate goes back to zero when the source is idle.

In the time-based proposal, the feedback brought by the RM cell is ignored because of rule 5b which ignores one feedback after the ACR decrease. Hence, the next two bursts are sent at ICR, before the next RM cell is sent which brings back feedback requesting an increase. Note that the sending of this second RM cell does not decrease the ACR since it is already at ICR. Hence, on the average, one out of every three bursts is sent at a higher rate.

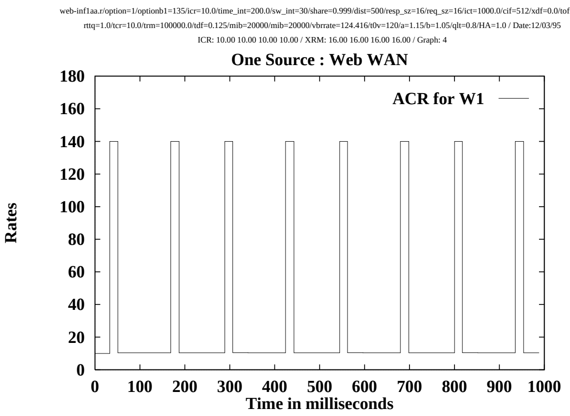
In the count-based proposal, the rate-increase feedbacks are always ignored. Hence, the ACR successively reduces to ICR and remains at ICR. Over the long term, all short bursts are sent out at ICR only.



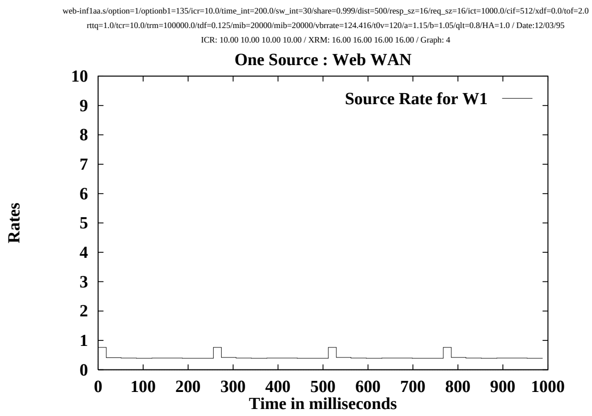
(a) Count-Based UILI : ACR



(b) Count-Based UILI : Source Rates



(c) Time-Based UILI : ACR



(d) Time-Based UILI : Source Rates

Figure 16: Small Bursts : One Bursty Source + Infinite Source Background

The simulation results showing the ACR and source rates for the small bursts for the time-based vs the count-based proposal are shown in Figure 16.

The count based performance can be improved by using a leaky bucket or GCRA [23] type of burst tolerance mechanism where small bursts can be sent at link rate irrespective of ACR or ICR. Other alternatives include choosing a small TDF or a larger ICR. An ICR of 10Mbps allows LANE traffic (the source of small bursts) to go through at full speed.

On the other hand, since the burst is very short, there is not a significant time difference in transmitting the burst at ACR and transmitting it at ICR (assuming ICR is not very small). The emphasis then shifts to supporting medium bursts and large bursts efficiently.

8.2.4 Medium Bursts

Medium bursts are expected in ATM backbone traffic or native mode ATM applications. Medium bursts contain more than Nrm cells, but the burst time is shorter than the round trip time. Though multiple RM cells are sent in a single burst, the network feedback for the burst arrives only after the burst has already been transmitted.

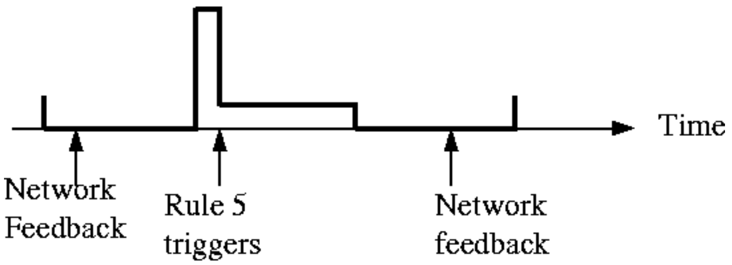
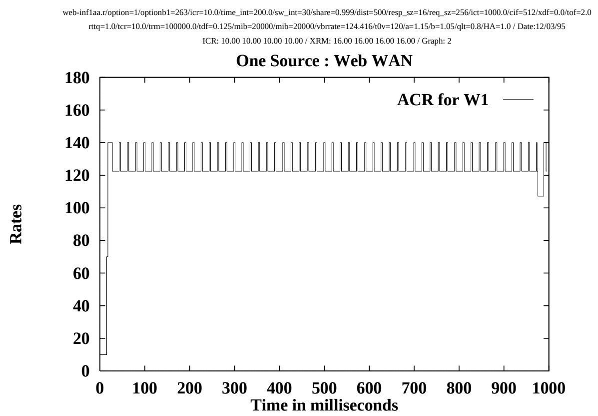
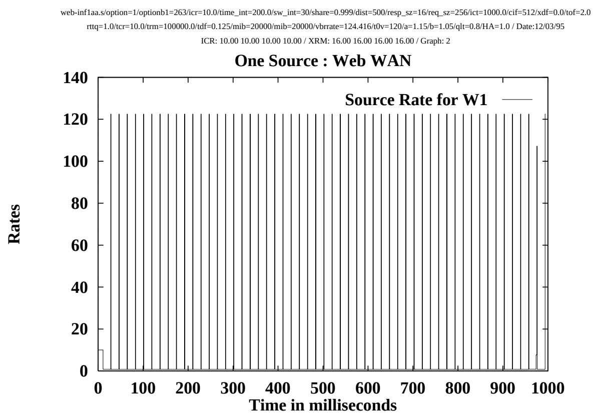


Figure 17: Effect of UILI on Medium Bursts

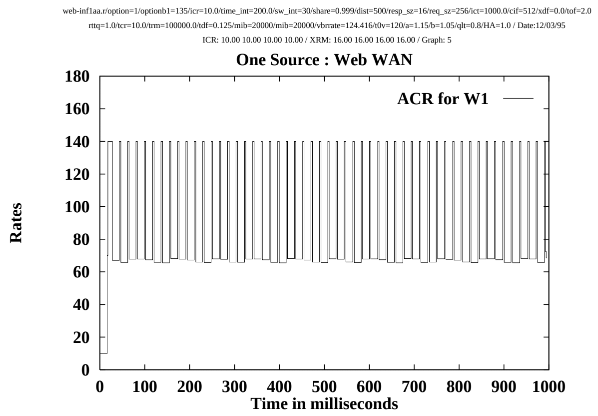
As shown in Figure 17, the UILI mechanism triggers once when the first RM cell is sent. In the time-based proposal, the amount of decrease is proportional to the idle time prior to the burst, while in the count-based UILI, the decrease is a constant amount. In the time-based proposal, if the idle time is large, almost the entire burst may be transmitted at ICR. Since, the count-based proposal sends the burst almost at ACR*(1-TDF), it provides better burst response.



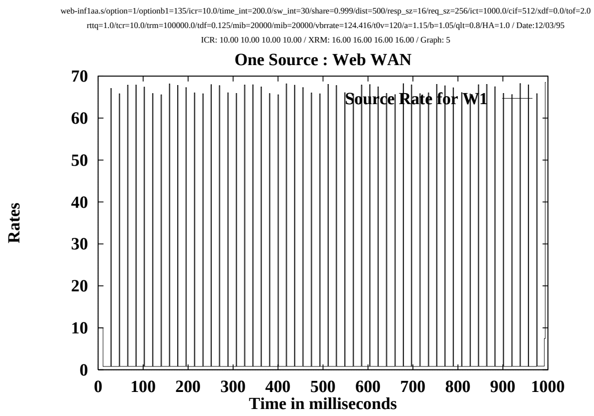
(a) Count-Based UILI : ACR



(b) Count-Based UILI : Source Rates



(c) Time-Based UILI : ACR



(d) Time-Based UILI : Source Rates

Figure 18: Medium Bursts : One Bursty Source + Infinite Source Background

Simulation results in Figure 18 show that the average source rate experienced by the bursts is higher for the count-based option (120 Mbps) compared to the time-based option (68 Mbps).

8.2.5 Large Bursts

Large bursts are expected to be seen in backbone ATM links. Large bursts have a burst time larger than the round trip time. The network feedback returns to the source before the burst completes transmission.

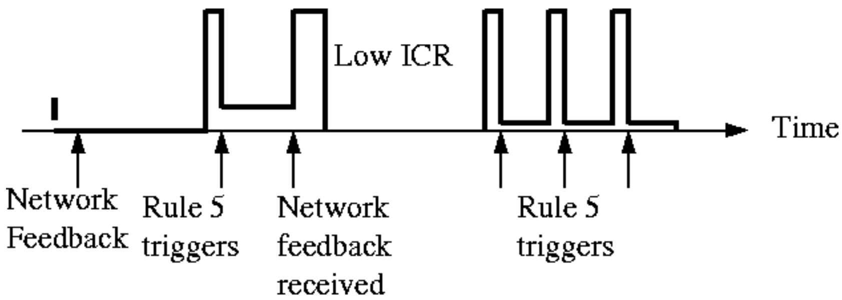
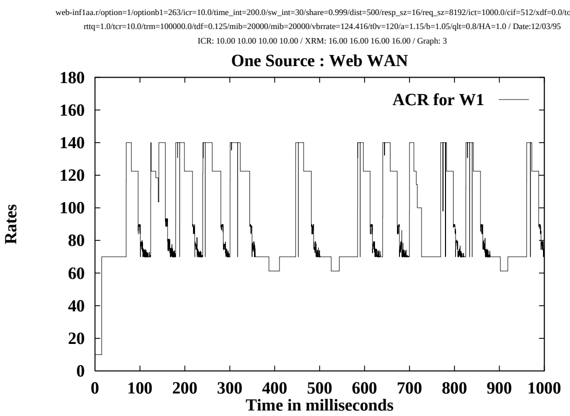
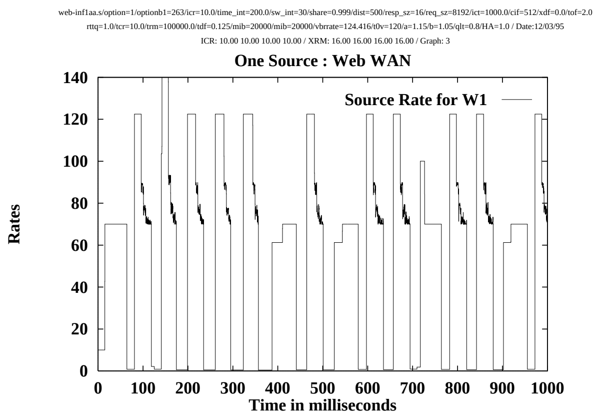


Figure 19: Effect of UILI on Large Bursts

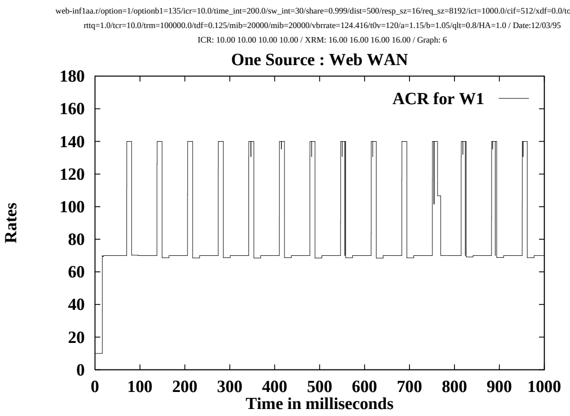
Figure 19 shows the behavior of large bursts with the August 1995 proposal. When the burst starts, UILI triggers when the first RM cell is sent, and brings the rate to ICR. Some part of the burst is transmitted at ICR. When network feedback is received, the ACR increases to the network directed value. Since ICR is not very low, there are no further oscillations, since normal increase is not hampered. However, if ICR is very low, UILI is triggered after the ACR increase, when the first RM cell is sent, bringing the rate down to ACR again. The cycle is repeated, as UILI triggers multiple times during the transmission of the burst. In the latter case, effective throughput is low, and burst response time is high.



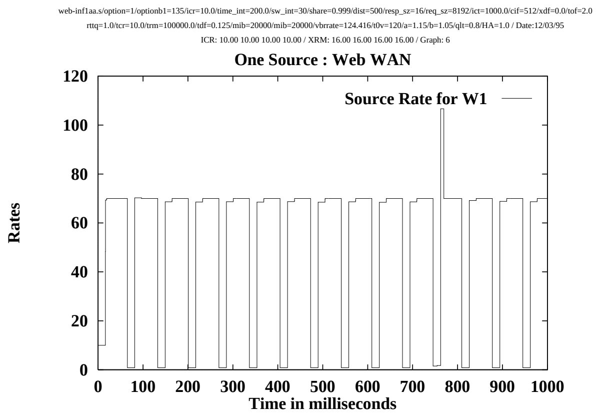
(a) Count-Based UILI : ACR



(b) Count-Based UILI : Source Rates



(c) Time-Based UILI : ACR



(d) Time-Based UILI : Source Rates

Figure 20: Large Bursts : One Bursty Source + Infinite Source Background

The time-based UILI avoids the multiple triggering of UILI. It triggers once when the burst starts, and reduces the ACR proportional to the idle time. The count-based UILI also triggers once, and reduces the ACR by a constant value. The comparative simulation results are shown in Figure 20.

Since the burst size is large, for large idle times (¿ RTT), the count-based technique may result in a larger queue in the network compared to the time-based technique.

However, for such large idle times, there are no RM cells of the source in the network. In such situations, SES Rule 6 provides the network protection against large loads. SES Rule 6 triggers when TBE cells are sent and no BRM cells are received. When SES Rule 6 triggers, ACR is reduced by a proportional factor, CDF for every extra FRM cell sent. The count-based proposal, in conjuction with SES Rule 6 protects the network under these circumstances.

9 Summary

The Use-It or Lose-It (UILI) issue was debated in the ATM Forum for over a year. The central problem was that of ACR Retention where sources may send at a rate smaller than their allocations for a period of time and then overload the network by using their entire ACR allocation suddenly. Source-bottlenecked and Bursty sources fall into such a category. The challenge was to develop a UILI scheme which gives good performance for both these types of sources, and protects the network against sudden overloads. It was felt that such a solution held the key to the success of the ABR service.

We have surveyed a number of UILI alternatives proposed over the year and why ideas in each of them was accepted or rejected. The final proposals could be classified as a source-based joint proposal and a switch-based proposal. The ATM Forum adopted the switch based proposal, with a simple source timeout. NIC manufacturers can optionally implement the joint source-based proposal which is briefly described in Informative appendix I.8 of the ATM Traffic Management 4.0 specifications. This paper traces the history and comprehensively evaluates the various options and their tradeoffs.

References

[1] R. Jain, "ABR Service on ATM Networks: What is it?," *Network World*, June 24th, 1995.[1]

[2] A.W.Barnhart, "Changes Required to the Specification of Source Behavior," *AF-TM 95-0193*[2], February 1995.

[1] All our papers and ATM Forum contributions are available through http://www.cis.ohio-state.edu/~jain

[2] Throughout this section, AF-TM refers to ATM Forum Traffic Management sub-working group contributions.

[3] A.W.Barnhart, "Evaluation of Source Behaviour Specification," *AF-TM 95-0267R1*, April 1995.

[4] Lawrence G. Roberts, "RM Return Failure Condition," *AF-TM 95-0340*, April 1995.

[5] John B. Kenney, "Problems and Suggested Solutions in Core Behavior," *AF-TM 95-0564R1*, May 1995.

[6] John B. Kenney, Srini W. Seetharam, "Minor Change in Reduction Formula Fixes Flaw," *AF-TM 95-0745R1*, June 1995.

[7] Anna Charny, Gunter Leeb, Michael Clarke, "Some Observations on Source Behavior 5 of the Traffic Management Specification," *AF-TM 95-0976R1*, August 1995.

[8] Bennett J., Fendick K., Ramakrishnan K.K., Bonomi F., "RPC Behaviour as it Relates to Source Behavior 5," *AF-TM 95-0568R1*, May 1995.

[9] R. Jain, S. Kalyanaraman, S. Fahmy, F. Lu, "Out-of-Rate RM Cell Issues and Effect of Trm, TOF, and TCR," *AF-TM 95-973R1*, August 1995.

[10] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, F. Lu, "Bursty ABR Sources," *AF-TM 95-1345*, October 1995.

[11] A.W.Barnhart, "Evaluation and Proposed Solutions for Source Behavior # 5," *AF-TM 95-1614*, December 1995.

[12] L.G.Roberts, "Operation of Source Behavior # 5," *AF-TM 95-1641*, December 1995.

[13] S. Liu, M. Procanik, T. Chen, V.K. Samalam, J. Ormond, "An analysis of source rule # 5," *AF-TM 95-1545*, December 1995.

[14] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, F. Lu, "A Fix for Source End System Rule 5," *AF-TM 95-1660*, December 1995.

[15] K.K. Ramakrishnan, P. P. Mishra and K. W. Fendick, "Examination of Alternative Mechanisms for Use-it-or-Lose-it," *AF-TM 95-1599*, December 1995.

[16] R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, F. Lu, S. Srinidhi, "Comments on "Use-it or Lose-it"," *AF-TM 96-0178*, February 1996.

[17] H. T. Kung, et al, "Adaptive Credit Allocation for Flow-Controlled VCs," AF-TM 94-0282, March 1994

[18] H. T. Kung, et al, "Flow Controlled Virtual Connections Proposal for ATM Traffic Management," AF-TM 94-0632R2, September 1994.

[19] J. Scott, et al, "Link by Link, Per VC Credit Based Flow Control," AF-TM 94-0168, March 1994.

[20] M. Hluchyj, et al., "Closed-Loop Rate-Based Traffic Management," AF-TM 94-0211R3, April 1994.

[21] R. Jain, S. Kalyanaraman and R. Viswanathan, "The OSU Scheme for Congestion Avoidance using Explicit Rate Indication," Tech Report OSU-CISRC-1/96-TR02, Dept. of CIS, Ohio State University, 1996.

[22] Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, and Ram Viswanathan, "ERICA Switch Algorithm: A Complete Description," *AF-TM 96-1172*, August 1996.

[23] ATM Forum, "ATM Traffic Management Specification Version 4.0," April 1996, available as ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.000.ps

[24] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent advances and a survey," invited submission to *Computer Networks and ISDN Systems*, October 1996, also AF-TM 95-0177, February 1995.

[25] R. Jain and S. Routhier, "Packet Trains - Measurement and a new model for computer network trafic," *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.

[26] Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, "Source Behavior for ATM ABR Traffic Management: An Explanation," *To appear in IEEE Communications Magazine*, November 1996.