To Appear: Proc. IEEE ICDS-16 Conference, Hongkong, May 1996.

A Multi-level Explicit Rate Control Scheme for ABR Traffic with Heterogeneous Service Requirements

Jörg Liebeherr † *Ian F. Akyildiz* ‡ *Alan Tai* †

† Computer Science Department
University of Virginia
Charlottesville, VA 22903

‡ Broadband and Wireless Networking Laboratory
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

The Available-Bit-Rate (ABR) service that is being standardized by the ATM Forum dynamically determines the maximum transmission rate, the explicit rate, *of a connection. A drawback of the dynamic control scheme for calculating the explicit rates is that it tries to allocate the same bandwidth to all ABR connections regardless of the application type of the connection. In this study a multi-level explicit rate scheme is proposed that can allocate different transmission rates bounds to connections. ABR traffic is controlled at three levels. At the topmost level, bandwidth is dynamically regulated between CBR, VBR, and ABR traffic sources. At the next level, bandwidth is controlled between different classes of ABR traffic. At the lowest level, bandwidth is distributed among connections belonging to the same ABR traffic class. The effectiveness of the proposed scheme is demonstrated in simulation experiments.*

1 Introduction

The *Available-Bit-Rate* (ABR) service class currently being defined by the ATM Forum [10] completes the suite of services available in ATM networks. The goal of the ABR service is to efficiently support traffic types that were traditionally carried in packet switching networks, i.e., highly bursty traffic with only vaguely defined quality-of-service requirements. To establish an ABR connection both maximum traffic demand (*peak cell rate*) and the minimum throughput requirements (*minimum cell rate*) need to be specified. During the lifetime of an ABR connection, the ATM network can adjust the maximum traffic rate of the connection to any value in the range between the minimum and the peak cell rate.

Several mechanisms for controlling ABR traffic have been proposed to the ATM Forum. Prominent among these mechanisms are the *credit-based* congestion control approach and the *rate-based* flow control approach. Credit-based congestion control of ABR traffic is based on a modified window flow control mechanism implemented on a per-link basis [7]. In contrast, the rate-based approach [9, 11] controls congestion on an end-to-end basis. Here, the ATM switches give feedback information to the traffic sources which, in turn, adjust their traffic rate. The ATM Forum decided in September 1994 to adopt the rate-based approach in preference over the credit-based approach.

In this study we address the problem of determining the explicit rate for ABR connections in an ATM network with rate-based traffic control. The ATM Forum has adopted the so-called *max-min fairness* or *bottleneck flow control* [1, 4] scheme as objective for the rate control algorithm [5, 10]. Max-min fairness attempts to give each ABR connection the same maximum throughput allocation on each link in the network. However, since ABR connections may result from a heterogeneous set of network applications, ranging from interactive bulk data transfers to video-conferencing applications, the explicit rate calculation should distinguish between and give differential treatment to different classes of ABR connections.

In this study we present an explicit rate scheme for ATM networks that can give differential treatment to different types of ABR connections. The scheme determines the explicit rate of ABR connections by simultaneously satisfying several control objectives. Each ABR connection belongs to one **traffic class** where the class assignment of the connection is based on the application type, on traffic parameters, or on extraneous factors, such as the location of the traffic source or a pricing scheme. The main advantage of our scheme over previously proposed enhancements to max-min fairness is that our method enables to completely decouple the bandwidth allocation for connections from different traffic classes. To our knowledge, our work is the first scheme that provides explicit rates for multiple ABR traffic classes where the explicit rate of a class is calculated independently from the other classes.

We control the availability of bandwidth at three levels.

[0] The work of Jörg Liebeherr and Alan Tai was supported in part by the National Science Foundation under Grant No. NCR-9309224.

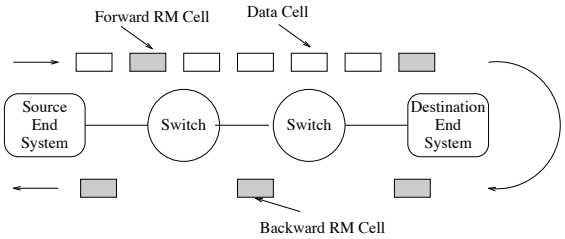


Figure 1: Rate-Based Traffic Control.

• **Service-Level Bandwidth Control**, the highest control level, determines the capacity to available to ABR traffic on each link. This capacity allocated to ABR traffic is dependent on the current demand due to CBR and VBR traffic. [1]

• **Class-Level Bandwidth Control** distributes the available ABR bandwidth on a link among multiple traffic classes. Class-level bandwidth control attempts to distribute unused bandwidth by increasing the class guarantees and capacities of traffic classes with a high bandwidth demand.

• **Connection-Level Bandwidth Control** involves the allocation of bandwidth to individual ABR connections within a traffic class at a link. A *share* provides the maximum bandwidth available to each connection from this class at each link. The maximum end-to-end throughput of a connection is limited by the link with the smallest share on the connection's route.

The remainder of the paper is structured as follows. In Section 2 we review the ABR traffic control scheme of the ATM Forum. In Section 3 we characterize the multi-level bandwidth control scheme. In Section 3, we how to modify the existing ABR traffic control protocol to incorporate our multi-level bandwidth control. In Section 5 we show simulation experiments that demonstrate the effectiveness of our scheme. Finally, we conclude our results in Section 6.

2 Rate-Based Traffic Management of ABR Traffic

Since the first proposals for closed-loop rate-based traffic control within the Traffic Management Group of the ATM Forum in 1994, many additions to the protocol have significantly enhanced its functionality. The current draft version of the Traffic Management document [10] uses more than two dozen parameters for the control algorithm. Excellent discussions of the rate-based approach for ABR traffic can be found in [2, 5].

Rate-based traffic control in ATM consists of a closed feedback loop involving the source end system, the destination end system, and the intermediate ATM switches. The basic steps of the control loop are shown in Figure 1. The source end system periodically generates *Resource Management (RM)* cells that are interleaved with the stream of data cells. The RM cells travel to the destination as *forward RM cells*. The destination turns the RM cells around and sends them back to the source as *backward RM cells*. On its round trip, the RM cell collects congestion information from the switches and the destination. This information is used by the source to adjust its traffic rate.

Three different feedback control schemes are used in the rate-based scheme:

1. *Explicit Forward Congestion Notification (EFCN)* [8, 11]: If switches are congested they set the *Explicit Forward Congestion Indication (EFCI)* flag in the headers of regular data cells. The destination, upon receiving a data cell with the EFCI flag set, sends the congestion information to the source by setting a flag in a backward RM cell.

2. *Backward Explicit Congestion Notification (BECN)* [9]: Here, congested switches can return an congestion indication to the sources directly by generating backward RM cells with a congestion flag set.

Both EFCN and BECN are *binary feedback* schemes in that the information returned to a source system merely consists of a single bit. In contrast, with explicit rate setting, the network informs the sources about the maximum permitted traffic rate.

3. *Explicit Rate Setting* [5, 10]: Here a source regularly emits RM cells with *Explicit Rate* (`ER`) set to the desired transmission rate. The ATM switches reduce the value in the `ER` field and return the RM cell back to the source. When the RM cell returns, the `ER` field contains the maximum traffic rate that is permitted by the network.

The actual implementation of the feedback control loop proposed by the ATM Forum is too complex to be presented here [10]. Consideration of long propagation delays, long idle times of sources, RM cell losses, and low bandwidth connections have significantly increased the complexity of the basic feedback schemes.

3 Multi-level Bandwidth Control of ABR Traffic

In this section, we develop a formal framework for the proposed multi-level explicit rate scheme for ABR traffic. We state the objectives of bandwidth control at three levels. At the lowest, we control the bandwidth available to ABR connections within the same traffic class. This level of control is currently well-understood and applied by the ATM Forum to calculate the explicit rate of ABR connections. At the second control level, we dynamically control the bandwidth available to the ABR traffic classes. The control method takes into account the current bandwidth use of the traffic classes: classes with a high bandwidth demand can temporarily borrow bandwidth from traffic classes with a low bandwidth demand. Finally, at the topmost level, we

[1] We ignore the presence of UBR traffic classes in this study.

control the availability of bandwidth to all ABR connections in the network. The goal of the multi-level control scheme is to find for each connection the maximum traffic rate which complies to the control objectives, i.e., the explicit rate. We consider an ATM network where switches are connected by unidirectional ATM links. Each ABR connection in this network has a fixed route with an unidirectional traffic flow and is assigned to exactly one traffic class. We introduce the following notation:

$\mathcal{L}$	Set of (unidirectional) links in the ATM network.
C_l	Capacity of link l.
P	Number of traffic classes.
$\mathcal{F}_p$	Set of all class-p connections ($1 \leq p \leq P$).
$\mathcal{R}_i$	Route of ABR connection i; $\mathcal{R}_i = (l_{i_1}, l_{i_2}, \ldots, l_{i_K})$ where $l_{i_k} \in \mathcal{L}$ is the kth link on the route of connection i.
Δ_{lp}	Set of connections in class p with link l on their route ($\Delta_{lp} = \{i \mid l \in \mathcal{R}_i,\ i \in \mathcal{F}_p\}$).

The traffic demand of an ABR connection i is expressed in terms of the *peak cell rate*, denoted by $\mathtt{PCR}_i$, and the *minimum cell rate*, denoted by $\mathtt{MCR}_i$. If the network does not have sufficient bandwidth to satisfy $\mathtt{MCR}_i$, then the connection will not be established.

The *maximum throughput* of a connection allowed by the bandwidth control scheme is called the *explicit rate*, and denoted by $\mathtt{ER}_i$ for connection i. $\mathtt{ER}_i$ is the (theoretical) maximum traffic rate allowed by the multi-level bandwidth control mechanism, irrespective of other throughput constraints, e.g., due to a BECN or FECN scheme, in the network. We have the following relation:

$$\mathtt{MCR}_i \leq \mathtt{ER}_i \leq \mathtt{PCR}_i$$

The bandwidth control scheme for ABR traffic consists of imposing bounds on the explicit rate $\mathtt{ER}_i$ by enforcing a set of control parameters for each link in the network. The control parameters used in this study are as follows:

• For **connection-level bandwidth control** the network enforces class-dependent throughput bounds on all connections at all links. The class-p *share* for link l, denoted by $Share_{lp}$, is the throughput bound for all ABR connections from class p at link l. Formally:

$$\mathtt{ER}_i \leq Share_{lp} \qquad \text{for all } i \in \Delta_{lp}$$

• **Class-level bandwidth control** enforces throughput bounds for the aggregate bandwidth used by connections from a class at a link. We use $C_{lp}^{\mathtt{ABR}}$ to denote the available capacity for all class-p connections at link l, so-called *class capacity*. For all traffic classes p ($1 \leq p \leq P$) we have:

$$\sum_{i \in \Delta_{lp}} \mathtt{ER}_i \leq C_{lp}^{\mathtt{ABR}} \qquad \text{for all } l \in \mathcal{L}$$

• **Service-level bandwidth control** bounds the aggregate throughput of ABR connections from all classes on a link by a so-called *ABR capacity*, denoted by $C_l^{\mathtt{ABR}}$, that is,

$$\sum_{p=1}^{P} \sum_{i \in \Delta_{lp}} \mathtt{ER}_i \leq C_l^{\mathtt{ABR}} \qquad \text{for all } l \in \mathcal{L}$$

3.1 Connection-Level Bandwidth Control

In this subsection we ignore the effects of class-level and service-level control. We do this by assuming that all class capacities for ABR traffic are fixed, i.e., $C_{lp}^{\mathtt{ABR}} \equiv const.$ In this case, the bandwidth left unused by some traffic class cannot be made available to other traffic classes. With the assumption of fixed class capacities, the connection admission control test for an ABR connection from class p verifies that the minimum cell rate $\mathtt{MCR}_j$ can be supported on all links on the route of the connection, i.e.,

$$\sum_{i \in \Delta_{lp}} \mathtt{MCR}_i \leq C_{lp}^{\mathtt{ABR}} \qquad \text{for all } l \in \mathcal{R}_j$$

Connection-level bandwidth control distributes the class capacity $C_{lp}^{\mathtt{ABR}}$ to the class- p connections on a link l. By enforcing shares $Share_{lp}$ for each traffic class p at the network links, the maximum end-to-end throughput of an ABR connection i is limited by the link l_i^* on the connection's route with the smallest share, i.e., $Share_{l_i^* p} = \min_{l \in \mathcal{R}_i} Share_{lp}$. We refer to link l_i^* as the *bottleneck link* of connection i. Since the maximum throughput of a connection is always bounded by the peak cell rate $\mathtt{PCR}_i$, we obtain the following throughput bound for connection i:

$$\mathtt{ER}_i = \min\Big(\mathtt{PCR}_i, Share_{l_i^* p}\Big)$$

Note that a connection-level control scheme that is based on enforcing *shares* at the network links implements an intuitive notion of *fairness*, in the sense that all connections from the same class with the same bottleneck link have identical throughput constraints [3, 4].

A control scheme that selects the values of the fair shares too small may waste a significant portion of the bandwidth. Thus, a bandwidth control scheme will attempt to make the values of the fair shares as large as possible. With fixed class capacities, such a bandwidth control scheme is identical to *max-min fairness*, a scheme that has been considered for ABR traffic control in several different versions [5].

We determine the values of the maximal fair shares as follows. Given the share values on each link (not necessarily maximal), we partition the set of class-p ABR connections on a link l into three groups: *underloaded* connections, *overloaded* connections, and *restricted* connections. Let the set of underloaded connections, denoted by $\mathbf{U}_{lp}$, contain all class-p connections at link l that can satisfy their end-to-end bandwidth demand, i.e., $\mathtt{ER}_i = \mathtt{PCR}_i$. All other connections have bandwidth

requirements larger than their maximum throughput, i.e., $\mathtt{ER}_i < \mathtt{PCR}_i$; these connections are classified as overloaded or restricted. Connections that are 'overloaded at link l' have link l as their bottleneck. Connections at link l that are 'restricted' have their bottleneck on some link k on their route with $k \neq l$. For link l, let $\mathbf{O}_{lp}$ be the set of overloaded class-p connections, and let $\mathbf{R}_{lp}(k)$ be the set of restricted class-p connections that have their bottleneck at link k. The sets $\mathbf{U}_{lp}$, $\mathbf{O}_{lp}$ and $\mathbf{R}_{lp}(k)$ are specified as follows:

$$\begin{aligned}
\mathbf{U}_{lp} &= \left\{ i \in \Delta_{lp} \mid Share_{lp} \geq \mathtt{PCR}_i \,,\ i \notin \bigcup_{k \in \mathcal{L}} \mathbf{R}_{lp}(k) \right\} \\
\mathbf{O}_{lp} &= \left\{ i \in \Delta_{lp} \mid l = l_i^* \,,\ Share_{lp} < \mathtt{PCR}_i \right\} \\
\mathbf{R}_{lp}(k) &= \left\{ i \in \Delta_{lp} \mid k = l_i^* \,,\ Share_{kp} < \mathtt{PCR}_i \right\}, k \neq l
\end{aligned}$$

With the above definitions we can now characterize the maximal shares. Obviously, with maximal shares the entire class capacity C_{lp}^{ABR} on a link l is utilized if there is at least one connection in class p that is overloaded on this link. We obtain:

$$\begin{aligned}
C_{lp}^{\mathrm{ABR}} &= \sum_{i \in \Delta_{lp}} \mathtt{ER}_i = \sum_{i \in \Delta_{lp}} \min\left(\mathtt{PCR}_i, Share_{l_i^* p}^*\right) \\
&= |\mathbf{O}_{lp}| \cdot Share_{lp}^* + \sum_{i \in \mathbf{U}_{lp}} \mathtt{PCR}_i + \\
&\quad + \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share_{kp}^*
\end{aligned}$$

For links without overloaded class-p connections ($\mathbf{O}_{lp} = \emptyset$), we set $Share_{lp}^* = C_{lp}^{\mathrm{ABR}}$. Then we obtain the following values for the maximal shares:

$$Share_{lp}^* = \begin{cases} C_{lp}^{\mathrm{ABR}} & \text{if } \mathbf{O}_{lp} = \emptyset \\ \dfrac{C_{lp}^{\mathrm{ABR}} - \sum_{i \in \mathbf{U}_{lp}} \mathtt{PCR}_i - \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share_{kp}^*}{|\mathbf{O}_{lp}|} & \text{otherwise} \end{cases}$$

In other words, the maximal share is obtained by subtracting the throughput of the connections that are *not* overloaded from the class capacity, and by dividing the remaining bandwidth by the number of overloaded connections.

3.2 Class-Level Bandwidth Control

The bandwidth control scheme for calculating the explicit rates described so far has one major drawback: if the ABR connections in a class, say class p, do not consume the bandwidth C_{lp}^{ABR} that is available at link l, the unused bandwidth cannot be utilized by other traffic classes. Next we show how the drawback can be overcome by adapting the available capacity C_{lp}^{ABR} to the actual traffic demand.

In the scheme proposed here, the class capacity C_{lp}^{ABR} consists of two components: the *class guarantee* $Guar_{lp}$ and the *surplus bandwidth* $Surplus_l$. The class guarantee $Guar_{lp}$ is a fixed component and gives the minimum bandwidth that ABR connections from class p can use for transmission at link l. We assume $\sum_{p=1}^{P} Guar_{lp} = C_l^{\mathrm{ABR}}$, that is, the class guarantees divide the entire ABR bandwidth on a link l.

The surplus bandwidth, denoted by $Surplus_l$, gives the bandwidth in excess of the class guarantee that is temporarily made available to a class. Of course, this is only possible if some other classes do not utilize their respective class guarantees, i.e., if $Guar_{lq} - \sum_{i \in \Delta_{lq}} \mathtt{ER}_i > 0$ for some traffic classes $q \neq p$. Note that a traffic class may not be able to utilize the class guarantee at a link for two reasons. First, the total peak cell rate from all connections of the class could be less than its guarantee. Second, the throughput of class-p connections could be limited due to restrictions at other links.

In our class-level bandwidth control scheme, we reduce class capacity C_{lp}^{ABR} for a class p at link l whenever the connections from this class do not utilize their class guarantee. The resulting bandwidth that is made available is distributed evenly to those traffic classes that can take advantage of the additional bandwidth. The bandwidth is made available by adding a surplus $Surplus_l \geq 0$ to the class capacity. Formally, the class capacity C_{lp}^{ABR} at link l for class p is set to:

$$C_{lp}^{\mathrm{ABR}} = \min\left(\sum_{i \in \Delta_{lp}} \mathtt{ER}_i, Guar_{lp} + Surplus_l \right)$$

The above equation assumes the enforcement of shares $Share_{lp}$ for all connections, as discussed in Subsection 3.1. Even though the concepts of 'shares' and 'surplus' are independent, we assume that the bandwidth control scheme enforces maximal shares on all links.

Since in the worst case $Surplus_l = 0$, each connection class with a sufficiently high traffic load and no other limiting constraints can always obtain the class guarantee as its class capacity. Therefore, the following connection admission control test for a new connection j with route $\mathcal{R}_j$ guarantees that all connections can receive their minimum cell rate $\mathtt{MCR}_i$ at all times:

$$\sum_{i \in \Delta_{lp}} \mathtt{MCR}_i \leq G_{lp} \qquad \text{for all } l \in \mathcal{R}_j$$

A goal of a bandwidth control scheme is to select the surplus values $Surplus_l$ as large as possible. If the surplus on a link l is chosen maximally, denoted by $Surplus_l^*$, then the entire link bandwidth can be made available for transmission if there is at least one overloaded connection in some traffic class on this link. Note that only traffic classes with overloaded connections at link l will be able to utilize their maximum bandwidth $Guar_{lp} + Surplus_l$. Assuming that at least one such class exists on a link l, i.e., $|\mathbf{O}_{lq}| > 0$, and that

the maximal shares $Share^*_{lp}$ are available for all links l, then a bandwidth control scheme which enforces the maximal fair shares and maximal surplus values satisfies the following equation:

$$\begin{aligned} C_l^{\text{ABR}} &= \sum_{\mathbf{O}_{lq}\neq\emptyset} \Big(Guar_{lq} + Surplus_l \Big) + \\ &\quad + \sum_{\mathbf{O}_{lq}=\emptyset} \left(\sum_{i\in\mathbf{U}_{lq}} \texttt{ER}_i + \sum_{k\in\mathcal{L}} \sum_{i\in\mathbf{R}_{lq}(k)} \texttt{ER}_i \right) \\ &= \sum_{\mathbf{O}_{lq}\neq\emptyset} \Big(Guar_{lq} + Surplus_l \Big) + \\ &\quad + \sum_{\mathbf{O}_{lq}=\emptyset} \Big(\sum_{i\in\mathbf{U}_{lq}} \texttt{PCR}_i + \sum_{k\in\mathcal{L}} |\mathbf{R}_{lq}(k)| \cdot Share^*_{kq} \Big) \end{aligned}$$

If the link does not have any overloaded connections, that is, $|\mathbf{O}_{lq}| = 0$ for all traffic classes, the surplus is selected to some large value, i.e., $Surplus_l = C_l^{\text{ABR}}$. Then we obtain that a bandwidth control scheme with maximal surplus satisfies:

$$Surplus_l^* = \begin{cases} C_l^{\text{ABR}} & \text{if } \bigcup_{q=1}^{P} \mathbf{O}_{lq} = \emptyset \\ \dfrac{1}{|\{q \mid 1\le q\le P,\ \mathbf{O}_{lq}\neq\emptyset\}|} \cdot \Bigg(C_l^{\text{ABR}} - \sum_{\mathbf{O}_{lq}\neq\emptyset} Guar_{lq} - \\ \quad - \sum_{\mathbf{O}_{lq}=\emptyset} \left(\sum_{i\in\mathbf{U}_{lp}} \texttt{PCR}_i + \sum_{k\in\mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share^*_{kp} \right) \Bigg) & \text{otherwise} \end{cases}$$

3.3 Service-Level Bandwidth Control

So far we have not accounted for the fact that the bandwidth available to ABR traffic is dependent on the bandwidth allocation to CBR and VBR connections. Service-level control adjusts the link bandwidth available to ABR traffic to the demands of CBR and VBR traffic. The control method is simple: CBR and VBR traffic is given priority over ABR traffic whenever possible.

To prevent ABR traffic becoming completely preempted, we introduce C_l^{min} as a lower bound for the ABR bandwidth available at link l. In addition to C_l^{min}, ABR traffic can obtain the bandwidth not used by connections with CBR or VBR service. Denoting by Γ_l^{CBR} and Γ_l^{VBR}, the current allocation at link l of CBR and VBR traffic, respectively, the bandwidth available to ABR traffic on a link l is set to:

$$C_l^{\text{ABR}} = \max\left(C_l^{min},\ \ C_l - \Gamma_l^{\text{CBR}} - \Gamma_l^{\text{VBR}} \right)$$

Note that changing C_l^{ABR} will typically require recalculation of the class guarantees G_{lp}.

If the ATM network wants to ensure that all ABR connections can satisfy their minimum cell rate, the following connection admission control test should be executed for all links that are on the route of a new connection j:

$$\sum_{p=1}^{P} \sum_{i\in\Delta_{lp}} \texttt{MCR}_i \le C_l^{min} \qquad \text{for all } l \in \mathcal{R}_j$$

4 Protocol Mechanisms for Multi-level Bandwidth Control

In this section we present how the proposed multi-level bandwidth control scheme from the previous section can be incorporated into the framework of the ABR traffic management protocol developed by the ATM Forum [10]. The protocol mechanisms described here are mainly modifications or additions to the ABR control protocol.

4.1 Modifications to the RM Cell Format

We require only a minor modification to the RM cell format described in [10]. All bit flags described in [10] are unchanged, and the use of the ER field is similar to, but not identical with [10]. The only addition to the RM cell format is the following field:

BNK The *Bottleneck (BNK) Field* contains a unique identification of an ATM switch or the destination system and identifies the bottleneck link of the connection. The field is set by the intermediate switches on the forward pass of the RM cell.

4.2 Source and Destination Behavior

If a class-p connection issues a forward RM cell, it sets the ER field to the peak cell rate, i.e., ER = PCR. And the bottleneck field to BNK = *nil*.

The state of an ABR connection, can be determined from the last backward RM cell that has returned to the source. If BNK = *nil*, then the connection is '*underloaded*' (in this case, the ER field will be set to ER = PCR). On the other hand, if the content of the last backward RM cell was BNK = S and ER = *rate* (with PCR > *rate*), then the connection i is '*overloaded*' at switch S and '*restricted*' at all other switches on the connection's route. Underloaded connections can transmit data at their peak cell rate. Overloaded connections can transmit at most at the rate given by the ER field of the last RM cell. Note that a source need not be aware of its state.

The destination plays no particular role in the calculation of the explicit rate. It receives forward RM cells and returns them back to the source as backward RM cells.

4.3 Switch Behavior

A switch maintains information on each ABR connection that has a route on one of the outgoing links of

the switch [2]. The information for a connection i consists of a rate field $\texttt{MaxRate}_i$ that contains the switch's current knowledge of the maximum allowed traffic rate of connection i and an *overload flag* $\texttt{OV}_i$ which is set when connection i is overloaded at this switch.

In addition, the switch maintains a set of variables needed for calculating the throughput bounds of the connections: $Share_p$ contains the maximum cell rate at which connections from class p can transmit at this switch, the bandwidth guarantee G_p and the $Surplus$ are used to calculate the total bandwidth available for traffic from class p at this switch. Finally, C^{ABR} is the total available capacity for ABR traffic at this switch.

The following operations are performed at a switch, say with identification T, when it receives a forward RM cell from a class-p connection i. The switch first compares the ER field with its value for $Share_p$. If $Share_p \geq \texttt{ER}$, then the switch does not perform any operations. On the other hand, if $Share_p < \texttt{ER}$, then the maximum rate at which connection i wants to transmit exceeds the maximum allowed rate for class-p connections at switch T. Therefore, the switch modifies the fields of the backward RM cell by setting:

$$\texttt{ER} = Share_p \qquad \texttt{BNK} = \texttt{T}$$

Thus, the switch sets the explicit cell rate to the maximum rate at this switch, and identifies itself as the bottleneck of the connection.

The following steps are performed at a switch, say switch S, when it receives a backward RM cell from class-p connection i. If the bottleneck field of the RM cell is set to $\texttt{BNK} = nil$, that is, the connection is underloaded, or to $\texttt{BNK} = \texttt{S}$ ($\texttt{S} \neq \texttt{T}$), that is, the connection is overloaded at some other switch, switch S updates its information on connection i by setting:

$$\texttt{MaxRate}_i = \texttt{ER} \qquad \texttt{OV}_i = 0$$

Note that ER for an underloaded connection ($\texttt{BNK} = nil$) is set to the the peak cell rate $\texttt{PCR}_i$.

If the bottleneck field of the backward RM cell that arrives to switch T is set to $\texttt{BNK} = \texttt{T}$, the information on the connection is updated to:

$$\texttt{MaxRate}_i = 0 \qquad \texttt{OV}_i = 1$$

4.4 Operations at Update Intervals

Periodically, a switch uses the values of $\texttt{MaxRate}_i$ and $\texttt{OV}_i$ to calculate new throughput bounds for the connections. Each switch uses timers to keep track of three different time intervals: The *share interval*, the *surplus interval*, and the *ABR capacity interval*. We assume that the surplus interval is a multiple of the share intervals, and the capacity interval is a multiple of the surplus interval.

[2] To simplify the notation, we assume in the following that a switch has exactly one outgoing link.

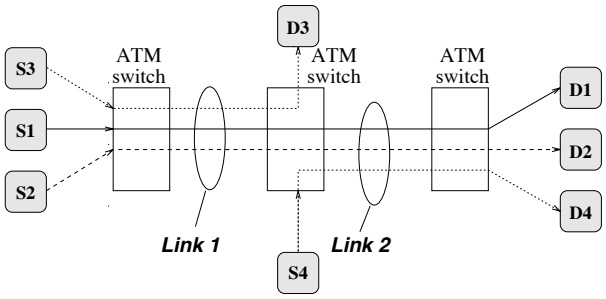


Figure 2: Simulated Network.

At the end of each share interval, switch S calculates the following sums for each traffic class p:

$$\begin{aligned} \texttt{OL}_p &:= \textstyle\sum_{i \in \mathcal{F}_p} \texttt{OV}_i \\ \texttt{RATE}_p &:= \textstyle\sum_{i \in \mathcal{F}_p} \texttt{MaxRate}_i \end{aligned}$$

From Section 3 it becomes clear that $\texttt{OL}_p$ contains the number of class-p connections that are *overloaded* at the local switch S. Likewise $\texttt{RATE}_p$ contains the maximum traffic rate from connections that are either *underloaded* or *restricted* at switch S, i.e., we can set for an outgoing link l

$$\begin{aligned} \texttt{RATE}_p &\iff \sum_{i \in \mathbf{U}_{lp}} \texttt{PCR}_i + \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share_{kp} \\ \texttt{OL}_p &\iff |\mathbf{O}_{lp}| \end{aligned}$$

Therefore, after obtaining the values for $\texttt{OL}_p$ and $\texttt{RATE}_p$, the switch can calculate the value for $Share_p$ using the expression given at the end of Section 3.1.

At the end of an *class interval* the switch additionally recalculates $Surplus$ using the expression given at the end of Section 3.2.

At the end of an *ABR traffic interval* the switch obtains new values for C^{ABR} using the expression given at the end of Section 3.3.

5 Simulation Experiments

To provide insight into the dynamics of the multi-level bandwidth scheme from Section 4, we present simulation experiments to show the transient behavior during changes of the network load. The simulations were implemented using the REAL (version 4.0) network simulator [6]. The implementation of the source, destination, and switch behavior of the ABR traffic control protocol is based on the Traffic Management Specification Version 4.0 from November 1995 [10]. We do not consider the effects of higher-level protocols.

As shown in Figure 2, the simulated network consists of four source end systems *S1 – S4* and four destination end systems *D1 – D4*. There are three ATM switches connected by two links with a capacity of $C_l = 155$ Mb/s each. The scheduling discipline at the switches is assumed to be FIFO, and the buffer capacity is set to 2000 cells. The propagation delay of the links is varied in the range between 20 μs and 1 ms, corresponding to a distance of 6 km to 300 km. The access links of the sources to the ATM switch have a capacity of 155 Mb/s with negligible propagation delay.

Source – Dest. System	Route	Traffic Class	PCR (Mb/s)	Start Time (ms)
S1–D1	(*L1, L2*)	*I*	10	0
			70	5
S2–D2	(*L1, L2*)	*III*	65	25
S3–D3	(*L1*)	*III*	80	75
S4–D4	(*L2*)	*II*	50	150

Table 1: Connection Parameters.

Switch Parameter	Value ($cells$)
Buffer capacity	2000
NI threshold	200
CI threshold	1500

Source Parameter	Value
Nrm	32 cells
AIR	100
RDF	512
TOF	8 ms
Xrm	64

Table 2: Parameters of ABR Traffic Control.

We set the capacity available to ABR traffic to $C_l^{min} = C_l^{\text{ABR}} = 150$ Mb/s. The four ABR connections are from three different traffic classes: class *I*, class *II*, and class *III*. The class guarantees are identical on each link and set to:

class-*I*	class-*II*	class-*III*
$G_I = 30$ Mb/s	$G_{II} = 45$ Mb/s	$G_{III} = 75$ Mb/s

The parameters of the four connections in Figure 2 are summarized in Table 1. Since each end system is the source or destination of at most one connection, we will use the name of the source to identify a connection, e.g., the connection that begins at end system *S1* is referred to as connection *S1*. All connections are initially idle and start to transmit at time specified in Table 1. We assume that the time interval between cell transmissions is constant; also, the transmission rate of a cell is assumed to include the cell header.

The parameters for the ABR traffic control scheme are set to the values shown in Table 2. The table contains parameters for the switches and parameters for the sources.

The values for the minimum cell rate and the initial cell rate of all connections are set to MCR = 10 Mb/s and ICR = 7 Mb/s. All RM cells are sent *"in-band"*, that is, the transmission of RM cells is not discounted to the traffic rate.

Here, we only present simulations that show connection-level and class-level bandwidth control. We set the length of both the *share intervals* and the *surplus intervals* to 1 ms.

We first demonstrate the effectiveness of connection-level and class-level bandwidth control. By setting the propagation delays of the ATM links to small values, that is, 20 μs or approximately 6 km. The simulation results are summarized in Figure 3 which depicts two graphs that show the bandwidth (in Mb/s) utilized by each connection on the two links *L1* and *L2* . Each data point in the graph corresponds to the amount of data that is transmitted during a share update interval of 1 ms. The experimental results have been verified to match the expected values from Section 3. We now discuss the outcome of the simulation in detail.

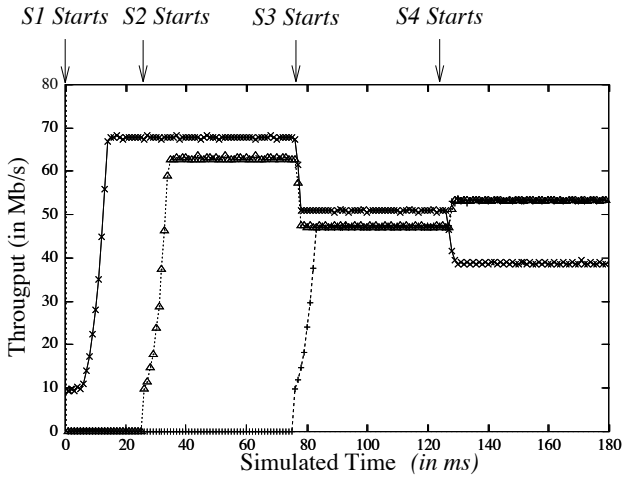
- All connections are initially idle. At $t = 0$, connection *S1* from class-*I* becomes active with a peak cell rate of PCR = 10 Mb/s. This value is increased at $t = 5$ to PCR = 70 Mb/s. Connection *S1* exceeds the its bandwidth guarantee of class *I* but can 'borrow' extra bandwidth from the other classes.

- At $t = 25$, class-*III* connection *S2* begins transmission with PCR = 65 Mb/s. Since sufficient bandwidth is guaranteed to class *III*, *S2* can transmit at its peak cell rate.

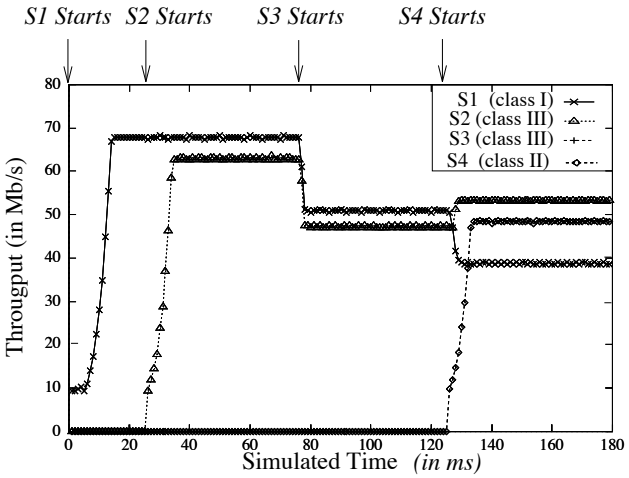
- At $t = 75$, connection *S3* from class *III* starts to transmit on *L1* with PCR = 80 Mb/s. Then, traffic classes *I* and *III* require all of their respective bandwidth guarantees on *L1*. Since there is no class-*II* traffic on *L1*, there is a surplus bandwidth of 45 Mb/s ($= G_{II}$) on this link. Therefore, class-level bandwidth control takes effect and evenly divides the surplus bandwidth between classes *I* and *III*. Since *S1* is the only connection in class *I*, it obtains its class guarantee and one half of the surplus, resulting in a throughput of $30 + 22.5 = 52.5$ Mb/s. For traffic class *III* the available bandwidth on *L1* after class-level bandwidth control is $75 + 22.5 = 97.5$ Mb/s. Since there are two class-*III* connections on *L1*, connection-level control splits the bandwidth between *S2* and *S3*. As a result, both connections obtain a throughput of 48.75 Mb/s.

- At time $t = 125$, connection *S4*, from class *II* becomes active on *L2* with a peak rate of PCR = 50 Mb/s, and requires its entire bandwidth guarantee of $G_{II} = 45$ Mb/s at *Link 2*. The reduced 'surplus' on *Link 2* decreases the throughput available to *S1*, and causes a shift of connections *S1*'s bottleneck from *L1* to *L2*. This in turn, makes bandwidth available for the class-*III* connections on *L1*, yielding a throughput increase for *S2* and *S3*.

Finally we investigate the impact of the propagation delay on the effectiveness of our bandwidth control. Figure 4 depicts the simulation results if the propagation delay is set to 1000 μs per link, corresponding to a length of about 300 km, yielding a maximum round-trip delay of 4 ms. Note that the maximum round-trip delay is larger than the length of the update interval of 1 ms. We see in Figure 4 that at times $t = 75$ and at $t = 125$ the network requires a considerable time to converge to stable throughput values. Nonetheless, it can be seen in Figure 4 that the protocol stabilizes at the correct values.

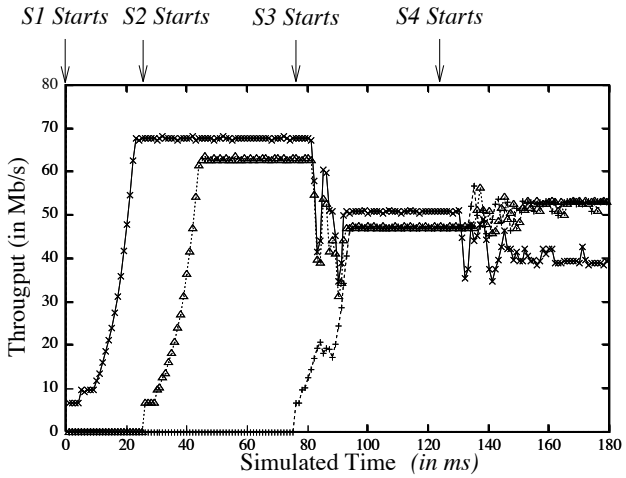


Throughput on Link 1

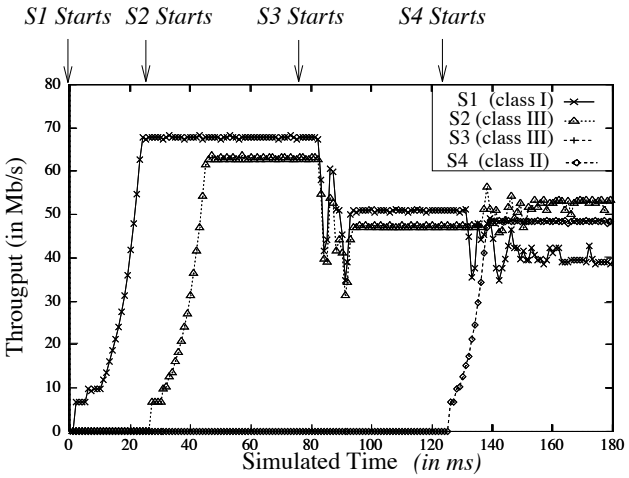


Throughput on Link 2

Figure 3: Multi-Level Bandwidth Control (propagation delay is 20 μs per link).



Throughput on Link 1



Throughput on Link 2

Figure 4: Multi-Level Bandwidth Control (propagation delay is 1000 μs per link).

References

[1] D. Bertsekas and R. Gallager. *Data Networks, 2nd Ed.* Prentice Hall, 1992.

[2] F. Bonomi and W. Fendick. The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service. *IEEE Network*, 9(2):25–39, March/April 1995.

[3] D.-M. Chiu and R. Jain. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. *Computer Networks and ISDN Systems*, 17:1–14, 1989.

[4] J. M. Jaffe. Bottleneck Flow Control. *IEEE Transactions on Communications*, 29(7):954–962, July 1981.

[5] J. Jain. Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey. *to appear in: Computer Networks and ISDN Systems*, February 1995. Invited Submission.

[6] S. Keshav. REAL: A Network Simulator. Technical Report 88/472, Computer Science Department, University of California, Berkeley, December 1988.

[7] H. T. Kung and R. Morris. Credit-Based Flow Control for ATM Networks. *IEEE Network*, 9(2):40–48, March/April 1995.

[8] B. A. Makrucki. On the Performance of Submitting Excess Traffic to ATM Networks. In *Proc. IEEE Globecom'91*, December 1991.

[9] P. Newman. Backward Explicit Congestion Notification for ATM Local Area Networks. In *Proc. IEEE Globecom'93*, pages 719–723, November 1993.

[10] S. S. Sathaye. Draft ATM Forum Traffic Management Specification Version 4.0. ATM Forum/95-0013R8, November 1995.

[11] N. Yin and M. G. Hluchyj. On Closed-Loop Rate Control for ATM Cell Relay Networks. In *Proc. IEEE INFOCOM'94*, pages 99–108, June 1994.