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

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

† Department of Computer Science
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 currently being standardized by the ATM Forum supports data traffic without firm quality-of-service demands. The rate at which an ABR traffic source can transmit is, among others, dynamically determined by a feedback mechanism which calculates an upper bound for the transmission rate, the *explicit rate*, of a connection. A drawback of existing schemes for calculating the explicit rate of ABR connections is that they do not distinguish among different types of ABR connections, and instead, try to give equal treatment to all ABR connections. In this study it is shown how to give differential treatment to different types of ABR connections. A multi-level explicit rate scheme for ABR traffic is presented that controls ABR traffic in an ATM network at three levels. At the topmost level, bandwidth is dynamically regulated between CBR, VBR, and ABR traffic sources based on the utilization and the priority given to the services. At the next level, bandwidth is controlled between different classes of ABR traffic, such as video traffic classes, bulk data traffic classes, and so on. At the lowest level, bandwidth is distributed among connections belonging to the same ABR traffic class. Modifications to the ABR traffic control protocol are presented that achieve the multi-level control scheme in a general network. The effectiveness of the protocol is demonstrated in simulation experiments.

*Key Words: ATM Networks, Available Bit Rate (ABR), Traffic Control, Bandwidth Allocation, Congestion Control, Fairness.*

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

# 1 Introduction

The *Available-Bit-Rate* (ABR) service class currently being defined by the ATM Forum [22] 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. The ABR service is a lower-quality service as compared to the *Constant-Bit-Rate* (CBR) and *Variable-Bit-Rate* (VBR) services in that it does not provide guarantees on cell delays or cell delay variations. On the other hand, the ABR service is superior to the *Unspecified-Bit-Rate* (UBR) service which is intended for non-real time applications without a need for tightly constrained quality-of-service.

Upon establishment of an ABR connection, the user submits to the ATM network a specification of its maximum traffic demand (*peak cell rate*), and its minimum throughput requirements (*minimum cell rate*). The network performs an admission test and verifies that sufficient bandwidth is available to support the minimum cell rate of the connection. If so, the network accepts the new connection. The quality-of-service guarantees of an ABR connection are limited to guarantees on the throughput at the minimum cell rate and a bound on the cell loss rate. 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.

The main challenge for defining the ABR service class is to devise effective congestion control methods that adjust the permissible traffic rate of ABR connections to the currently existing load from CBR and VBR connections, as well as from other ABR connections. 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. The credit-based *Flow Controlled Virtual Circuit* mechanism achieves congestion control of ABR traffic with a modified window flow control mechanism on a per-link basis [13, 14]. For each hop on the route of the connection, the upstream switch maintains a *credit balance* which is decremented every time a cell is transmitted. A switch is allowed to transmit a cell only if the credit balance is positive. The downstream switch periodically sends control cells to the upstream switch to increase the credit balance of the upstream switch. In contrast, the closed-loop rate-based approach [16, 21, 25] controls congestion on an end-to-end basis. Here, the intermediate switches give feedback information to the traffic source which, in turn, adjusts its traffic rate. In some proposals, the feedback information given to the source only consists of a single bit that indicates the presence of congestion in the network (*binary feedback*) [16, 26]; other proposals suggest to tell ABR traffic sources directly the maximum allowed transmission rate (*explicit rate schemes*) [21].

Both the credit-based and the rate-based approaches have inherent merits and drawbacks [20]. The main advantage of the rate-based over the credit-based approach is that it does not require a complicated switch architecture. The ATM Forum decided in September 1994 to adopt the rate-based approach in preference over the credit-based approach. Since then, many details of the algorithms to be used have been added with finalization expected in August 1995.

In this study we consider ATM networks with a rate-based congestion control scheme and address the problem of determining the explicit rate for ABR connections in an ATM network with rate-based traffic control. An explicit rate mechanism implements a specified policy which describes the maximum bandwidth allocation available for individual ABR connections. The policy should make bandwidth allocation independent of the location of the source and the route of a connection. Since early on, the ATM Forum has adopted the so-called *max-min fairness* or *bottleneck flow control* [3, 6, 9] scheme. This scheme gives each ABR connection the same maximum throughput allocation on each link in the network. Since the end-to-end throughput of a connection is bounded by the link on the connection's route with the smallest allocation (the *bottleneck*), all connections with the same bottleneck have identical throughput constraints.

A whole set of feasible implementations of max-min fairness has been presented to the ATM Forum [5, 21, 11]. The major drawback of max-min fairness is that it tries to give equal treatment to all ABR connections 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. We will show that this is feasible without significantly increasing the complexity of the ABR traffic control protocol. We propose an explicit rate scheme for ABR traffic that operates at multiple levels. 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 such the minimum or peak cell rate, 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, e.g., [24], 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. The three levels of control are illustrated in Figure 1 for a single link.

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

• **Class-Level Bandwidth Control** is the second level of control. It distributes the available ABR bandwidth on a link among multiple traffic classes. For instance, in Figure 1 we distinguish

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

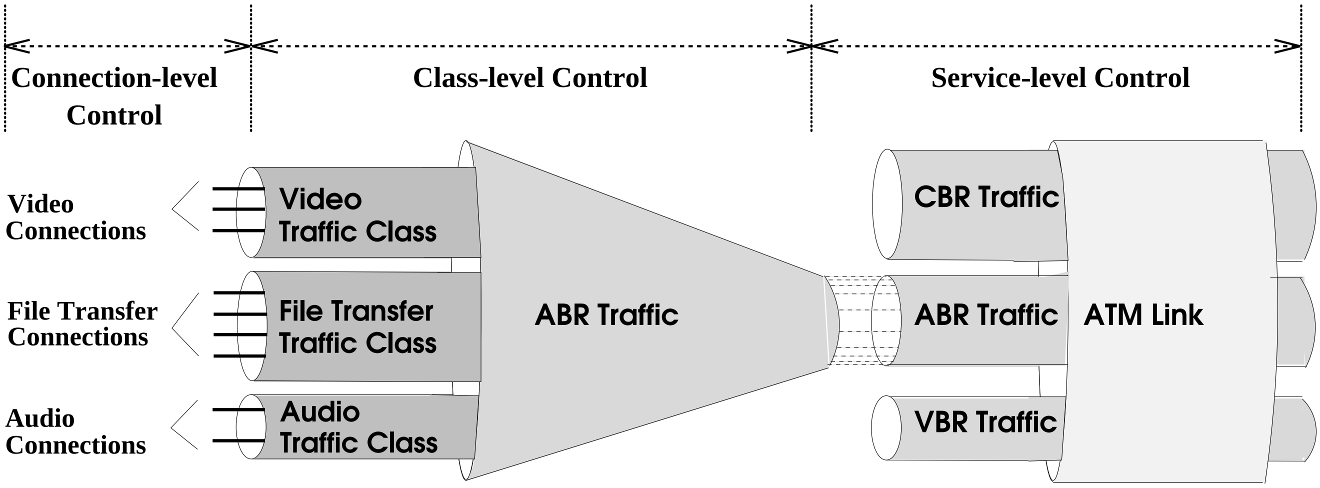



Figure 1: Multi-Level Bandwidth Control at an ATM Link.

three different traffic classes; connections that perform file transfers, connections with video traffic, and connections with audio traffic.

For each link, a *class guarantee* specifies the minimum bandwidth guarantee for this class. The *class capacity* specifies the maximum bandwidth currently available to the connections in this class. 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. In Figure 1, connection-level bandwidth control would determine the bandwidth available to individual video, file transfer, or audio connections.

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

In this section we review the main aspects of the rate-based traffic management approach for ABR traffic as specified by the Traffic Management Group of the ATM Forum. Since our proposed explicit rate scheme extends the work of the ATM Forum, it is necessary to review the ATM

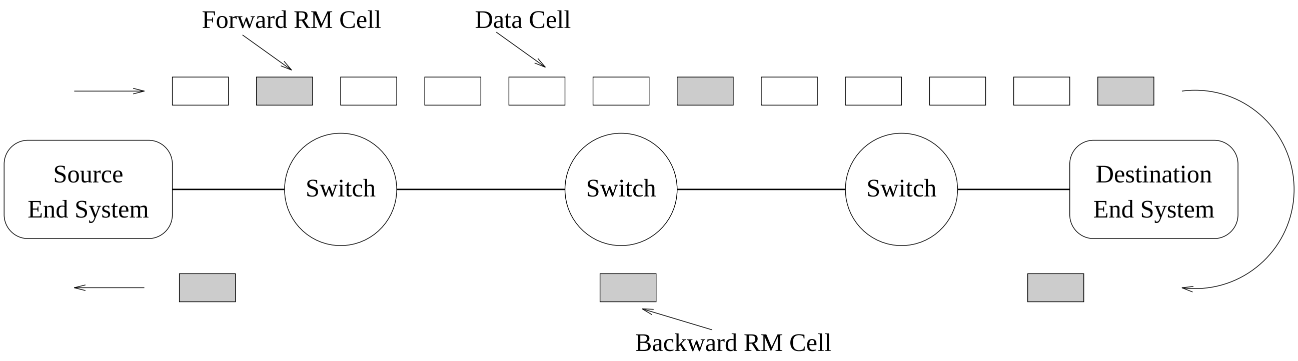



Figure 2: Closed-Loop Traffic Control.

Forum's traffic management approach to a certain level of detail.

Since the first proposals for closed-loop rate-based traffic control [7], many additions to the protocol have significantly enhanced its functionality. The current draft version of the Traffic Management document [22], expected to be finalized in August 1995, uses more than a dozen parameters for the control algorithm (see Appendix A for a complete list). The following review is by no means comprehensive and we refer to the current draft specification for details [22]. Excellent discussions of the rate-based approach for ABR traffic can be found in [4, 10].

## 2.1 Closed-Loop Traffic Control

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, reflecting the state of the ATM Forum discussions since April 1994 [7], are shown in Figure 2. 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)* [15, 25]: 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 Congestion Indication (`CI`) flag in a backward RM cell.

2. *Backward Explicit Congestion Notification (BECN)* [2, 17, 18]: With BECN, congested switches

can return congestion indication to the sources directly (without the cooperation of the destination) by generating backward RM cells with the `CI` 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, 8]: If a source generates an RM cell, it initializes an *Explicit Rate* (`ER`) field with the desired transmission rate. This `ER` field is reduced by the intermediate switches on either or both of the forward and return trip of the RM cell. 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 more complex than presented here [22]. 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. Note that the feedback schemes described above do not reveal how the source should adapt its traffic to the feedback information. Nonetheless, the way in which the source adjusts its traffic rate has a crucial impact on the stability of the feedback control loop. The ATM Forum adopted the *additive increase/multiplicative decrease* scheme [19] where traffic is increased in fixed increments, while traffic decreases are proportional to the traffic rate. Another important issue for the effectiveness of the feedback mechanism deals with the frequency of feedback information. The solution eventually adopted by the ATM Forum gives sources positive feedback information at a rate proportional to their respective traffic rates.

## 2.2 The ATM Forum Specification

Next we discuss how the closed-loop rate control is implemented according to the specification of the ATM Forum [22]. We first discuss the RM cell format, and then present the operations performed by sources, destinations, and switches. In Appendix A we give an overview chart of the acronyms used in the context of ABR traffic control. Note that the completion of the traffic management specification, scheduled for August 1995, is pending and may still undergo changes.

*(a) RM Cell Format.*

The RM cell contains fields for binary feedback (both BECN and FECN) as well as explicit rate feedback. Binary feedback is given to the source via the *Congestion Indication* (`CI`) *flag* and the *No Increase* (`NI`) *flag*. A source cannot increase its traffic rate if any of the flags is set, and must decrease its rate if the `CI` flag is set. Recall that the EFCI field in a data cell is also used for binary feedback.

For explicit rate setting, the RM cell provides three fields which are initialized by the source.

- The *Explicit Rate* (ER) *field* is initialized by the source to its peak cell rate. The ER field can be reduced by the switches and the destination end system. The value of the ER field that is returned to the source, is a throughput bound for the traffic source which cannot be exceeded by the source.
- The *Minimum Cell Rate* (MCR) *field* is filled by the source with the minimum guaranteed traffic rate that was negotiated during connection establishment. The field is not modified in the network.
- The *Current Cell Rate* (CCR) *field* contains the source's current notion of the its maximum transmission rate. Like the MCR field, the CCR field is set by the source and left unchanged in the network.

The switches use the ER, MCR, and CCR fields to calculate the throughput bounds for sources, i.e., the explicit rates that are returned to the source in the ER field. Finally, the RM cell contains a *queue length (*QL*) field* which is updated by the switches to the maximum number of cells currently queued at any switch on the route of a connection.

*(b) Behavior of Source End Systems.*

During connection setup a new source negotiates with the ATM network the maximum traffic rate PCR (*Peak Cell Rate*), the minimum traffic rate MCR (*Minimum Cell Rate*) guaranteed by the network, and an initial transmission rate ICR (*Initial Cell Rate*). The following parameters are needed at the source to calculate the explicit rate: Nrm determines the number of data cells between two RM cells, and the *Additive Increase Rate* (AIR) and *Rate Decrease Factor* (RDF) are used to implement an additive increase and multiplicative decrease congestion control scheme. At all times, the *Allowed Cell Rate* (ACR) at a source gives the maximum transmission rate currently supported by the network. (Both ICR and ACR range between MCR and PCR.)

A source generates at least one RM cell after every Nrm cells transmitted. The source initializes the RM cell by setting the ER field to PCR, and the CCR field of the cell to the current value of ACR.

The following updates are performed when a source receives a backward RM cell, based on the content of the CI flag, the NI flag and, the ER field. If CI=1, then ACR is decreased by ACR·Nrm/RDF; ACR is not decreased below MCR. If CI=0 and NI=1, then ACR is not changed. Finally, if CI=0 and NI=0 then the cell can increase the ACR by AIR·Nrm, up to, but not exceeding, the value in the ER field. If a source has not received a single backward RM cell after sending several RM forward cells, ACR is also reduced by ACR·Nrm/RDF.

*(c) Behavior of Destination End Systems*

As indicated in Figure 2, the destination end system returns the received RM cells to the sender. By setting a flag, the destination changes an RM cells from a 'forward' to a 'backward' direction. If a destination receives an RM cell immediately after a data cell with the EFCI flag set, then the destination sets CI=1 in the backward cell. If the destination itself experiences congestion, it may reduce the rate in the ER field and/or set CI=1 and NI=1. For BECN, the destination is allowed to generate backward RM cells.

*(d) Behavior of Switches*

The traffic management specification distinguishes between two types of switches. EFCI-based switches only participate in binary feedback control and do not participate in explicit rate setting. EFCI-based switches, if congested, set the EFCI bit in data cells, and/or set CI=1 in backward RM cells. ER-based switches take part in the explicit rate setting by reducing (never increasing!) the ER field of both or either of the forward and backward RM cells. Several algorithms have been proposed that determine how a switch sets the ER field. Most of the proposals implement one of the fairness schemes that are discussed in detail in the next section [1, 5, 21]. Similar to the destination, a congested switch can generate a backward RM cell with the CI bit set.

# 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 in 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 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. Recall from Section 2 that the maximum allowed transmission rate of a connection may be smaller than the explicit rate due to additional feedback congestion schemes, such as EFCN or BECN. In accordance to the ABR traffic control scheme, we assume that explicit rate setting and binary feedback schemes operate independently. The maximum throughput, the so-called *allowed cell rate*, is the smaller of the rates permitted by the explicit rate and the binary scheme.

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 ($1 \leq p \leq P$). |
| $\mathcal{R}_i = (l_{i_1}, l_{i_2}, \ldots, l_{i_K})$ | Route of ABR connection $i$; $l_{i_k} \in \mathcal{L}$ is the $k$th link on the route of connection $i$. |
| $\Delta_{lp} = \{ i \mid l \in \mathcal{R}_i,\ i \in \mathcal{F}_p \}$ | Set of connections in class $p$ with link $l$ on their route. |

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 ability of the ATM network to support $\mathtt{MCR}_i$ is ensured through appropriate connection admission control functions that are tested during the connection establishment phase.

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:

$$\mathtt{MCR}_i \leq \mathtt{ER}_i \leq \mathtt{PCR}_i \tag{1}$$

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:

$$\mathtt{ER}_i \leq Share_{lp} \qquad \text{for all } i \in \Delta_{lp} \tag{2}$$

- **Class-level bandwidth control** consists of the enforcement of throughput bounds for the aggregate bandwidth used by connections from a class at a link. We use $C_{lp}^{\mathrm{ABR}}$ 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:

$$\sum_{i \in \Delta_{lp}} \mathtt{ER}_i \leq C_{lp}^{\mathrm{ABR}} \qquad \text{for all } l \in \mathcal{L} \tag{3}$$

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

$$\sum_{p=1}^{P} \sum_{i \in \Delta_{lp}} \mathtt{ER}_i \leq C_l^{\mathrm{ABR}} \qquad \text{for all } l \in \mathcal{L} \tag{4}$$

In the following subsections we describe the multi-level bandwidth control scheme in detail. In Subsection 3.1 we describe a method for selecting throughput bounds for connection-level bandwidth control. In Subsection 3.2 we discuss class-level bandwidth control and present a scheme that varies the bandwidth available to ABR traffic classes according to the bandwidth needs in each class. In Subsection 3.3 we present a service-level control scheme which adapts the total capacity available to ABR traffic according to the demands of CBR and VBR connections.

### 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., $C_{lp}^{\mathrm{ABR}} \equiv \mathit{const.}$ 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 $\mathtt{MCR}_i$ can be supported on all links on the route of the connection. Hence, for a connection $j$, the following condition must be tested at the time of connection establishment:

$$\sum_{i \in \Delta_{lp}} \mathtt{MCR}_i \leq C_{lp}^{\mathrm{ABR}} \qquad \text{for all } l \in \mathcal{R}_j \tag{5}$$

Connection-level bandwidth control distributes the class capacity $C_{lp}^{\mathrm{ABR}}$ 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 $\mathtt{PCR}_i$ (see equation (1)), We obtain the following throughput bound for connection $i$:

$$\mathtt{ER}_i = \min\Big(\mathtt{PCR}_i, Share_{l_i^* p}\Big) \tag{6}$$

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, 6, 9, 27].

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 [1, 5, 21].

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:

$$\mathbf{U}_{lp} = \Big\{ i \in \Delta_{lp} \ \big| \ Share_{lp} \geq \mathtt{PCR}_i \ , \ \ i \notin \bigcup_{k \in \mathcal{L}} \mathbf{R}_{lp}(k) \Big\} \tag{7}$$

$$\mathbf{O}_{lp} = \Big\{ i \in \Delta_{lp} \ \big| \ l = l_i^* \ , \ \ Share_{lp} < \mathtt{PCR}_i \Big\} \tag{8}$$

$$\mathbf{R}_{lp}(k) = \Big\{ i \in \Delta_{lp} \ \big| \ k = l_i^* \ , \ \ Share_{kp} < \mathtt{PCR}_i \Big\} \qquad \text{for } k \neq l \tag{9}$$

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

$$\begin{aligned} C_{lp}^{\mathrm{ABR}} &= \sum_{i \in \Delta_{lp}} \mathtt{ER}_i \\ &= \sum_{i \in \Delta_{lp}} \min \Big( \mathtt{PCR}_i, Share^*_{l_i^* p} \Big) \\ &= |\mathbf{O}_{lp}| \cdot Share^*_{lp} + \sum_{i \in \mathbf{U}_{lp}} \mathtt{PCR}_i + \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share^*_{kp} \end{aligned} \tag{10}$$

For links without overloaded class-$p$ connections ($\mathbf{O}_{lp} = \emptyset$), we set $Share^*_{lp} = C_{lp}^{\mathrm{ABR}}$. Then we obtain with (10) the following values for the maximal shares:

$$Share^*_{lp} = \begin{cases} C_{lp}^{\mathrm{ABR}} & \text{if } \mathbf{O}_{lp} = \emptyset \\ \dfrac{C_{lp}^{\mathrm{ABR}} - \sum\limits_{i \in \mathbf{U}_{lp}} \mathtt{PCR}_i - \sum\limits_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share^*_{kp}}{|\mathbf{O}_{lp}|} & \text{otherwise} \end{cases} \tag{11}$$

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. It is noteworthy that for the calculation of maximal shares, no knowledge is required on individual connections. Precisely, three pieces of information must be at hand to determine the maximal share for class $p$ on a link: (1) the maximal throughput of class-$p$ connections that are overloaded on some other link, (2) the peak cell rate of underloaded class-$p$ connections, and (3) the number of overloaded class-$p$ connections on the link.

## 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 $C_{lp}^{\text{ABR}}$ 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 $C_{lp}^{\text{ABR}}$ to the actual traffic demand.

In the scheme proposed here, the class capacity $C_{lp}^{\text{ABR}}$ 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 $\sum_{p=1}^{P} Guar_{lp} = C_l^{\text{ABR}}$, 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 $Guar_{lq} - \sum_{i\in\Delta_{lq}} \texttt{ER}_i > 0$ 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 $C_{lp}^{\text{ABR}}$ 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 $C_{lp}^{\text{ABR}}$ at link $l$ for class $p$ is set to:

$$C_{lp}^{\text{ABR}} = \min\left(\sum_{i\in\Delta_{lp}} \texttt{ER}_i, Guar_{lp} + Surplus_l\right) \tag{12}$$

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 $\texttt{MCR}_i$ at all times:

$$\sum_{i\in\Delta_{lp}} \texttt{MCR}_i \leq G_{lp} \qquad \text{for all } l \in \mathcal{R}_j \tag{13}$$

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:

$$\begin{aligned} C_l^{\text{ABR}} &= \sum_{\mathbf{O}_{lq} \neq \emptyset} \Big( Guar_{lq} + Surplus_l \Big) + \sum_{\mathbf{O}_{lq} = \emptyset} \left( \sum_{i \in \mathbf{U}_{lq}} \texttt{ER}_i + \sum_{k \in \mathcal{L}} \sum_{i \in \mathbf{R}_{lq}(k)} \texttt{ER}_i \right) && (14) \\ &= \sum_{\mathbf{O}_{lq} \neq \emptyset} \Big( Guar_{lq} + Surplus_l \Big) + \sum_{\mathbf{O}_{lq} = \emptyset} \left( \sum_{i \in \mathbf{U}_{lq}} \texttt{PCR}_i + \sum_{k \in \mathcal{L}} |\mathbf{R}_{lq}(k)| \cdot Share^*_{kq} \right) && (15) \end{aligned}$$

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, with (14) we obtain that a bandwidth control scheme with maximal surplus satisfies:

$$Surplus_l^* = \begin{cases} C_l^{\text{ABR}} & \text{if } \bigcup_{q=1}^{P} \mathbf{O}_{lq} = \emptyset \\ \dfrac{C_l^{\text{ABR}} - \sum_{\mathbf{O}_{lq} \neq \emptyset} Guar_{lq} - \sum_{\mathbf{O}_{lq} = \emptyset} \left( \sum_{i \in \mathbf{U}_{lp}} \texttt{PCR}_i + \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share^*_{kp} \right)}{|\{q \mid 1 \leq q \leq P,\ \mathbf{O}_{lq} \neq \emptyset\}|} & \text{otherwise} \end{cases} \tag{16}$$

In a bandwidth control scheme with maximal shares $Share^*_{lp}$ and maximal surplus values $Surplus^*_l$ both conditions in (11) and (16) are simultaneously satisfied.

### 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 $\Gamma_l^{\text{CBR}}$ and $\Gamma_l^{\text{VBR}}$, 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:

$$C_l^{\text{ABR}} = \max\Big( C_l^{min}, \quad C_l - \Gamma_l^{\text{CBR}} - \Gamma_l^{\text{VBR}} \Big) \tag{17}$$

Note that changing $C_l^{\text{ABR}}$ 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 $j$:

$$\sum_{p=1}^{P} \sum_{i \in \Delta_{lp}} \mathtt{MCR}_i \leq C_l^{min} \qquad \text{for all } l \in \mathcal{R}_j \tag{18}$$

In summary, we have presented a set of objectives for controlling the bandwidth available to ABR traffic in an ATM network. The objectives define a multi-level bandwidth control scheme. Maximal shares control the bandwidth allocated to ABR connections in the same class, adaptive class capacities control the bandwidth available to the traffic classes, and an adaptive ABR capacity regulates the total bandwidth available to ABR traffic. In the next section we show how to modify the traffic control protocol developed by the ATM Forum [22] to implement multi-level bandwidth control.

# 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 [22]. The protocol mechanisms described here are mainly modifications or additions to the ABR control protocol reviewed in Section 2. Therefore, we focus in our discussion on the differences of our scheme to the Traffic Management Specifications (Version 4.0). We only discuss the calculation of the explicit rates. We assume the existence of a binary feedback scheme to detect and react to short-term congestion, as described in Section 2.

## 4.1 Modifications to the RM Cell Format

We require only a minor modification to the RM cell format described in [22]. Overall, our scheme requires fewer parameters than the current ATM Forum proposal. All bit flags, e.g., CI and NI, described in Section 2 of the RM cell are left unchanged. The use of the ER field is similar to, but not identical with, the use of the ER field in [22]. The difference will become clear when we describe the switch behavior in Subsection 4.4. We do not require the `CCR` and `MCR` fields. The only addition to the RM cell format described in [22] 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 Behavior

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

The state of an ABR connection (see Section 3), can be determined from the last backward RM cell that has returned to the source. If $\texttt{BNK} = nil$, then the connection is '*underloaded*' (in this case, the ER field will be set to $\texttt{ER} = \texttt{PCR}$). On the other hand, if the content of the last backward RM cell was $\texttt{BNK} = S$ and $\texttt{ER} = rate$ (with $\texttt{PCR} > rate$), 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. As in the ATM Forum draft [22], the maximum transmission rate can be lower than the explicit rate due to binary congestion information. Note that a source need not be aware of its state.

### 4.3 Destination Behavior

The destination plays no particular role in the calculation of the explicit rate. As described in Section 2, the destination turns around forward RM cells and returns them to the source as backward RM cells. If the destination can experience congestion, it should perform the same operations as outlined in the Subsections 4.4 and 4.5.

### 4.4 Switch Behavior

In conformance with the ATM Forum, [22] ATM switches are classified as either *EFCI*-based or *ER*-based. Since *EFCI*-based switches are not involved in the calculation of explicit rates (see Section 2) we only describe the behavior of *ER*-based switches.

An *ER*-based 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 (see Section 3): $\texttt{Share}_p$ contains the maximum cell rate at which connections from class $p$ can transmit at this switch, the bandwidth guarantee $\texttt{Guar}_p$ and the Surplus are used to calculate the total bandwidth available for traffic from class $p$ at this switch (see Section 3). Finally, $\texttt{Cap}^{\text{ABR}}$ 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 $\texttt{Share}_p$. If $\texttt{Share}_p \geq \texttt{ER}$, then the switch does not perform any operations. On the other

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

hand, if $\texttt{Share}_p < \texttt{ER}$, 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:

$$\begin{aligned} \texttt{ER} &= \texttt{Share}_p \\ \texttt{BNK} &= \texttt{T} \end{aligned} \tag{19}$$

In other words, the switch sets the explicit cell rate to the maximum rate at this switch, and identifies itself as the bottleneck of the connection. Clearly, if a backward RM cell returns to the source, the BNK field contains the switch with the smallest value for $\texttt{Share}_p$ on the route of the connection, i.e., the bottleneck.

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:

$$\begin{aligned} \texttt{MaxRate}_i &= \texttt{ER} \\ \texttt{OV}_i &= 0 \end{aligned} \tag{20}$$

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:

$$\begin{aligned} \texttt{MaxRate}_i &= 0 \\ \texttt{OV}_i &= 1 \end{aligned} \; 21 \tag{21}$$

### 4.5 Operations at Update Intervals

Periodically, a switch uses the values on the ABR connections 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, i.e.,

$$\begin{aligned} \textit{ABR capacity interval} &= M \cdot \textit{surplus interval} \\ &= M \cdot N \cdot \textit{share interval} \end{aligned} \tag{22}$$

for some integers $M, N \geq 1$. At the end of each *share interval*, a switch, say switch S, recalculates for each of its outgoing links the maximum allowable cell transmission rate $\texttt{Share}_p$ for connections from traffic class $p$. At the end of a *surplus interval* it recalculates in addition the surplus value Surplus. At the end of an *ABR capacity interval* the switch also recalculates the capacity available for ABR traffic, denoted by $\texttt{Cap}^{\text{ABR}}$.

At the end of each share interval, switch $\mathtt{S}$ calculates the following sums for each traffic class $p$:

$$\begin{array}{lll} \mathtt{OL}_p & := & \sum_{i \in \mathcal{F}_p} \mathtt{OV}_i \\ \mathtt{RATE}_p & := & \sum_{i \in \mathcal{F}_p} \mathtt{MaxRate}_i \end{array} \tag{23}$$

From equation (21), it should become clear that $\mathtt{OL}_p$ contains the number of class-$p$ connections that are *overloaded* at the local switch $\mathtt{S}$. Likewise, according to equation (20), $\mathtt{RATE}_p$ contains the maximum traffic rate from connections that are either *underloaded* or *restricted* at switch $\mathtt{S}$. After obtaining the values for $\mathtt{OL}_p$ and $\mathtt{RATE}_p$, the following calculations are performed.

1. At the end of a *share interval* the switch calculates the new value for $\mathtt{Share}_p$ for all $1 \leq p \leq P$ as follows:

$$\mathtt{Share}_p = \begin{cases} \mathtt{Cap}^{\mathrm{ABR}} & \text{if } \mathtt{OL}_p = 0 \\ \frac{\mathtt{Guar}_p + \mathtt{Surplus} - \mathtt{RATE}_p}{\mathtt{OL}_p} & \text{otherwise} \end{cases} \tag{24}$$

2. At the end of an *class interval* the switch recalculates $\mathtt{Surplus}$ as follows.

$$\mathtt{Surplus} = \begin{cases} \mathtt{Cap}^{\mathrm{ABR}} & \text{if } \mathtt{OL}_p = 0 \text{ for all } p \\ \dfrac{\mathtt{Cap}^{\mathrm{ABR}} - \displaystyle\sum_{\mathtt{OL}_q > 0} \mathtt{Guar}_q - \sum_{\mathtt{OL}_q = 0} \mathtt{RATE}_q}{\displaystyle\sum_{\mathtt{OL}_q > 0} \mathtt{Guar}_q} & \text{otherwise} \end{cases} \tag{25}$$

3. At the end of an *ABR traffic interval* the switch obtains new values for $\mathtt{Cap}^{\mathrm{ABR}}$ similar to equation (17). Assuming that $\mathtt{Cap}$ is the total capacity of an outgoing link, $\mathtt{Cap}^{\mathrm{min}}$ is the minimum capacity available to ABR traffic, and assuming that the cell rates currently allocated to CBR and VBR traffic are given by $\mathtt{RATE}^{\mathrm{CBR}}$ and $\mathtt{RATE}^{\mathrm{VBR}}$, respectively, the switch calculates:

$$\mathtt{Cap}^{\mathrm{ABR}} = \max(\, \mathtt{Cap}^{\mathrm{min}},\ \ \mathtt{Cap} - \mathtt{RATE}^{\mathrm{CBR}} - \mathtt{RATE}^{\mathrm{VBR}}) \tag{26}$$

Note that the above equations (24)–(26) are completely based on equations (11), (16), and (17), and the following relationships hold for an outgoing link $l$ at a switch:

$$\begin{array}{llll} \mathtt{Cap}^{\mathrm{ABR}} & \equiv C_l^{\mathrm{ABR}} & \mathtt{Share}_p & \equiv Share^*_{lp} \\ \mathtt{Cap}^{\mathrm{min}} & \equiv C_l^{\mathrm{min}} & \mathtt{Surplus}_p & \equiv Surplus^*_l \\ \mathtt{Cap} & \equiv C_l & \mathtt{OL}_p & \equiv |\mathbf{O}_{lp}| \\ \mathtt{Guar}_p & \equiv G_{lp} & \mathtt{RATE}_p & \equiv \sum_{i \in \mathbf{U}_{lp}} \mathrm{PCR}_i + \displaystyle\sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot Share^*_{kp} \end{array} \tag{27}$$

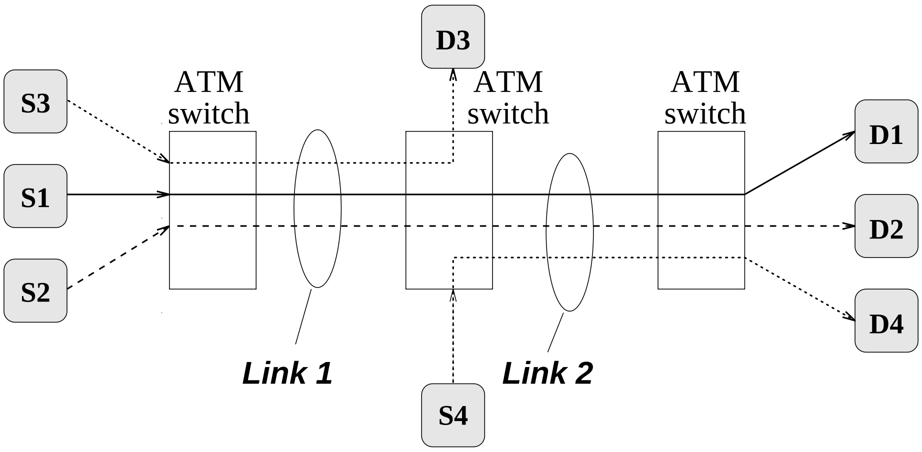



Figure 3: Simulated Network.

# 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 [12].

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 March 1995 [23]. We do not consider the effects of higher-level protocols.

As shown in Figure 3, 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 $\mu$s 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.

We set the capacity available to ABR traffic to $C_l^{min} = C_l^{\mathrm{ABR}} = 150$ 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 3 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.

| Source End System | Destination End System | Route | Traffic Class | Peak Cell Rate | Start Time (*in ms*) |
|---|---|---|---|---|---|
| *S1* | *D1* | (*Link 1, Link 2*) | *I* | 10 Mb/s | $t = 0$ |
| | | | | 70 Mb/s | $t = 5$ |
| *S2* | *D2* | (*Link 1, Link 2*) | *III* | 65 Mb/s | $t = 25$ |
| *S3* | *D3* | (*Link 1*) | *III* | 80 Mb/s | $t = 75$ |
| *S4* | *D4* | (*Link 2*) | *II* | 50 Mb/s | $t = 150$ |

Table 1: Connection Parameters.

| Switch Parameter | Value |
|---|---|
| *Buffer capacity* | 2000 cells |
| *NI threshold* | 200 cells |
| *CI threshold* | 1500 cells |

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

Table 2: Parameters of ABR Traffic Control.

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 minimum the cell rate and the initial cell rate of all connections are set to

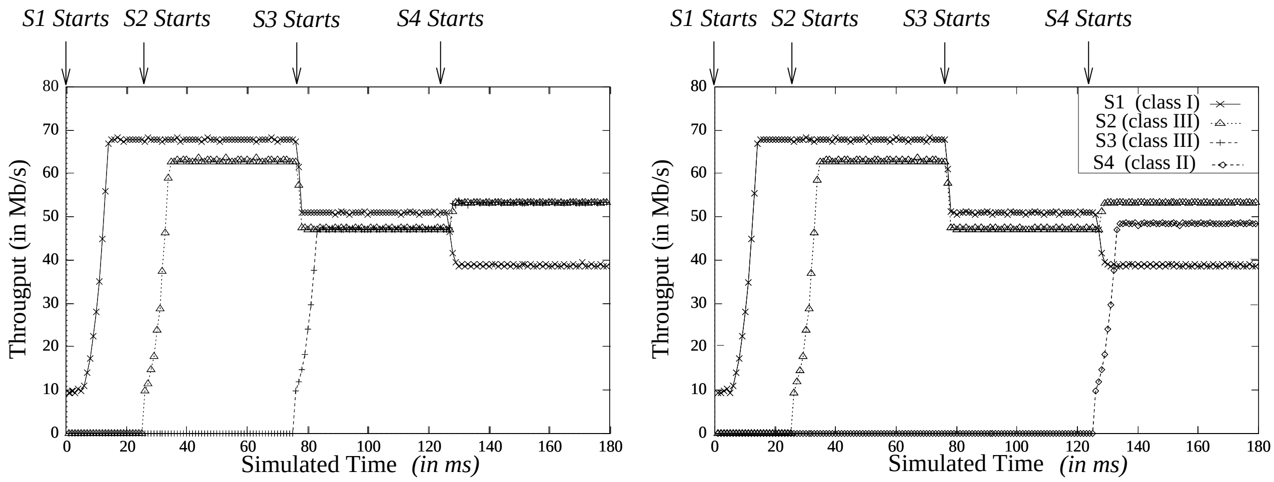
$$\texttt{MCR} = 10 \text{ Mb/s} \qquad \text{and} \qquad \texttt{ICR} = 7 \text{ Mb/s}.$$

All RM cells are sent *"in-band"*, that is, the transmission of RM cells is not discounted to the traffic rate.

While all levels of bandwidth control have been tested, we only present simulation that show connection-level and class-level bandwidth control. In our experiments we set the length of both the *share intervals* and the *surplus intervals* to 1 ms. In this case, both surplus values and share values are changed simultaneously.

## 5.1 Experiment 1

This experiment demonstrates 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 $\mu$s or approximately 6 km, we eliminate the effects of long propagation delays. The simulation results are summarized in Figure 4 which depicts two graphs that show the bandwidth (in Mb/s) utilized by




Throughput on Link 1

Throughput on Link 2

Figure 4: Multi-Level Bandwidth Control (propagation delay is 20 $\mu$s per link).

each connection on the two links *Link 1* and *Link 2* . 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 with the theoretically expected values from Section 3 and found to be accurate. 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. Note that connection *S1* exceeds its bandwidth guarantee of class *I*. However, as none of the other classes utilize their bandwidth guarantee, class *I* can 'borrow' extra bandwidth from the other classes. This allows *S1* to transmit at its offered load.

- 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 *Link 1* with `PCR`= 80 Mb/s. Then, traffic classes *I* and *III* require all of their respective bandwidth guarantees on *Link 1*. Since there is no class-*II* traffic on *Link 1*, 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 *Link 1* after class-level bandwidth control is $75 + 22.5 = 97.5$ Mb/s. Since there are two class-*III* connections on *Link 1*, connection-level control splits the bandwidth between *S2* and *S3*. As a result, both connections obtain a throughput of 48.75 Mb/s. Note in Figure 4 that the theoretically correct values are assumed very quickly.

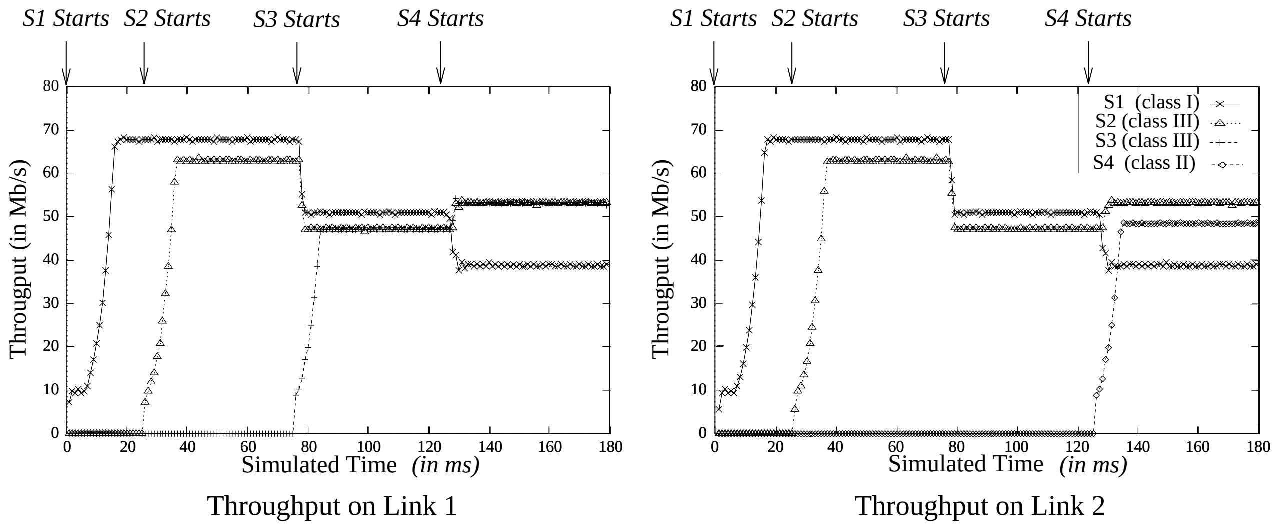



Figure 5: Multi-Level Bandwidth Control (propagation delay is 200 $\mu$s per link).

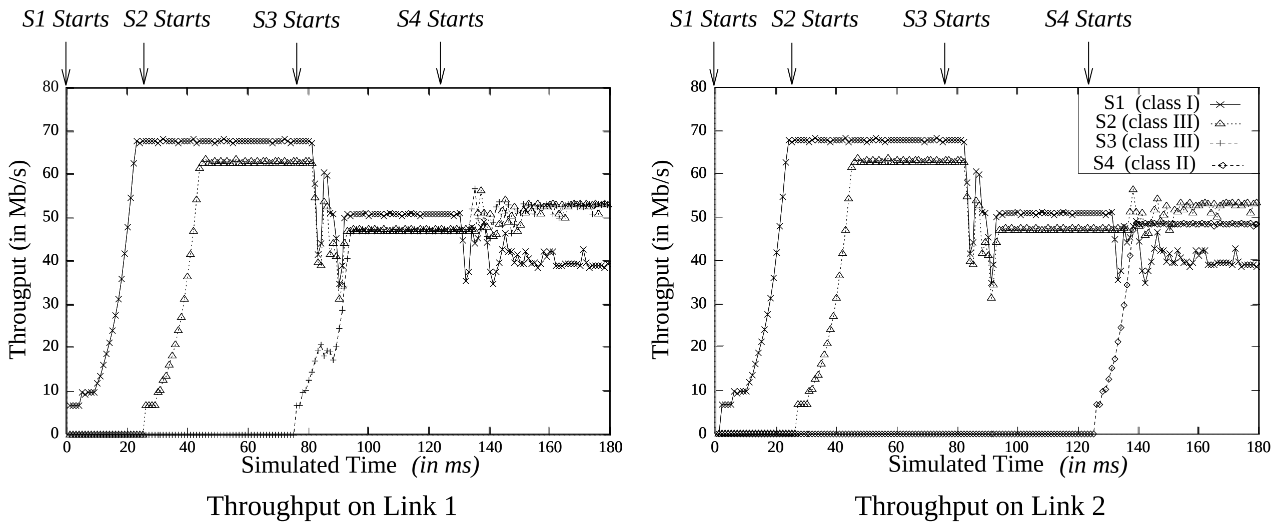



Figure 6: Multi-Level Bandwidth Control (propagation delay is 1000 $\mu$s per link).

- At time $t = 125$, connection *S4*, from class *II* becomes active on *Link 2* 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 *Link 1* to *Link 2*. This in turn, makes bandwidth available for the class-*III* connections on *Link 1*, yielding a throughput increase for *S2* and *S3*.

## 5.2 Experiment 2

In this experiment, we investigate the impact of the propagation delay on the effectiveness of multi-level bandwidth control. In Figure 5 we show the results of the simulations for a propagation delay of 200 $\mu$s per link, or a length of about 60 km. With these values, the maximum round-trip propagation delay is given by 800 $\mu$s, which is still less than the time period of the share update interval of 1 ms. Therefore, the similarity of the simulation results in Figures 4 and 5 is not surprising.

The effects of the propagation delays become more visible when we select the maximum round-trip delay to a value that is larger than the length of the update interval of 1 ms. Figure 6 depicts the simulation results if the propagation delay is set to 1000 $\mu$s per link, corresponding to a length of about 300 km. In this case, the maximum round-trip delay is given by 4 ms. We see in Figure 6 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 6 that the protocol stabilizes at the correct values.

# 6 Conclusions

The rate-based scheme for traffic control of ABR connections proposed by the ATM Forum does not consider that ABR connections may have fundamentally different service requirements. In the current traffic management specification, the explicit rate calculated for ABR connections does not distinguish among different types of ABR connections. In this study we have proposed a solution to this problem by presenting a multi-level bandwidth control scheme for ABR connections in an ATM network. Our proposed scheme uses the notion of *traffic classes* and assignseach ABR connection to exactly one traffic class; connections in the same traffic class receive the same service by the ATM network. Our proposed explicit rate scheme enables the control of bandwidth at three levels. At the highest level, the *service level*, we control the allocation of ABR bandwidth under consideration of the traffic from VBR and CBR connections. At the next level, the *class level*, we control the bandwidth made available to different ABR traffic classes. Each traffic class obtains a bandwidth guarantee at a network link, and the guarantees can be temporarily increased by borrowing bandwidth from traffic classes that do not utilize their guarantee. At the lowest level, the *connection level*, we control the bandwidth available to single ABR connections such that all connections in the same traffic class satisfy specific fairness conditions. We have shown the theoretical underpinnings for the multi-level bandwidth control scheme. We have discussed how to implement multi-level bandwidth control into the traffic management scheme that is currently being finalized by the ATM Forum. Finally, we presented simulation experiments to show that our scheme can indeed achieve bandwidth control simultaneously at multiple levels and satisfy the control objectives at every stage.

# Acknowledgments

Our simulation program is based on a simulator for packet-switching networks written by Debo Sarkar from Hughes Network Systems.

## References


[1] A. W. Barnhart. Use of the Extended PRCA with Various Switch Mechanisms. ATM Forum/94-0898, September 1994.

[2] A. W. Berger, F. Bonomi, and K. W. Fendick. Proposal for Backward Congestion Feedback at the ATM UNI. ATM Forum/93-839R1, September 1993.

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

[4] 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.

[5] A. Charny, D.D. Clark, and R. Jain. Congestion Control with Explicit Rate Indication. ATM Forum/94-0692, July 1994.

[6] 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.

[7] M. G. Hluchyj et.al. Closed-Loop Rate-Based Traffic Management. ATM Forum/94-0211R3, April 1994.

[8] M. G. Hluchyj et.al. Closed-Loop Rate-Based Traffic Management. ATM Forum/94-0438R2, September 1994.

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

[10] J. Jain. Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey. Manuscript, January 1995.

[11] R. Jain, S. Kalyanramana, and R. Viswanathan. The EPRCA+ Scheme. ATM Forum/94-0988, September 1994.

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

[13] H. T. Kung, T. Blackwell, and A. Chapman. Credit-Based Flow Control for ATM Networks: Credit Update Protocol, Adaptive Credit Allocation, and Statistical Multiplexing. In *Proc. ACM SIGCOMM'94*, pages 101–114, September 1994.

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

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

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

[17] P. Newman. Traffic Management for ATM Local Area Networks. *IEEE Communications Magazine*, 8(32):44–50, August 1994.

[18] P. Newman and G. Marshall. BECN Congestion Control. ATM Forum/94-789R1, July 1993.

[19] K. K. Ramakrishnan and R. Jain. A Binary Feedback Scheme for Congestion Avoidance in Computer Networks. *ACM Transactions on Computer Systems*, 8(2):158–181, May 1990.

[20] K. K. Ramakrishnan and P. Newman. Integration of Erate and Credit Schemes for ATM Flow Control. *IEEE Network*, 9(2):49–56, March/April 1995.

[21] L. Roberts. Enhanced PRCA Proportional Rate-Control Algorithm. ATM Forum/94-0735R1, August 1994.

[22] S. S. Sathaye. Draft ATM Forum Traffic Management Specification Version 4.0. ATM Forum/95-0013R4, June 1995.

[23] S. S. Sathaye. Draft ATM Forum Traffic Management Specification Version 4.0. ATM Forum/95-0013R1, March 1995.

[24] K.-Y. Siu and H.-Y. Tzeng. Adaptive Proportional Rate Control for ABR Service in ATM Networks. Technical Report 1102, University of California, Irvine, 1994.

[25] 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.

[26] N. Yin and M. G. Hluchyj. On Closed-Loop Rate Control for ATM Cell Relay Networks. ATM Forum/94-0438R2, September 1994.

[27] M. Zukerman and S. Chan. Fairness in ATM Networks. *Computer Networks and ISDN Systems*, 26(1):109–117, September 1993.

# A ABR Traffic Control Parameters

| Source Parameters | | |
|---|---|---|
| PCR | Peak Cell Rate | Maximum traffic rate; specified by the source and enforced by the network. |
| MCR | Minimum Cell Rate | Minimum acceptable traffic rate; specified by the source and guaranteed by the network. |
| ICR | Initial Cell Rate | Traffic rate for sources after idle periods. |
| ACR | Allowed Cell Rate | Current maximum traffic rate of a source. |
| AIR | Additive Increase Rate | Additive bound for increasing the traffic rate. |
| RDF | Rate Decrease Factor | Rate factor used for multiplicative decrease of traffic rate. |
| Nrm | Number of RM cells | Used to determine the number of data cells between consecutive RM cells. |
| RTT | Round-Trip Time. | Used to adjust the increase in cell rate. |
| TOF | Time Out Factor | Maximum time between RM cells before an enforced rate decrease. |
| Xrm | RM Return Failure | Number of missing RM cells that cause an enforced rate decrease . |

| Bit Flags of RM Cells | | |
|---|---|---|
| DIR | Direction of RM cell | 0 = forward RM cell, 1 = backward RM cell. |
| BN | Backward Explicit Congestion Notification | 0 = switch generated RM cell, 1 = source generated RM cell. |
| CI | Congestion Indication | 0= no congestion, 1 = congestion. |
| BS | Block Start | 1 = new block of cells, 0= otherwise. |
| NI | No Increase | 1 = no additive increase is allowed, 0 = otherwise. |
| EFCI | Explicit Forward Congestion Indication | One-bit field in an ATM cell header. |

| Rate Fields of RM Cells | | |
|---|---|---|
| ER | Explicit Rate | Set by the source to the peak cell rate; reduced by the switches or the destination end system. |
| CCR | Current Cell Rate | Set by the source to its most recent cell transmission rate. |
| MCR | Minimum Cell Rate | Set by source to minimum cell rate. |
| QL | Queue Length | Current maximum queue length for this connection at any of the intermediate switches or the end system. |