Resource Requirements for ABR Explicit Rate Flow Control: Deterministic and Probabilistic Analyses

Ivy Hsu, [1] Stephan Robert[2], and Jean Walrand [3]

Keywords: ATM, Available Bit Rate Service, Explicit Rate control

Abstract

ATM Available Bit Rate (ABR) service is intended to offer low cell loss for non-real-time data sources that can respond to closed-loop flow control. ATM Forum Traffic Management Specification Version 4.0 defines the various parameters used in the ABR flow control, as well as the source, destination, and switch behaviors. However, the switch designers and service providers are free to choose the method of congestion control to implement and the ABR Quality of Service (QoS) objective to offer. This paper addresses the interaction among the flow control algorithm, the switch resource requirements, and the resulting QoS characteristics.

In this paper we propose and evaluate an Explicit Rate (ER) algorithm. The objective of this algorithm is to maintain the total buffer occupancy of all ABR connections to be close to a given threshold. By maintaining a non-zero queue, the ABR service can achieve a high utilization. The switch periodically determines its desirable ER value, based on the available capacity, the ABR buffer occupancy, and the number of active ABR sources.

We develop analyses that relate ABR resources to QoS objectives for this algorithm. The first approach is a deterministic, conservative analysis. It provides formulas for determining the ABR buffer and capacity requirements that can achieve zero buffer overflow.

The second analysis determines an upper bound on the buffer overflow probability when the above requirements are not met. The result is most effective when the number of active sources is a small fraction of the total ABR connections. Numerical examples show that by slightly relaxing the loss constraint, resource requirements can be significantly reduced.

1 Introduction

The Available Bit Rate (ABR) Service Class defined by ATM Forum is intended for data applications that can respond to feedback control information generated by networks. In the events of network congestion, the congested network elements can use such feedback information to signal the ABR sources for rate reduction. It is envisioned that both ABR

[1]Bay Architecture Lab, Nortel Networks. 4401 Great America Pkwy. SC1-06, Santa Clara, CA 95052. email: ivyhsu@nortelnetworks.com

[2]Swisscom AG., Switzerland. email: stephan.robert@swisscom.com

[3]Department of EECS, University of California at Berkeley, Berkeley, CA 94720. email: wlr@eecs.berkeley.edu

and Unspecified Bit Rate (UBR) Services can be used to serve delay-insensitive applications by effectively utilizing the network resources unused by the Constant Bit Rate (CBR) and Variable Bit Rate (VBR) Service Classes.

The feedback information is primarily carried in ABR Resource Management (RM) cells that are generated by ABR source end systems and turned around and returned to the sources by ABR destination end systems. RM cells that flow from source to destination are referred to as Forward RM (FRM) cells and RM cells that flow from destination to source are referred to as Backward RM (BRM) cells.

The closed-loop control can be performed in at least one of four ways:

(1) In the simplest case, the congested switch can set the Explicit Forward Congestion Indication (EFCI) bit in the forward data cells as they pass by. This signals the destination to set the Congestion Indication (CI) bit in the next BRM cell and thus propogates the congestion information back to the source.

(2) In the approach known as Relative Rate Marking, the congested switch may set CI bit or the No Increase (NI) bit in FRM and/or BRM cells. This allows the switch to selectively signal rate decrease for certain VCs.

(3) In the approach known as Explicit Rate (ER) Marking, the congested switch calculates the desirable rate from the sources and indicates it in the ER field of the RM (typically BRM) cells if the desirable rate is less than the current ER.

(4) The switch may also act as a Virtual Source (VS) and Virtual Destination (VD) pair and thus segment the ABR control loop.

ATM Forum Traffic Management Specification Version 4.0 [1] defines the various parameters used in the ABR flow control. It also specifies the source, destination and switch behaviors. For example, it defines in detail the transmission priority of data and RM cells and how a source should behave in response to BRM information. However, many crucial elements governing the characteristics of the feedback control are left as implementatation and/or network operation specific, including

- The definition of "congestion" at each switch;
- The algorithm for calculating ER for switches that implement Explicit Rate Marking;
- The coupling between adjacent control segments associated with VS/VD; and
- The choice of ABR control parameters.

Consequently, how the ABR feedback control scheme ensures its QoS objective is also implementation specific, since QoS is tightly coupled with the flow control mechanism as well as the buffering and service scheduling mechanisms of the switch. Unlike UBR, which has been considered a best-effort service with no numerical QoS commitments, ABR service is intended to offer low cell loss ratios for conforming traffic. Therefore, how to relate the trio of flow control algorithm, switch resource requirements, and the resulting QoS remains an active research area of great interest.

In this paper we propose and evaluate an ER flow control algorithm. The objective of this algorithm is to maintain the total buffer occupancy of all ABR connections to be close to a given threshold. The net drift of the buffer occupancy therefore should be negative when it is above the threshold, and positive if it is below. By maintaining a non-zero queue, the ABR service can achieve a high bandwidth utilization.

The switch periodically determines its desirable ER value based on the total buffer occupancy, the current bandwidth available to serve ABR traffic, and the number of active ABR sources. This ER is used to update the ER field in the BRM cells of active sources. Sources that change from inactive to active state are bounded by a fixed rate for at most two update periods. This helps prevent sudden surges in the arrival rate and consequently the buffer occupancy.

Based on this algorithm, we develop analyses that relate ABR resource requirements to QoS objectives. The QoS objectives addressed in this paper are cell losses: either zero cell loss or a controlled probability of overflow. The first approach that we present is a deterministic, conservative analysis for achieving a zero cell loss objective. It provides formulas for determining the required buffer and capacity for ABR service. If the ABR buffer size and the bandwidth available to serve the ABR buffer are kept above these levels, zero buffer overflow can be guaranteed.

The second analysis determines an upper bound on the buffer overflow probability when the above resource requirements are not met. It takes into account the fact that a significant part of the resource requirements in the deterministic analysis are used to accommodate sources that just become active but are not yet governed by the ER algorithm. If at any one time only a small fraction of inactive sources become active, the deterministic approach may be too conservative. Based on this assumption, we show analytically how a switch can significantly reduce the resource requirements, while only slightly relaxing the loss constraint.

This paper builds upon the groundwork layed out in [6] and [9]. It removes several simplifying assumptions and thus better reflects the key characteristics of ABR mechanism: (1) One complexity in ABR analysis is the fact that the closed-loop control information (i.e., the RM cells) is generated asynchronously among the sources, and the frequency of the RM cells varies with the rate of the source. This is accounted for in this paper. (2) In practical systems, switches cannot continuously recalculate the desired ER values. In this paper we assume that the ER value is only updated periodically. (3) Another extension is a much more detailed probabilistic analysis, which offers a bound on the buffer overflow probability directly as a function of the system parameters. The result also leads to a more flexible trade-off between the buffer and bandwidth resource requirements.

This paper is organized in the following manner: In Section 2 we propose a simple switch algorithm for periodically calculating the desirable ER value. This algorithm forms the basis for the subsequent analyses. Section 3 presents a deterministic analysis that identifies a set of resource requirements (minimum buffer and bandwidth available to the ABR queue) that can guarantee zero loss for conforming sources. In Section 4 we relax this zero loss constraint and show that the resource requirements can be much reduced when we accept a small amount of loss. An upper bound on the buffer overflow probability is presented. Section 5 offers some numerical examples to illustrate the applicability of the formulas. We remark on areas of extensions for this work in Section 6.

2 Definition of ER Algorithm

2.1 System Model and Assumptions

A typical ATM switch may contain multiple queuing stages with per-VC and/or per-QoS buffering and bandwidth scheduling. In this paper we focus only on the buffer and bandwidth available to the ABR class, which we model as a single buffered link. Let b be the total buffer space available to the ABR connections. The buffer is drained at a time-varying rate $c(t)$.

Define n_{max} as the total number of ABR connections that share this buffered link. Define τ_j^f as the delay for a cell to propogate from the j^{th} source to the buffered link. Similarly, define τ_j^b as the delay in the return path from the buffered link to the j^{th}. Let the round-trip time between the j^{th} source and the buffered link be $\tau_j = \tau_j^f + \tau_j^b$. Define $\tau = max_{j=1}^{n_{max}}\{\tau_j\}$ as the longest round trip delay among all sources.

In this work, we do not require specific traffic models for the interfering CBR and VBR sources, which affect the behavior of the available ABR capacity $c(t)$. Instead, we take the approach of assuming a single general constraint on $c(t)$'s rate of decrease. We assume that the system satisfies the following constraint: $dc(t)/dt \geq -\rho$, for a positive constant ρ. Since a decrease in the amount of available bandwidth translates to a decrease in ER, which can only be conveyed after a finite delay, a decrease in $c(t)$ can lead to a surge in buffer occupancy. As we will see in Section 3, the sharper we allow this rate decrease, the larger is the buffer needed to absorbed such surge.

The intuition on the $dc(t)/dt$ constraint is as follows: In many practical switch implementations, the QoS buffers are served in a manner that approximates Generalized Processor Sharing (GPS) [7]. (e.g., Weighted Fair Queuing [10] and Weighted Round Robin [5]). In GPS service scheduling schemes, each queue is guaranteed its allocated bandwidth, while any bandwidth unused by one queue can be shared by others. For a CBR queue, the total Peak Cell Rate (PCR) of all CBR connections determines its allocated bandwidth. For a VBR queue, often the total Effective Bandwidth (EB) plays this role [3]. Changes in the number of connections due to call setup or release affect the bandwidth allocation in the service scheduler. A constraint ρ on $-dc(t)/dt$ reflects the rate of processing new calls and changing the service scheduler to accommodate new CBR and VBR connections.

2.2 Source Model

According to ATM Forum TM 4.0 [1], the source must generate a FRM cell if one of the following two conditions holds: (1) at least two in-rate cells have been sent and at least T_{rm} time has elapsed since the tranmission of the last FRM cell, (2) N_{rm} - 1 in-rate cells have been sent since the last FRM cell. In-rate cells refer to cells with Cell Loss Priority (CLP) set to 0. They can be either data cells or Forward or Backward RM cells.

These criteria indicate that the rate at which feedback control opportunities (RM cells) are generated is in general proportional to the data rate (condition 2 above). The only exception is when the data rate is very low (when condition 1 applies). The parameter T_{rm} is used to ensure that the interval between successive RM cells does not become excessively large in that case, so long as the source still has data to send (the two in-rate cells requirement corresponds to one BRM cell and one data cell between any two FRM cells from the source). The default value for T_{rm} is 100 msec.

In Section 3 we assume that the interval between successive RM cells, either forward or backward, is upper bounded by a constant σ for all sources. This assumption is met and $\sigma = T_{rm}$ if the requirement for two in-rate cells is always met, and if the interval between successive RM cells is not altered along the control loop. In Section 4 we draw a finer distinction between busy and idle sources, and only require that the bound σ be observed for busy sources.

According to [1], upon receiving a BRM cell, a source may need to adjust its Allowed Cell Rate (ACR) depending on the contents of three fields in the BRM cell: Congestion Indication (CI), No Increase (NI), and Explicit Rate (ER). The CI and NI bits are used for switches with EFCI or relative rate marking. The new ACR should be calculated as follows:

- if CI = 1: ACR = max{MCR, min{ACR - ACR * RDF, ER}};
- if CI = 0 and NI = 0: ACR = max{MCR, min{PCR, ACR + PCR*RIF, ER}};
- if CI = 0 and NI = 1: ACR = max{MCR, min{ACR, ER}},

where MCR is the Minimum Cell Rate of the connection, RIF is the Rate Increase Factor, and RDF is the Rate Decrease Factor. RIF and RDF control the source rate for the EFCI or relative rate rarking switches.

In this work we focus on networks in which all switches are capable of ER marking and that ACR is only dictated by the ER field in the BRM cells. ER offers more effective control and avoids the phenomenon of rate oscillation often associated with binary feedback control.

The rate of a new ABR source or a source that has been idle for some time must start with a pre-negotiated Initial Cell Rate (ICR). For our system, we require that the ICR of any ABR source must satisfy the inequality $r^* \leq ICR < r^* + R$, where the two parameters r^* and R are used in the ER algorithm below. r^* is the threshold that qualifies an ABR source as active, and R provides a range of variation for ICR. They are further described in Section 2.3. Furthermore, the limit on the idle time permitted before the Allowed Cell Rate (ACR) must start with ICR again is known as ACR Decrease Time Factor (ADTF). For this system let $ADTF = P > \sigma + \tau$, where P is the update interval of the ER algorithm given below.

Finally, denote $r_j(t)$ as the instantaneous rate, i.e., the Current Cell Rate (CCR), of the j^{th} source at time t. Then $r_j(t-\tau_j^f)$ is the instantaneous rate of the j^{th} source observed by the buffered link at time t.

2.3 Switch ER Algorithm

In this section we present a simple ER algorithm to be implemented at the buffered link.

One important concept of feedback control that has been given little attention in many existing ABR ER schemes is that the frequency of adaptation should not exceed the frequency of the feedback opportunities [1, 4]. Note that the buffered link may wait as long as σ time unit before it receives a BRM cell for a specific virtual circuit, and it takes up to τ time unit before the effect of the new ER is observed at the buffered link.

Thus in this algorithm, we require that the buffered link recalculates its desired explicit rate value periodically with period $P \geq \sigma + \tau$. Let $\{P_i,\ i = 1, 2, \ldots\}$ denote the set of successive recalculation times, where $P_{i+1} = P_i + P$. By setting $P \geq \sigma + \tau$, we ensure that a new ER calculated at P_i would have been communicated back to all sources and be reflected in the instantaneous rates of all sources observed at the buffered link by time P_{i+1}.

At time P_i, the desired ER value that the buffered link calculates will depend on the following factors: the number of "active" sources $N(P_i)$, the capacity that is available to serve this ABR queue $c(P_i)$, and the queue occupancy $q(P_i)$.

At the time of explicit rate update, a source is considered active if it generates traffic at a rate higher than r^* for some time in the previous interval. In other words, r^* is the threshold rate above which a source is considered active. Define the set of active sources $A(P_i)$ as

$$A(P_i) = \{j : r_j(s - \tau_j^f) > r^* \text{ for some } s \in [P_{i-1}, P_i)\},$$

and the number of active sources $N(P_i)$ as

$$N(P_i) = \sum_{j=1}^{n_{max}} 1\{r_j(s - \tau_j^f) > r^* \text{ for some } s \in [P_{i-1}, P_i)\}.$$

The objective of the algorithm is to maintain a target buffer occupancy q^*. Define

$$f(q(P_i)) = -k \cdot sgn(q(P_i) - q^*).$$

The purpose of $f(q(P_i))$ is to add a positive drift when the queue occupancy falls below the target occupancy q^*, and a negative drift when it is above q^*. In the analysis below, we will see the role played by the rate factor k. The desired ER value over the period $[P_i, P_{i+1})$ is then set to be

$$e(P_i) = \frac{a(P_i)}{N(P_i) \vee 1},$$

where $a(P_i) = f(q(P_i)) + c(P_i)$. In other words, the available capacity is increased or decreased by a constant k, depending on the buffer occupancy (note that k has the same unit as $c(P_i)$). The operation $N(P_i) \vee 1 = max\{N(P_i), 1\}$ is to ensure that $e(P_i)$ has a meaningful value when the number of active sources is zero.

However, instead of directly using $e(P_i)$ in all BRM cells that it receives over $[P_i, P_{i+1})$, the buffered link determines the explicit rate $e_j(t)$ for source j in the following manner: For any BRM cell received for source j within the next period $t \in [P_i, P_{i+1})$,

$$e_j(t) = \begin{cases} e(P_i) & \text{if source } j \in A(P_{i-1}) \text{ and } j \in A(P_i) \\ min\{e(P_i), r^* + R\} & \text{otherwise} \end{cases} \quad (1)$$

In other words, a source is allowed to transmit at the explicit rate only if it has been active for the two previous intervals. Otherwise the allowed rate is upper bounded by $r^* + R$. Of course, this ER is inserted in the BRM cell only if it is less than the original value in the ER field. Note that the ER field of FRM cells originally generated by a source is always set to PCR.

The factor R helps constrain the rate of the sources that have just become active. The rationale is that if very few sources are active at the explicit rate calculation, ER can be large. If many of the previously inactive sources suddenly become active and are allowed to ramp up to ER as soon as possible, it can cause a large buffer backlog. By keeping the ACR of these sources within a constant initially, the algorithm allows the buffered link time to recognize these sources in its accounting of the number of active sources and also allows time for the corresponding ER to be propogated back to all sources.

This issue is also often ignored in existing ER schemes. For example, in the source behaviors specified in ATM Forum TM 4.0 [1], a source that just becomes active only needs to maintain its ICR until it receives the first BRM cell. It is expected that this duration corresponds to the source-to-destination round trip time. However, this assumes that the first BRM cell already contains the "correct" rate that this source should transmit at. Since practical ER schemes only update periodically, improper values in the initial BRM cells lead to unnecessary surges in input rate and consequently large oscillations in buffer occupancy until the ER algorithm catches up. The algorithm proposed here compensates for this by treating BRM cells intelligently.

From the implementation point of view, this implies that the switch needs to keep track of the list of active sources from the previous two update intervals and treat their backward RM cells differently from the rest. As we show in subsequent analyses, the factors r^* and R provide a degree of trade-off between how high the initial rates can be (important for fast transfer of short files) and how much network resources need to be reserved for ABR to absorb the surges caused by newly active sources.

3 Deterministic Analysis

In many ATM switches, the multiple QoS classes share the buffer and bandwidth resources dynamically with some minimum guarantees. For example, the bandwidth resource is allocated in a Generalized Processor Sharing (GPS) manner [7] that guarantees a minimum bandwidth for each QoS class queue while sharing the spare bandwidth efficiently. The buffer pool is shared in a similar manner, with some minimum and maximum bounds for each queue. Thus it is imperative for both switch designers and network operators to understand the relationship between the given traffic and these minimum allocation parameters, for both resource management and call admission control purposes.

For VBR service, much progress has been made in the area of effective bandwidth. Given the traffic characteristics, the effective bandwidth results allow a switch to determine the bandwidth and buffer requirements for achieving its QoS objectives (see, for example, [3]). Since ABR service is a closed-loop control, its resource requirements naturally depend primarily on the control mechanism and other system parameters such as the control delay.

In this section, we make a deterministic and conservative estimate of the amount of buffer and capacity resources that the network must reserve for ABR service with the above ER algorithm in order to guarantee zero loss.

The net drift of buffer occupancy $q(t)$ at the ABR buffered link is the difference between the total input rate and the output rate:

$$\frac{dq(t)}{dt} = \begin{cases} \sum_{j=1}^{n_{max}} r_j(t-\tau_j^f) - c(t) & \text{for } q(t) > 0 \\ \sum_{j=1}^{n_{max}} [r_j(t-\tau_j^f) - c(t)] \vee 0 & \text{for } q(t) = 0 \end{cases}$$

In the following we partition the sources into the ones whose rates are above $(r^* + R)$ and the ones below as seen by the buffered link:

$$\begin{aligned} \sum_{j=1}^{n_{max}} r_j(t-\tau_j^f) &= \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) \leq r^* + R\} \cdot r_j(t-\tau_j^f) \\ &+ \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^* + R\} \cdot r_j(t-\tau_j^f) \\ &\leq (r^* + R) \cdot n_{max} + \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^* + R\} \cdot r_j(t-\tau_j^f). \end{aligned}$$

Now consider the sources with rate $r_j(t-\tau_j^f) > r^* + R$ for $t \in [P_i, P_{i+1})$. Note that the active source j is bounded by the ER field in the last BRM cell that it receives prior to $t-\tau_j^f$. Because the intervals between RM cells vary over time, this BRM cell may carry either $e(P_i)$ or $e(P_{i-1})$. The reason is as follows: Since $t-\tau_j^f \geq P_{i-1}+\sigma+\tau_j^b$ for all j, we ensure that there is at least one BRM cell that reaches all sources after P_{i-1}, so that no source is governed by an ER older than $e(P_{i-1})$. On the other hand, for $t-\tau_j^f-\tau_j^b \geq P_i$, source j may have received $e(P_i)$.

Therefore

$$\begin{aligned} r_j(t-\tau_j^f) &\leq \max\{e(P_{i-1}), e(P_i)\} \\ &= \max\{\frac{a(P_{i-1})}{N(P_{i-1}) \vee 1}, \frac{a(P_i)}{N(P_i) \vee 1}\} \\ &\leq \frac{\max\{a(P_{i-1}), a(P_i)\}}{\min\{N(P_{i-1}), N(P_i)\} \vee 1} \end{aligned}$$

Next note that for a source to have rate $r_j(t-\tau_j^f) > r^*+R$ over the time interval $[P_i, P_{i+1})$, it must have been active over both intervals $[P_{i-2}, P_{i-1})$ and $[P_{i-1}, P_i)$. This is because of both Eq. 1 and the fact that if a source is idle for more than P, it must restart with its ICR, which is less than r^*+R. Thus we have $N(P_{i-1}) \geq \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^*+R\}$ and $N(P_i) \geq \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^*+R\}$. So $\min\{N(P_{i-1}), N(P_i)\} \vee 1 \geq \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^*+R\}$. Thus

$$\begin{aligned} & \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^*+R\} \cdot r_j(t-\tau_j^f) \\ \leq\ & \sum_{j=1}^{n_{max}} \frac{\max\{a(P_{i-1}), a(P_i)\}}{\min\{N(P_{i-1}), N(P_i)\}} \cdot 1\{r_j(t-\tau_j^f) > r^*+R\} \\ \leq\ & \max\{a(P_{i-1}), a(P_i)\} \cdot 1. \end{aligned}$$

Note that $f(q(P_i))$ is either k or $-k$, and that the rate of decrease in $c(t)$ is bounded by ρ. If we set $k = 2\rho P + n_{max} \cdot (r^*+R)$, we have

$$\begin{aligned} & \sum_{j=1}^{n_{max}} r_j(t-\tau_j^f) - c(t) \\ \leq\ & (r^*+R) \cdot n_{max} + \max\{a(P_{i-1}), a(P_i)\} - c(t) \\ \leq\ & (r^*+R) \cdot n_{max} + \max\{f(q(P_{i-1})) + c(P_{i-1}) - c(t), f(q(P_i)) + c(P_i) - c(t)\} \\ \leq\ & \begin{cases} 2k & \text{if } q(P_{i-1}) < q^* \text{ or } q(P_i) < q^* \\ 0 & \text{if } q(P_{i-1}) \geq q^* \text{ and } q(P_i) \geq q^* \end{cases} \end{aligned}$$

In other words, the net drift of the buffer occupancy $dq(t)/dt$ over the interval $[P_i, P_{i+1})$ can be upper bounded in one of two ways depending on how the occupancy compares to q^* at times P_{i-1} and P_i. This allows us to bound $q(t)$ at all times by $q(t) \leq q^* + 4kP$. The argument is as follows.

The net drift over $[P_i, P_{i+1})$ can be positive only if $q(P_{i-1}) < q^*$ or $q(P_i) < q^*$. Suppose we start with $q(P_{i-1}) < q^*$. The net drift over $[P_{i-1}, P_i)$ is bounded by $2k$ and $q(P_i) \leq q^* + 2kP$. However, note that the actual occupancy $q(P_i)$ may be either greater or less than q^*. First consider the case where $q(P_i) \geq q^*$. The net drift over $[P_i, P_{i+1})$ may still be positive and $q(t) \leq q^* + 4kP$ for all $t \in [P_i, P_{i+1})$. Again the actual occupancy at P_{i+1} may be greater or less than q^*. If it is greater, then the occupancy has exceeded q^* for two consecutive recalculation times and the net drift in the next interval, $[P_{i+1}, P_{i+2})$, must be negative (i.e., upper bounded by zero) and the occupancy must decrease. If $q(P_{i+1}) < q^*$, we can repeat the same argument as for $q(P_{i-1}) < q^*$ for the next two intervals. Similarly for the other case where $q(P_i) < q^*$.

Thus the buffer occupancy can exceed q^* for at most two consecutive recalculation times before the function $f(q(\cdot))$ leads to a negative drift. The most that $q(t)$ can grow up to is $q^* + 4kP$. In addition, note that in order for the ER value $e(P_i)$ to remain non-negative for all i, it is necessary that the ABR queue capacity $c(P_i) \geq k$ for all i. Combining the two, we have shown the following:

Proposition 1 *The ABR queue can ensure zero cell loss if the ER algorithm parameter k satisfies*

$$k = 2\rho P + n_{max}(r^* + R),$$

and its minimum bandwidth and buffer allocations, denoted as c_{min} and b_{min}, can be maintained above the following:

$$c_{min} \geq k,$$

$$b_{min} \geq q^* + 4kP = q^* + 4P(2\rho P + n_{max}(r^* + R)).$$

Intuitively, the requirements reflect the resources required to absorb the two perturbations to the closed-loop control: The first term $2\rho P$ corresponds to the effect of the interfering CBR and VBR traffic on the available ABR capacity. The second term $n_{max}(r^* + R)$ accounts for the newly active sources in their initial bursts before the ER algorithm can respond.

4 Probabilistic Analysis

One drawback with the deterministic solution is that if the total number of ABR connection is large and only a small fraction of the sources happens to be non-idle at any time, the $n_{max}(r^* + R)$ estimate may be overly conservative. It may be desirable to relax the lossless constraint consider in Section 3 and trade off a small probability of loss for a reduction in the resource requirements (or conversely an increase in the number of connections supported).

In this section, we introduce a probabilistic analysis that relates buffer overflow probability to resource allocation. The numerical results in the next section demonstrate that indeed by allowing a very small overflow probability, the resource requirements can be significantly reduced.

4.1 Additional Source Assumptions

At any time t, all ABR sources, as seen at the buffered link in question, can be identified as being in one of three possible states: (1) idle state, with $r_j(t - \tau_j^f) = 0$; (2) ER state, where the source's allowed cell rate is controlled by a previously calculated explicit rate; and (3) initial state, where the source recently transitions from idle to active and its rate is still upper bounded by $r^* + R$. We refer to sources in the third state as the "new sources" and denote such set as $A_{new}(t)$.

Next define $N_{new}(P_i)$ as the total number of new sources in internal $[P_i, P_{i+1})$:

$$N_{new}(P_i) = \sum_{j=1}^{n_{max}} 1\{j \in A_{new}(s) \text{ for some } s \in [P_i, P_{i+1})\}.$$

Thus a source is counted toward $N_{new}(P_i)$ if it is in the initial state for some interval over $[P_i, P_{i+1})$. Note that since a source in the ER state cannot transition into the initial state until it has been idle for more than P time unit, $N_{new}(P_i)$ can be no greater than (n_{max} - the number of sources that are in the ER state and are still active at time P_i). In this section we assume the following: for source j that does not belong to the set of sources that are active and in the ER state at time P_i, denote

$$\gamma = P[j \in A_{new}(s) \text{ for some } s \in [P_i, P_{i+1})],$$

for all time intervals $[P_i, P_{i+1})$. Hence we ignore the correlation between successive update intervals.

The number of new sources in each update interval therefore has a binomial distribution. Furthermore, this number is statistically smaller than a random variable with binomial (n_{max}, γ) distribution. For n_{max} large and γ small, this binomial distribution can be approximated by a Poisson distribution with mean m, where $m = n_{max} \cdot \gamma$.

4.2 Analysis

For $t \in [P_i, P_{i+1})$, again we partition the sources as in Section 3:

$$\begin{aligned}
\sum_{j=1}^{n_{max}} r_j(t-\tau_j^f) &= \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) > r^* + R\} \cdot r_j(t-\tau_j^f) \\
&+ \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) \leq r^* + R\} \cdot r_j(t-\tau_j^f) \\
&\leq \max\{a(P_{i-1}), a(P_i)\} + \sum_{j=1}^{n_{max}} 1\{r_j(t-\tau_j^f) \leq r^* + R\} \cdot r_j(t-\tau_j^f) \\
&\leq \max\{a(P_{i-1}), a(P_i)\} + (r^* + R) \cdot N_{new}(P_i).
\end{aligned}$$

Thus for $t \in [P_i, P_{i+1})$,

$$\begin{aligned}
\frac{dq(t)}{dt} &= \sum_{j=1}^{n_{max}} r_j(t - \tau_j^f) - c(t) \\
&\leq \max\{f(q(P_{i-1})) + c(P_{i-1}) - c(t),\ f(q(P_i)) + c(P_i) - c(t)\} + (r^* + R) \cdot N_{new}(P_i) \\
&\leq \begin{cases} k + 2\rho P + (r^* + R) \cdot N_{new}(P_i) & \text{if } q(P_{i-1}) < q^* \text{ or } q(P_i) < q^* \\ -k + 2\rho P + (r^* + R) \cdot N_{new}(P_i) & \text{if } q(P_{i-1}) \geq q^* \text{ and } q(P_i) \geq q^* \end{cases}
\end{aligned}$$

The probability that buffer overflow occurs in a given update interval can be upper bounded:

$$\begin{aligned}
& P[\text{buffer overflow in } [P_i, P_{i+1})] \\
= \quad & P[q(P_{i-1}) < q^*] \cdot P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) < q^*] \\
+ \quad & P[q(P_{i-1}) \geq q^*] \cdot P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) \geq q^*] \\
\leq \quad & P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) < q^*] + P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) \geq q^*]
\end{aligned}$$

We next bound the two terms separately:

$$\begin{aligned}
& P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) < q^*] \\
\leq \quad & P[q^* + (k + 2\rho P) \cdot 2P + (r^* + R) \cdot P \cdot (N_{new}(P_{i-1}) + N_{new}(P_i)) > b] \\
= \quad & P[N_{new}(P_{i-1}) + N_{new}(P_i) > \frac{b - q^* - (k + 2\rho P) \cdot 2P}{(r^* + R) \cdot P}]
\end{aligned}$$

Since $N_{new}(P_{i-1})$ and $N_{new}(P_i)$ are both statistically smaller than Poisson(m), $N_{new}(P_{i-1}) + N_{new}(P_i)$ is statistically smaller than Poisson($2m$). Thus the first term is upper bounded by the tail distribution of Poisson($2m$).

The second term is bounded as follows:

$$\begin{aligned}
& P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) \geq q^*] \\
\leq \quad & P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) \geq q^*, q(P_i) \geq q^*] \\
+ \quad & P[\text{overflow in } [P_i, P_{i+1}) \mid q(P_{i-1}) \geq q^*, q(P_i) < q^*] \\
\leq \quad & P[-k + 2\rho P + (r^* + R) \cdot N_{new}(P_i) > 0] \\
+ \quad & P[q^* + (k + 2\rho P) \cdot P + (r^* + R) \cdot P \cdot N_{new}(P_i) > b]
\end{aligned}$$

$$= \quad P[N_{new}(P_i) > \frac{k - 2\rho P}{r^* + R}] + P[N_{new}(P_i) > \frac{b - q^*}{(r^* + R)P} - \frac{k + 2\rho P}{r^* + R}]$$

The first inequality is obtained by applying Bayes's theorem again and using the fact that $P[q(P_i) \geq q^* \,|\, q(P_{i-1}) \geq q^*] \leq 1$ and $P[q(P_i) < q^* \,|\, q(P_{i-1}) \geq q^*] \leq 1$. The first term of the second inequality is as follows: Although we cannot upper-bound the queue size at the beginning of the interval, overflow for this case can only possibly happen if the drift $dq(t)/dt$ is greater than zero. Since for any $t \in [P_i, P_{i+1})$, $dq(t)/dt \leq -k + 2\rho P + (r^* R) \cdot N_{new}(P_i)$ when $q(P_{i-1}) \geq q^*$ and $q(P_i) \geq q^*$, we have $P[\text{overflow in } [P_i, P_{i+1}) \,|\, q(P_{i-1}) \geq q^*, q(P_i) \geq q^*] \leq P[dq(t)/dt > 0 \text{ for some } t \in [P_i, P_{i+1}) \,|\, q(P_{i-1}) \geq q^*, q(P_i) \geq q^*] \leq P[-k + 2\rho P + (r^* + R) \cdot N_{new}(P_i) > 0]$.

Thus $P[\text{buffer overflow in } [P_i, P_{i+1})]$ can be upper bounded by the sum of three tail probability values. Note that in general such sum can be great than 1. A necessary but not sufficient condition for the bound to be non-trivial is the following: $k > 2\rho P$ and $b > q^* + (k + 2\rho P) \cdot 2P$.

Proposition 2 *Under the probabilistic assumption for the sources, if the ABR buffer size b satisfies $b > q^*$ and the ABR available capacity $c(t)$ satisfies $c(t) \geq k$ and $dc(t)/dt \geq -\rho$, the buffer overflow probability can be upper-bounded as*

$$\begin{aligned}
& P[\text{buffer overflow in } [P_i, P_{i+1})] \\
\leq \quad & P[N_{new}(P_{i-1}) + N_{new}(P_i) > \frac{b - q^* - (k + 2\rho P) \cdot 2P}{(r^* + R) \cdot P} \\
+ \quad & P[N_{new}(P_i) > \frac{k - 2\rho P}{r^* + R}] + P[N_{new}(P_i) > \frac{b - q^*}{(r^* + R)P} - \frac{k + 2\rho P}{r^* + R}] \\
\leq \quad & F_{2m}(\frac{b - q^* - (k + 2\rho P) \cdot 2P}{(r^* + R) \cdot P}) + F_m(\frac{k - 2\rho P}{r^* + R}) + F_m(\frac{b - q^*}{(r^* + R)P} - \frac{k + 2\rho P}{r^* + R}) \quad (2)
\end{aligned}$$

where $F_x(\cdot)$ is the tail probabilty of a Poisson distribution with mean x, and $m = n_{max} \cdot \gamma$.

5 Numerical Examples

In this section we look at a few numerical examples to illustrate the effects of the ER algorithm parameters on the ABR resource requirements. Furthermore, we demonstrate how the probabilistic approach can lead to a much smaller resource requirements.

Consider the case where the buffered link has a OC-3c (150 Mbps) capacity and that the system has the following set of parameters:

- $\rho = 10$ Mbps/s
- $r^* = 50$ Kbps
- $R = 50$ Kbps
- $\tau = 0.1$ sec
- $\sigma = 0.1$ sec

Thus the smallest P would be $\sigma + \tau = 0.2$ sec.

5.1 Deterministic Approach

Case (1): First consider the case where the total number of ABR sources $n_{max} = 100$.

By setting $k = 2\rho P + n_{max}(r^* + R) = 2 \cdot 10 \cdot 0.2 + 100 \cdot 0.1 = 14$ Mbps, zero loss from the ABR buffer can be guaranteed if the minimum capacity requirement $c_{min} \geq k = 14$ Mbps and the minimum ABR buffer size $b_{min} \geq q^* + 4kP = q^* + 11.2$ Mbits, which amounts to be about 26K cells in addition to the target buffer occupancy q^*. This means that the minimum capacity that needs to be reserved for the ABR queue is less than 10% of the total capacity. The buffer requirement is also well within the range of many current ATM switches.

Case (2): Next consider the case where the total number of ABR sources $n_{max} = 10,000$. This can happen particularly in the Permanent Virtual Connection (PVC) environment, where a large number of VCs are provisioned and only a small fraction of them are expected to be active at any one time.

The resource requirements in this case become $c_{min} \geq k = 2 \cdot 10 \cdot 0.2 + 10000 \cdot 0.1 = 1{,}004$ Mbps and $b_{min} = q^{*}+$ 803.2 Mbits, or q^{*} + 1894K cells. Note that the minimum bandwidth requirement exceeds the total capacity of the buffered link and the buffer requirement also well exceeds the range of practical systems. This demonstrates the need for a probabilistic approach that accounts for the "new" sources less conservatively.

5.2 Probabilistic Approach

Case (3): Now suppose there are $n_{max} = 10,000$ sources and $\gamma = 0.01$, so that on the average the number of new sources in an update interval is limited to $m = n_{max} \cdot \gamma = 100$.

Consider the case where the ABR buffer size $b = q^{*} + 11.2$ Mbits and capacity $c(t) \geq k = 14$ Mbps. These are the buffer and capacity requirements in Case (1). We have

$$
\begin{aligned}
& P[\text{overflow in } [P_i, P_{i+1})] \\
\leq \; & P[N_{new}(P_{i-1}) + N_{new}(P_i) > 200] + P[N_{new}(P_i) > 100] + P[N_{new}(P_i) > 380] \\
\leq \; & 0.9546,
\end{aligned}
$$

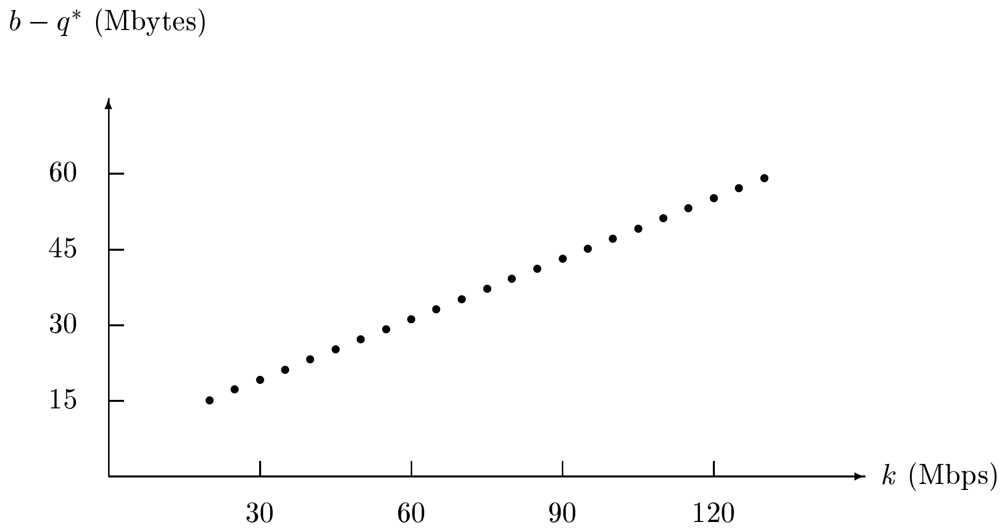
which shows that the given resources are too small to provide satisfactory performance. However, in the next example, a slight increase of the resources is sufficient to significantly reduce the overflow probability.

Case (4): We still assume that $n_{max} = 10,000$ and $\gamma = 0.01$. Let $b = q^{*}+$ 22.4 Mbits and $c(t) \geq k = 20$ Mbps.

$$
\begin{aligned}
& P[\text{overflow in } [P_i, P_{i+1})] \\
\leq \; & P[N_{new}(P_{i-1}) + N_{new}(P_i) > 640] + P[N_{new}(P_i) > 160] + P[N_{new}(P_i) > 880] \\
\leq \; & 1.262 \times 10^{-8}
\end{aligned}
$$

Note that this is a significant reduction compared to the deterministic resource requirements.

Figure 1: The ABR resource requirements $b - q^*$ and $c_{min} = k$ sufficient for achieving an overflow probability of 10^{-7}



Case (5): We will use the same $n_{max} = 10,000$ and $\gamma = 0.01$. Figure 1 shows the combination of buffer size (the required buffer amount above the occupancy target q^*) and k factor for achieving an overflow probability of 10^{-7}. The minimum ABR capacity should exceed k. Note that the buffer requirement increases with increasing k, since a larger k reflects a greater fluctuation in explicit rate assignment (i.e., the algorithm becomes more sensitive to queue occupancy when the rate factor k is large).

6 Concluding Remarks

This paper introduces one simple explicit rate algorithm that takes into account system parameters such as feedback delays and frequency of RM cells. We show how resource requirements can be determined based on the rate algorithm parameters and loss objective. Such results can be used to determine resource assignment for a given number of ABR connections, or conversely to determine the number of acceptable ABR connections with a given resource allocation in Call Admission Control (CAC). We show that in order to maintain the loss objective, some minimum bandwidth and buffer should be reserved for the ABR queue.

In the remainder of this section, we offer some concluding remarks on further extensions.

6.1 Incorporating Minimum Cell Rate

ABR VC's are allowed to specify a Minimum Cell Rate (MCR). MCR is the minimum amount of bandwidth that the network guarantees to a particular VC. This can be specified at call set-up or provision time. The ER algorithm and the analyses in this paper so far deals with zero MCR for all VCs. They can be easily extended to cover non-zero MCR.

There are a few alternative definitions of fairness in the non-zero MCR situation. The most applicable one is known as "MCR plus equal share." In this case, a portion of the bandwidth available to ABR connections is reserved for the guaranteed MCRs. The remaining available bandwidth is then equally divided among all active ABR connections. The ER algorithm can thus be extended by applying it only to the remaining portion as follows.

- Define MCR_j as the MCR for source j and MCR_{total} as the sum of all MCRs
- Redefine $c(t)$ as ((the capacity available to the ABR queue at time t) - MCR_{total})
- Let $(e_j(t) + MCR_j)$ be the calculated ER for source j
- Redefine $r_j(t)$ as ((the instantaneous rate of source j) - MCR_j)
- The minimum capacity that must be reserved for ABR is therefore $c_{min} + MCR_{total}$

Since MCR is thus treated as a separate constant bit rate portion of an ABR connection, the remainder of the analyses still follows.

6.2 Effect of Under-Utilization

In this work we focus on a single ABR buffered link, so that the ER for every active source is the same, reflecting the fair share per source. In a practical network where a virtual circuit may traverse multiple switches, some ABR sources may generate less than their fair shares due to a number of factors: (1) the sources may be constrained by other link segments due to the ERs calculated there; (2) the traffic sources may be bursty in nature, so that they do not always have enough data to transmit at the allowed cell rates.

The presence of such sources can cause the ABR bandwidth to be under-utilized. Our simple ER algorithm partially accounts for this problem by setting the total allocated ABR bandwidth $a(P_i)$ to be $c(P_i) + k$ when the buffer occupancy is below q^*, so that the allocated bandwidth is more than the actual available bandwidth. However, when the number of under-utilizing ABR sources is large, the actual ABR bandwidth consumed may be less than the available ABR bandwidth even in that case.

In networks where Unspecified Bit Rate (UBR) sources are also present, it is often assumed that the residual bandwidths from higher classes (i.e., CBR, VBR, and ABR) can be consumed by UBR traffic and thus not considered wasted. We may view under-utilized ABR bandwidth as such. Alternatively, an ATM switch may attempt to address under-utilization by selectively increasing the ER of unconstrained ABR sources (i.e., sources that are bottlenecked at this switch). We outline one scheme as follows: To identify the amount of under-utilization and the number of unconstrained sources, an ATM switch first estimates the Current Cell Rate (CCR) of each ABR VC, which can be obtained either from the CCR field in the VC's RM cells or from on-line measurements. The CCR is compared with the ER assignment to estimate whether the source is under-utilizing its allocated bandwidth and the amount of under-utilization. The trade-off between the potential benefit of such approach and the increased complexity in ER assignment requires further investigation.

Acknowledgement

The authors would like to thank the reviewer's thorough review and comments on this paper.

References

[1] ATM Forum, "ATM Forum Traffic Management Specification Version 4.0", af-tm-0056.000, April, 1996.

[2] Flavio Bonomi and Kerry W. Fendick, "The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service", *IEEE Network Magazine*, March/April 1995, pp. 25-39.

[3] G. de Veciana, G. Kesidis, and J. Walrand, "Resource Management in Wide-Area ATM Networks Using Effective Bandwidths," IEEE Journal on Selected Areas in Communications, vol. 13, pp. 1081-1090, Aug. 1995.

[4] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey", Computer Networks and ISDN Systems, vol. 28, no. 13, pp. 1723-1738 Feb. 1995.

[5] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis, "Weighted Round-Robin Cell Multiplexing in a General-Purpose ATM Switch Chip," IEEE Journal on Selected Areas in Communications, vol. 9, no. 8, pp. 1265-1279, Aug. 1991.

[6] Jean-Yves Le Boudec, Gustavo de Veciana, and Jean Walrand, "QoS in ATM: Theory and Practice", *Proc. 35th IEEE Conference on Decision and Control*, pp. 773-778, Dec. 1996.

[7] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: the Single-Node Case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 334-357, June 1993.

[8] Kai-Yeung Siu and Hong-Yi Tzeng, "Intelligent Congestion Control for ABR Service in ATM Networks", *ACM Computer Communication Review*, vol. 24, no. 5, pp. 81-106, Oct. 1994.

[9] Ching-Fong Su, Gustavo de Veciana, and Jean Walrand, "Explicit Rate Flow Control for ABR Services in ATM Networks", *preprint*, available at http://www.ece.utexas.edu/ gustavo/papers.html.

[10] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet Switching Networks," *Proceedings of the IEEE*, Oct. 1995.