



ELSEVIER

Performance Evaluation 31 (1997) 107–131

**PERFORMANCE  
EVALUATION**  
An International  
Journal

## Multi-level rate-based flow control for ABR traffic

Ian F. Akyildiz<sup>a,\*</sup>, Jörg Liebeherr<sup>b,1</sup>, Ioanis Nikolaidis<sup>c</sup>

<sup>a</sup> *Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

<sup>b</sup> *Department of Electrical Engineering, Polytechnic University, Brooklyn, NY 11201, USA*

<sup>c</sup> *Department of Computing Science, University of Alberta, Edmonton, Alta., Canada T6G 2H1*

---

### Abstract

The rate control mechanism for *available-bit-rate* (ABR) traffic as defined by the ATM Forum does not distinguish between different types of ABR connections. However, since ABR traffic may result from a heterogeneous set of applications, there is an apparent need for a flow control scheme that can distinguish between and give differential treatment to different classes of ABR connections. In this study, a multi-level flow control scheme for ABR traffic is proposed that performs flow control of ABR traffic simultaneously at three levels: At the first level, the scheme determines the total available ABR bandwidth at a link. At the second level, the ABR bandwidth is distributed to different ABR traffic classes. At the third level, the scheme determines the bandwidth available to connections in same traffic class. It is shown that the multi-level flow control method completely decouples the explicit rate calculation of distinct traffic classes while achieving a high network utilization. Extensive simulations demonstrate that the multi-level flow control quickly adapts to load changes in the network. © 1997 Elsevier Science B.V.

**Keywords:** ATM network; Available-bit-rate (ABR); Traffic control; Bandwidth allocation; Congestion control; Fairness

---

### 1. Introduction

The *available-bit-rate* (ABR) service class standardized by the ATM forum [21] meets the service requirements of traffic types with only vaguely defined quality-of-service (QoS) requirements. An endsystem that establishes an ABR connection specifies its maximum required bandwidth, the *peak cell rate*, and minimum usable bandwidth, the *minimum cell rate*. During the lifetime of an ABR connection, the network can set the traffic rate of the connection to any value in the range between the minimum and the peak cell rate.

The difficulty of defining the ABR service class lies in finding flow control mechanisms that can adapt the cell rate of ABR traffic sources to dynamically changing network conditions. These flow control mechanisms must satisfy a number of requirements:

---

\* Corresponding author. Tel.: 404 894 5141; fax: 404 894 7883; e-mail: ian@ee.gatech.edu

<sup>1</sup> Supported in part by the National Science Foundation under Grant No. NCR-9309224 and NCR-9624106.

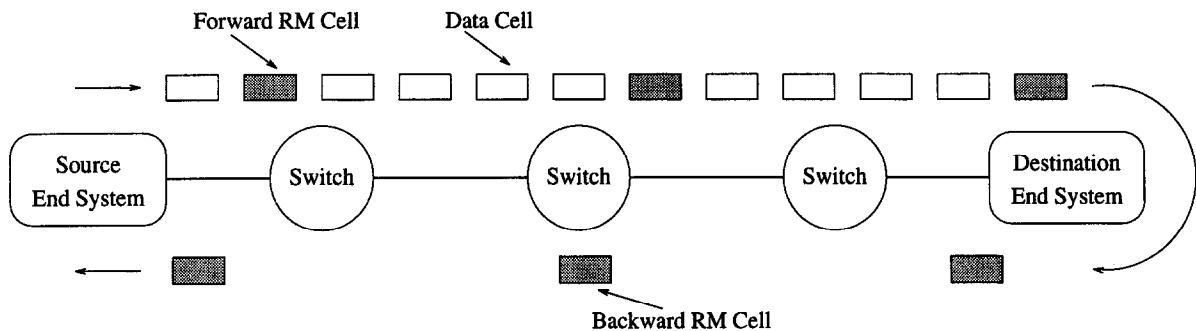


Fig. 1. Closed-loop traffic control.

- Adjust the ABR traffic to accommodate the bandwidth needs of CBR and VBR service classes which have more stringent QoS requirements.
- Detect and react to short-term congestion conditions that may arise due to rate fluctuations of VBR traffic.
- Adapt ABR traffic sources rapidly if bandwidth becomes available.
- Give fair treatment to all ABR traffic sources.

After considering a number of different flow control approaches for ABR traffic [14,15,17,18,23], the ATM Forum settled on a closed-loop rate-based scheme for ABR traffic with a control loop as shown in Fig. 1 [8]. An ABR source periodically sends *resource management* (RM) cells to the destination, which in turn sends the RM cells back to the source. On its roundtrip through the network, the RM cells collect congestion information which is given to the sources as feedback. An RM cell provides two types of feedback: *Binary feedback* indicates to the source the presence of congestion and *explicit rate feedback* informs the sources about its maximum permitted cell rate.

A major drawback of the currently existing explicit rate scheme [21] is that it tries to allocate equal bandwidth 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 a different treatment to different classes of ABR connections. In particular, the need to consider multiple traffic classes becomes apparent if ABR connections are used for (adaptive) multimedia applications with bandwidth requirements ranging from low-bandwidth audio encoders to high-quality compressed video streams.

We propose an explicit rate flow control scheme for ABR traffic that operates at multiple levels. We assume that each ABR connection is assigned to one *traffic class*, where the class assignment of a connection is based on the application type, on traffic parameters such as the minimum or peak cell rate, or on extraneous factors such as the location of the traffic source. The presented scheme is based on a multi-level bandwidth control management for internetwork traffic [1] and an earlier application of this framework to ABR flow control [16].

We control the availability of bandwidth at three levels, as illustrated in Fig. 2 for a single link:

- *Serve-level flow 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 demands of CBR and VBR traffic<sup>2</sup>.

<sup>2</sup> We ignore the presence of UBR traffic classes in this paper.

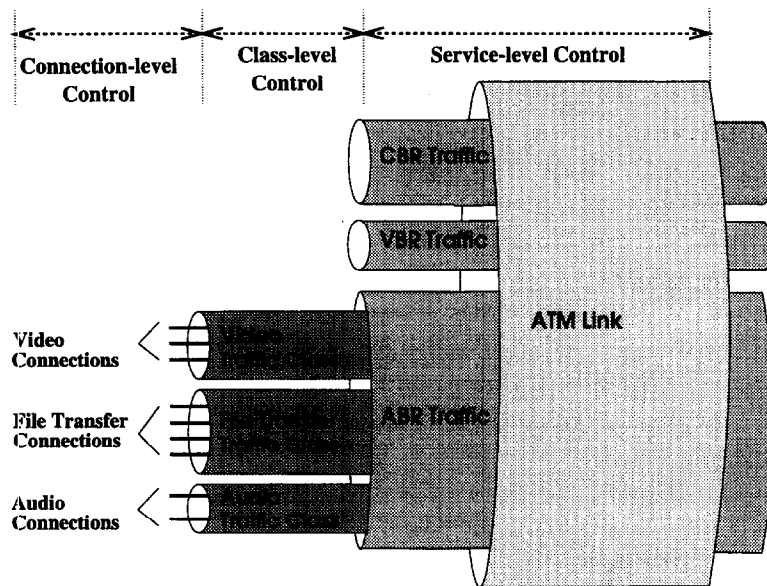


Fig. 2. Multi-level bandwidth control at an ATM link.

- *Class-level flow control* is the second level of control. It distributes the available ABR bandwidth on a link among multiple traffic classes. For instance, in Fig. 2 we show three different traffic classes; connections that perform file transfers, connections with video traffic, and connections with audio traffic. Class-level flow control attempts to distribute unused bandwidth to traffic classes with a high bandwidth demand.
- *Connection-level flow control* distributes the bandwidth that is available to a specific traffic class among all connections from this class. Connection-level flow control implements the *max-min fairness* or *bottleneck flow control* [2,7,10] scheme, specified in the ATM Traffic Management Specification of the ATM Forum [21]. It allocates an equal amount of bandwidth to all ABR connections that are bottlenecked at the same link. However, different from [21], our connection-level flow control implements max-min fairness independently for each traffic class.

The flow control operations at the different levels are highly interdependent, that is, any action of service-level flow control will have an impact on class-level flow control, and changes at the class-level will lead to changes at the connection-level. One of the features of our multi-level flow control scheme is that it accounts for the interdependencies of the multiple levels while maintaining full utilization of the available ABR bandwidth. In addition, by simulation we show that the multi-level flow control works well in conjunction with a binary feedback scheme.

The remainder of the paper is structured as follows. In Section 2 we review the rate-based flow control scheme for ABR traffic as specified by the ATM Forum. In Section 3 we characterize the multi-level flow control scheme. In Section 4, we indicate how to incorporate our multi-level flow control into the existing ABR flow control protocol. 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 rate-based flow control for ABR traffic as specified by the Traffic Management Group of the ATM Forum. Our review is based on Revision 11 of the Traffic Management Specification Version 4.0 [21]. Note that the review is not comprehensive. For a more complete discussion we refer to the ATM Forum documents [21] and a set of excellent survey papers [3,6,11,22].

### 2.1. The ATM Forum specification

During connection establishment, a source endsystem negotiates with the ATM network the *maximum (peak) cell rate* (PCR), the *minimum cell rate* (MCR), and an initial cell rate (ICR) for a new ABR connection. At all times the *allowed cell rate* (ACR) of an ABR connection gives the maximum transmission rate currently supported by the network with  $MCR \leq ICR \leq ACR \leq PCR$ . If an ABR connection is established, the source periodically generates *resource management* (RM) cells that are interleaved with the stream of data cells and sent to the destination. The destination turns the RM cells around and sends them back to the source. On its round trip, the RM cells collect congestion information from the switches and the destination. This information is used by the source to adjust its traffic rate.

An RM cell that is generated by an ABR source contains fields for binary feedback as well as for explicit rate feedback. The *congestion indication (CI) flag* and the *no increase (NI) flag* of an RM cell furnish the source with binary congestion information, i.e. they inform the source about the presence or absence of congestion in the network. The CI and NI flags are set by the intermediate switches or by the destination if they experience local congestion. The *explicit rate (ER) field* of an RM cell is used to return to the source the maximum tolerable cell rate, so-called *explicit rate*. When the source generates an RM cell it initializes the ER field to PCR. Switches or the destination participate in the explicit rate calculation by reducing (never increasing!) the content of the ER field. The value of the ER field that is returned to the source is a bound on the cell rate that cannot be exceeded by the source. When a source receives a returning RM cell, it performs the following operations<sup>3</sup>:

$$\begin{array}{ll} \text{if } CI = 1 & \text{then } ACR \leftarrow \max(ACR - RDF \cdot ACR, MCR), \\ \text{if } CI = 0 \text{ and } NI = 0 & \text{then } ACR \leftarrow \min(ACR + RIF \cdot PCR, ER), \end{array}$$

where *RIF* is a *rate increase factor* and *RDF* is a *rate decrease factor*.

### 2.2. Explicit rate calculations

The ATM Forum did not standardize a particular method for calculating the explicit rates ER as part of the rate-based flow control scheme. Note that in addition to the feedback protocol discussed above, an explicit rate scheme involves two additional components:

1. A *policy or fairness criteria* that specifies how the available bandwidth should be allocated to the ABR connections. An obvious requirement for such a policy is that the bandwidth allocation be independent of the location of the source and the route of a connection. Moreover, the rate calculation should treat all connections “fairly” in respect to a given set of fairness requirements.

<sup>3</sup> The given operations are simplified versions of the actual calculations (see [3,6,21]).

2. An *algorithm* executed by sources and switches for computing the explicit rate so as to satisfy the fairness criteria.

Within the ATM community, the most popular policy for setting the explicit rate is *max-min fairness* [2,7,10]. This policy gives each ABR connection the same maximum throughput allocation, so-called *share*, 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 share, the *bottleneck*, connections that have the same bottleneck link have identical throughput constraints. Denoting the number of bottlenecked flows at a link by  $n$ , and the bandwidth available to all bottlenecked flows by  $B$ , then the *share* of a connection  $i$ ,  $Share_i$ , is calculated as [21]:

$$Share_i = B/n.$$

Many algorithms for implementing max-min fairness were presented to the ATM Forum, including EPRCA, ERICA and others [4,5,11–13,19,21].

A major disadvantage of *max-min fairness* is that it enforces the same fairness policy on all ABR connections and does not differentiate among different types of ABR connections. However, since the bandwidth demands of ABR connections can vary by several orders of magnitude, more sophisticated fairness policies which can express service requirements for a large set of applications are urgently needed. Several improvements to max-min fairness have tried to address this issue [11,21]. For example in a *proportional scheme*, the share of a connection is proportional to its MCR. Denoting by  $MCR_i$  the MCR of connection  $i$  and by  $\sum_j MCR_j$  the sum of the MCR's from all connections, the share of connection  $i$  is calculated as [21]:

$$Share_i = B \cdot \frac{MCR_i}{\sum_j MCR_j}.$$

The proportional scheme can be generalized to a *weighted version* where each connection is assigned a weight  $w_j$ . Then the share is calculated by [21]:

$$Share_i = B \cdot \frac{w_i}{\sum_j w_j},$$

where  $\sum_j w_j$  is the sum of weights from all connections.

While these (and other) extensions of max-min fairness calculate different share values for ABR connections, they are limited in several ways. For example, the addition of a single ABR connection may reduce the share value of all other ABR connections. However, if ABR connections are to support a diverse set of applications, it is desirable to completely decouple the share calculations for different application types. These drawbacks will be overcome by our new multi-level explicit rate scheme. The key approach of our scheme is a decoupling of the explicit rate calculation for different ABR traffic classes.

### 3. Multi-level flow control of ABR traffic

The goal of the multi-level flow control scheme is to find for each ABR connection the maximum traffic rate which complies with the control objectives. Flow control is performed at three levels. At the connection level, we control the bandwidth available to ABR connections within the same traffic class. This level of flow control is currently well-understood and applied by the ATM Forum to calculate the explicit rate of

ABR connections. At the class level, we dynamically control the bandwidth available to each ABR traffic class; classes with a high bandwidth demand can temporarily borrow bandwidth from traffic classes with a low bandwidth demand. Finally, at the service level, we control the availability of bandwidth to all ABR connections in the network.

We consider an ATM network where switches are connected by unidirectional ATM links. Each ABR connection in this network has a fixed route with a unidirectional traffic flow and is assigned to exactly one traffic class. The first leg of the route of a connection is called the *source*. For the purposes of this paper it is convenient to view the source as a link that carries only one connection. We introduce the following notation:

$\mathcal{N} = \bigcup_{p=1}^P \mathcal{N}_p$	set of all connections; $\mathcal{N}_p$ is the set of all class- $p$ connections ( $1 \leq p \leq P$ )
$\mathcal{S} = \{s_i   i \in \mathcal{N}\}$	set of connection <i>sources</i>
$\mathcal{L}$	set of (unidirectional) links in the ATM network
$C_l$	capacity of a link or a source; the capacity of a source $s_i$ is given by the peak cell rate of the connection, i.e., $C_{s_i} = \text{PCR}_i$
$\mathcal{R}_i = (s_i, l_{i1}, l_{i2}, \dots, l_{iK})$	route of ABR connection $i$ ; $s_i$ is the source of connection $i$ , $l_{ik} \in \mathcal{L}$ is the $k$ th link on the route of connection $i$
$\Delta_{lp} = \{i   l \in \mathcal{R}_i, i \in \mathcal{N}_p\}$	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 its *peak cell rate*, denoted by  $\text{PCR}_i$ , and its *minimum cell rate*, denoted by  $\text{MCR}_i$ . If the network does not have sufficient bandwidth to satisfy  $\text{MCR}_i$ , then the connection will not be established. The ability of the ATM network to support  $\text{MCR}_i$  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 flow control scheme is called the *explicit rate*, and denoted by  $\text{ER}_i$  for connection  $i$ .  $\text{ER}_i$  is the maximum traffic rate allowed by the multi-level flow control, and we have  $\text{MCR}_i \leq \text{ER}_i \leq \text{PCR}_i$ . Recall from Section 2 that the allowed cell rate  $\text{ACR}$  of a connection may be smaller than the explicit rate due to additional binary feedback congestion schemes.

The multi-level flow control scheme for ABR traffic allows us to calculate the explicit rates  $\text{ER}_i$  from a set of control parameters for each link in the network. The control parameters in this study are as follows:

- *Connection-level flow control* enforces throughput bounds for all connections at all links. The *share* of a class- $p$  connection for link  $l$ , denoted by  $\text{Share}_{lp}$ , denotes the throughput bound for all class- $p$  connections at link  $l$ . Therefore, for each connection  $i \in \mathcal{N}_p$ :

$$\text{ER}_i \leq \text{Share}_{lp} \quad \text{for all } i \in \Delta_{lp}. \quad (1)$$

- *Class-level flow control* enforces throughput bounds for the aggregate bandwidth of connections from the same class. We use  $C_{lp}^{\text{ABR}}$  to denote the available capacity for all class- $p$  connections at link  $l$ , so-called *class capacity*. For each traffic class  $p$  we have

$$\sum_{i \in \Delta_{lp}} \text{ER}_i \leq C_{lp}^{\text{ABR}} \quad \text{for all } l \in \mathcal{L}. \quad (2)$$

- *Service-level flow control* bounds the aggregate throughput of ABR connections on a link by a so-called *ABR capacity*, denoted by  $C_l^{\text{ABR}}$ , i.e.,

$$\sum_{p=1}^P \sum_{i \in \Delta_{lp}} \text{ER}_i \leq C_l^{\text{ABR}} \quad \text{for all } l \in \mathcal{L}. \quad (3)$$

In the following sections we describe the multi-level flow control scheme in detail. In Section 3.1 we describe a method for selecting throughput bounds for connection-level flow control. In Section 3.2 we discuss class-level flow control and present a scheme that varies the bandwidth available to ABR traffic classes according to the bandwidth needs in each class. In Section 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 flow control

In this section 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.,  $C_{lp}^{ABR} \equiv \text{const}$ . In this case, the bandwidth left unused by some traffic class cannot be made available to other traffic classes. In the Section 3.2 we will show how the assumption of fixed class capacities can be relaxed. 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  $\text{MCR}_j$  can be supported on all links on the route of a connection  $j$ , i.e.,

$$\sum_{i \in \Delta_{lp}} \text{MCR}_i \leq C_{lp}^{ABR} \quad \text{for all } l \in \mathcal{R}_j. \quad (4)$$

Connection-level flow control distributes the class capacity  $C_{lp}^{ABR}$  to the class- $p$  connection on a link  $l$ . By enforcing shares  $\text{Share}_{lp}$  at each network link, the maximum end-to-end throughput of an ABR connection  $i$ , given by its explicit rate  $\text{ER}_i$ , is limited by the link on the connection's route:

$$\text{ER}_i = \min_{l \in \mathcal{R}_i} \text{Share}_{lp}. \quad (5)$$

The link at which the minimum is attained in Eq. (5) is called the *bottleneck link*, denoted by  $l_i^*$ . If  $l_i^* = s_i$ , we say that a connection is 'bottlenecked at the source'.

By enforcing that shares of all overloaded class- $p$  connections at a link are identical, we implement an intuitive notion of *fairness*, in the sense that all connections that have the same bottleneck link have identical throughput constraints [2,7,10,24]. Therefore, we will refer to the shares as *fair share*.

Given an assignment of share values on each link, the ABR connections for class- $p$  on a link  $l$  are labeled as *overloaded* or *restricted at link  $k$* . The set of overloaded connections, denoted by  $\mathbf{O}_{lp}$ , contains all class- $p$  connections at link or source  $l$  that have their bottleneck at  $l$ . Connections at link or source  $l$  that are 'restricted at  $k$ ', denoted by  $\mathbf{R}_{lp}(k)$ , have their bottleneck on some link  $k \in \mathcal{L} \cup \mathcal{S}$  with  $k \neq l$ :

$$\mathbf{O}_{lp} = \{i \in \Delta_{lp} \mid l_i^* = l\}, \quad (6)$$

$$\mathbf{R}_{lp}(k) = \{i \in \Delta_{lp} \mid l_i^* = k, k \in \mathcal{R}_i\} \quad \text{for } k \neq l, k \in \mathcal{L} \cup \mathcal{S}. \quad (7)$$

Since each source only has one connection, we have  $|\mathbf{O}_{sp}| \leq 1$  and  $|\mathbf{R}_{lp}(k)| \leq 1$  for each  $s \in \mathcal{S}$ .

Next we derive a flow control scheme where the values of the fair shares are selected maximally. In this case, the connection-level flow control scheme is identical to *max-min fair* flow control [2]. Note that fair shares need to be defined only for links with at least one overloaded connection. If such a connection exists, i.e.,  $\mathbf{O}_{lp} \neq \emptyset$ , then the entire available bandwidth can be allocated to the ABR connections:

$$C_{lp}^{ABR} = \sum_{i \in \Delta_{lp}} \text{ER}_i \quad (8)$$

$$= \sum_{i \in \Delta_{lp}} \min_{l \in \mathcal{R}_i} \text{Share}_{lp} \quad (9)$$

$$= \sum_{k \in \mathcal{L} \cup \mathcal{S}} \sum_{i \in \mathbf{R}_{lp}(k)} \text{Share}_{lp} + \sum_{i \in \mathbf{O}_{lp}} \text{Share}_{lp} \quad (10)$$

$$= \sum_{k \in \mathcal{L} \cup \mathcal{S}} |\mathbf{R}_{lp}(k)| \cdot \text{Share}_{lk} + |\mathbf{O}_{lp}| \cdot \text{Share}_{lp}. \quad (11)$$

Therefore, maximal fair shares, denoted by  $\text{Share}_{lp}^*$ , are obtained by

$$\text{Share}_{lp}^* = \begin{cases} \infty & \text{if } \mathbf{O}_{lp} \neq \emptyset, \\ \frac{1}{|\mathbf{O}_{lp}|} \left( C_{lp}^{\text{ABR}} - \sum_{k \in \mathcal{L}} |\mathbf{R}_{lp}(k)| \cdot \text{Share}_{kp}^* \right) & \text{otherwise.} \end{cases} \quad (12)$$

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

**Note.** If a connection  $i$  is bottlenecked at the source then the maximal share of the source, according to Eq. (12), is  $\text{Share}_{si}^* = \text{PCR}_i$ . This follows from the assumption that  $C_{si} = \text{PCR}_i$  and the fact that a source only has one connection.

It is noteworthy that the calculation of maximal shares does not require knowledge on individual connections. More precisely, two pieces of information must be at hand to determine the maximal share for class  $p$  on a link: (1) the bottleneck throughput of class- $p$  connections that are overloaded on some other link, and (2) the number of overloaded class- $p$  connections on the link.

### 3.2. Class-level flow control

The flow control scheme for calculating the explicit rates described so far has a major drawback. Namely, 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*  $g_{lp}$  and the *surplus*,  $\text{Surplus}_{lp}$ . The class guarantee  $g_{lp}$  is the guaranteed portion of the ABR capacity  $C_{lp}^{\text{ABR}}$  that is available to connections from class  $p$  at link  $l$ . We also define  $G_{lp} = g_{lp} C_{lp}^{\text{ABR}}$ , denoted as *guaranteed class rate*. We assume  $\sum_{p=1}^P g_{lp} = 1$ , i.e., the class guarantees divide the entire ABR capacity on a link  $l$ .

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

$$\sum_{i \in \Delta_{lp}} \text{MCR}_i \leq G_{lp} \quad \text{for all } l \in \mathcal{R}_j. \quad (13)$$

The surplus bandwidth, denoted by  $\text{Surplus}_p$ , gives the bandwidth in excess of the guaranteed class rate that is temporarily made available to class  $p$ . Of course, the surplus can be nonzero only if some other



classes do not utilize their respective class guarantees, i.e., if  $G_{lq} - \sum_{i \in \Delta_{lq}} ER_i > 0$  for some traffic classes  $q \neq p$ .

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

$$C_{lp}^{ABR} = \min \left( \sum_{i \in \Delta_{lp}} ER_i, G_{lp} + Surplus_{lp} \right). \quad (14)$$

The goal of connection-level flow control is to select the values for  $Surplus_{lp}$  as large as possible. Note that the connections on a link  $l$  utilize the maximum available bandwidth  $G_{lp} + Surplus_{lp}$ , only if there is at least one traffic class that contains overloaded connections. Assuming that such a class exists at link  $l$  and assuming that the maximal fair shares  $Share_{lp}^*$  are available for all links, then a class-level flow control scheme satisfies

$$C_l^{ABR} = \sum_{\mathbf{O}_{lq} \neq \emptyset} (G_{lq} + Surplus_{lp}) + \sum_{\mathbf{O}_{lq} = \emptyset} \sum_{i \in \mathcal{N}_q} ER_i \quad (15)$$

$$= \sum_{\mathbf{O}_{lq} \neq \emptyset} (G_{lq} + Surplus_{lp}) + \sum_{\mathbf{O}_{lq} = \emptyset} \sum_{k \in \mathcal{LUS}} |\mathbf{R}_{lq}(k)| \cdot Share_{kq}^*. \quad (16)$$

Denote the total surplus bandwidth at a link  $l$  by  $Surplus_l$ . From the above,  $Surplus_l$  is given by

$$Surplus_l := \sum_{\mathbf{O}_{lq} \neq \emptyset} Surplus_{lp} = C_l^{ABR} - \sum_{\mathbf{O}_{lq} \neq \emptyset} G_{lq} - \sum_{\mathbf{O}_{lq} = \emptyset} \sum_{k \in \mathcal{LUS}} |\mathbf{R}_{lp}(k)| \cdot Share_{kp}^*. \quad (17)$$

One can think of several methods for dividing the surplus  $Surplus_l$  among the traffic classes. For example,  $Surplus_{lp}$  could be selected proportionally to the total traffic load of a class or proportionally to the number of connections in a class. In this study, we divide the  $Surplus_l$  evenly among all classes with overloaded flows. Then we obtain from the above:

$$Surplus_{lp} = \begin{cases} Surplus_l & \text{if } \bigcup_{q=1}^P \mathbf{O}_{lq} = \emptyset, \\ \frac{Surplus_l}{|\{q | \mathbf{O}_{lq} \neq \emptyset\}|} & \text{otherwise.} \end{cases} \quad (18)$$

### 3.3. Service-level flow control

So far we have not accounted for the fact that the bandwidth available to ABR traffic depends on the bandwidth allocated 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 from 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

$$C_l^{\text{ABR}} = \max(C_l^{\text{min}}, C_l - \Gamma_l^{\text{CBR}} - \Gamma_l^{\text{VBR}}). \quad (19)$$

Since  $G_{lp} = g_{lp} C_l^{\text{ABR}}$ , changing  $C_l^{\text{ABR}}$  requires to recalculate the guaranteed class rates  $G_{lp}$ .

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

$$\sum_{p=1}^P \sum_{i \in \Delta_{lp}} \text{MCR}_i \leq C_l^{\text{min}} \quad \text{for all } l \in \mathcal{R}_j. \quad (20)$$

This concludes the discussion of the multi-level flow control of ABR traffic. Note that the multi-level flow control enables the ABR sources to fully utilize the available ABR capacity. In an actual protocol all the levels of flow control are simultaneously active. In the next section we show how to modify the flow control protocol developed by the ATM Forum [21] to implement our multi-level flow control scheme.

#### 4. Protocol mechanisms for multi-level flow control

In this section we discuss how to incorporate the multi-level flow control scheme from the previous section into the framework of ABR traffic management as developed by the ATM Forum [21]. The protocol mechanisms described here are mainly modifications or additions to the ABR control protocol reviewed in Section 2. Therefore, we focus 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, but assume the existence of a binary feedback scheme to detect and react to short-term congestion, as described in Section 2. In simulation experiments, presented in Section 5, we demonstrate how multi-level flow control and binary feedback schemes interoperate.

##### 4.1. Modifications to the RM cell format

We require only a minor modification to the RM cell format described in [21]. All bit flags, such as 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 [21]. The difference will become clear when we describe the switch behavior in Section 4.4. We do not require the CCR and MCR fields. The only addition to the RM cell format described in [21] is a so-called *bottleneck field*.

BNK – The *bottleneck field* contains a unique identification of the bottleneck link of the connection that issued the RM cell. The field is modified by the intermediate switches on the forward pass of the RM cell.

##### 4.2. Source and destination behavior

If the source of a class- $p$  connection, say  $\text{src}$ , issues an RM cell, it initializes the ER field to the peak cell rate, i.e.,  $\text{ER} = \text{PCR}$ , and the bottleneck field  $\text{BNK} = \text{src}$ . Both fields are modified by the intermediate switches and the destination.

The content of the ER and BNK fields of an RM cell that returns to the source are interpreted as follows. If  $\text{BNK} = \text{src}$ , then the connection is ‘bottlenecked at the source’. If the content is  $\text{BNK} = S$  then the connection  $i$  is ‘bottlenecked at link  $S$ ’.

Overloaded connections can transmit at most at the rate given by the ER field of the last RM cell. We will see below that connections that are bottlenecked at the source can transmit at their peak cell rate. As in the ATM Forum draft [21], the maximum transmission rate can be lower than the explicit rate due to a binary congestion scheme.

The destination plays no special role in the calculation of the explicit rate. As described in Section 2, the destination transmits each RM cell back to the source. If the destination can experience congestion, it should perform the same operations as outlined in Sections 4.4 and 4.5.

### 4.3. Switch behavior

Each switch maintains information on each ABR connection that has an outgoing route on one of its output ports. In the following we refer to an output port as a *link*. The information for a connection  $i$  that passes through a link consists of variables  $\text{MaxRate}_i$  and  $\text{OV}_i$ .  $\text{MaxRate}_i$  is the link’s current knowledge of the maximum allowed transmission rate of connection  $i$ , and  $\text{OV}_i$  is an *overload flag* which is set when connection  $i$  is bottlenecked at the link.

In addition, the link maintains a set of variables needed for calculating the throughput bounds of the connections:  $\text{Share}_p$  contains the maximum cell rate at which connections from class  $p$  can transmit at the link, the guaranteed class rate  $\text{Guar}_p$  and the *Surplus* are used to calculate the total bandwidth available for traffic from class  $p$  at this link. Finally,  $\text{Cap}^{\text{ABR}}$  is the total available capacity for ABR traffic at this link.

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

$$\text{ER} = \text{Share}_p, \quad \text{BNK} = L. \quad (21)$$

In other words, the switch sets the explicit cell rate to the maximum rate at the outgoing link, and identifies the link as the bottleneck of the connection. Clearly, if an RM cell returns to the source, the BNK field contains the link with the smallest value for  $\text{Share}_p$  on the route of the connection.

When an RM cell from class- $p$  connection  $i$  arrives to a switch on link  $L$  ( $L \neq K$ ), the switch updates the values of two local variables  $\text{MaxRate}_i$  and  $\text{OV}_i$ . The update operations depend on content of the BNK field in the RM cell. If the RM cell is set to  $\text{BNK} = \text{src}$  or to  $\text{BNK} = K$ , the switch updates the information on connection  $i$  that is kept for link  $L$  as follows:

$$\text{MaxRate}_i = \text{ER}, \quad \text{OV}_i = 0. \quad (22)$$

If the bottleneck field of the RM cell that arrives on link  $L$  is set to  $\text{BNK} = L$ , the information on the connection is updated to:

$$\text{MaxRate}_i = 0, \quad \text{OV}_i = 1. \quad (23)$$

#### 4.4. Operations at update intervals

Periodically, a switch uses the values of  $\text{MaxRate}_i$  and  $\text{OV}_i$  on the ABR connections to calculate new throughput bounds for its links. The time periods at which the values are calculated are determined by 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. A switch recalculates for each of its outgoing links the maximum allowable cell transmission rate  $\text{Share}_p$  for connections from traffic class  $p$ . At the end of a *surplus interval* it additionally recalculates the surplus value  $\text{Surplus}$ . At the end of an *ABR capacity interval* the switch also recalculates the capacity available for ABR traffic, denoted by  $\text{Cap}^{\text{ABR}}$ .

At the end of each share interval, a switch calculates for each outgoing link the following values for each traffic class  $p$ :

$$\text{OL}_p := \sum_{i \in \mathcal{N}_p} \text{OV}_i, \quad \text{RATE}_p := \sum_{i \in \mathcal{N}_p} \text{MaxRate}_i. \quad (24)$$

From Eq. (23), it is clear that  $\text{OL}_p$  contains the number of class- $p$  connections that are *bottlenecked* at the current link. Likewise, according to Eq. (22),  $\text{RATE}_p$  contains the maximum traffic rate from connections that are *bottlenecked* at some other link or at the source. After obtaining the values for  $\text{OL}_p$  and  $\text{RATE}_p$ , the following calculations are performed:

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

$$\text{Share}_p = \begin{cases} \infty & \text{if } \text{OL}_p = 0, \\ \frac{\text{Guar}_p + \text{Surplus} - \text{RATE}_p}{\text{OL}_p} & \text{otherwise.} \end{cases} \quad (25)$$

2. At the end of a *surplus interval* the switch re-calculates  $\text{Surplus}$  as follows:

$$\text{Surplus} = \begin{cases} \text{Cap}^{\text{ABR}} & \text{if } \text{OL}_p = 0 \text{ for all } p, \\ \frac{\text{Cap}^{\text{ABR}} - \sum_{\text{OL}_q > 0} \text{Guar}_q - \sum_{\text{OL}_q = 0} \text{RATE}_q}{|\{q | \text{OL}_q > 0\}|} & \text{otherwise.} \end{cases} \quad (26)$$

3. At the end of an *ABR capacity interval* the switch obtains new values for  $\text{Cap}^{\text{ABR}}$  similar to Eq. (19). Assuming that  $\text{Cap}$  is the total capacity of an outgoing link,  $\text{Cap}^{\text{min}}$  is the minimum capacity available to ABR traffic, and assuming that the cell rates currently allocated to CBR and VBR traffic are given by  $\text{Alloc}^{\text{CBR}}$  and  $\text{Alloc}^{\text{VBR}}$ , respectively, the switch calculates:

$$\text{Cap}^{\text{ABR}} = \max(\text{Cap}^{\text{min}}, \text{Cap} - \text{Alloc}^{\text{CBR}} - \text{Alloc}^{\text{VBR}}). \quad (27)$$

Note that Eqs. (25)–(27) correlate with Eqs. (12), (18), and (19) as follows:

$$\begin{aligned} \text{Cap}^{\text{ABR}} &\equiv C_l^{\text{ABR}}, & \text{Share}_p &\equiv \text{Share}_{lp}^*, \\ \text{Cap}^{\text{min}} &\equiv C_l^{\text{min}}, & \text{Surplus} &\equiv \text{Surplus}_l^*, \\ \text{Cap} &\equiv C_l, & \text{OL}_p &\equiv |\mathbf{O}_{lp}|, \\ \text{Guar}_p &\equiv G_{lp}, & \text{RATE}_p &\equiv \sum_{k \in \mathcal{LUS}} |\mathbf{R}_{lp}(k)| \cdot \text{Share}_{kp}^*. \end{aligned} \quad (28)$$

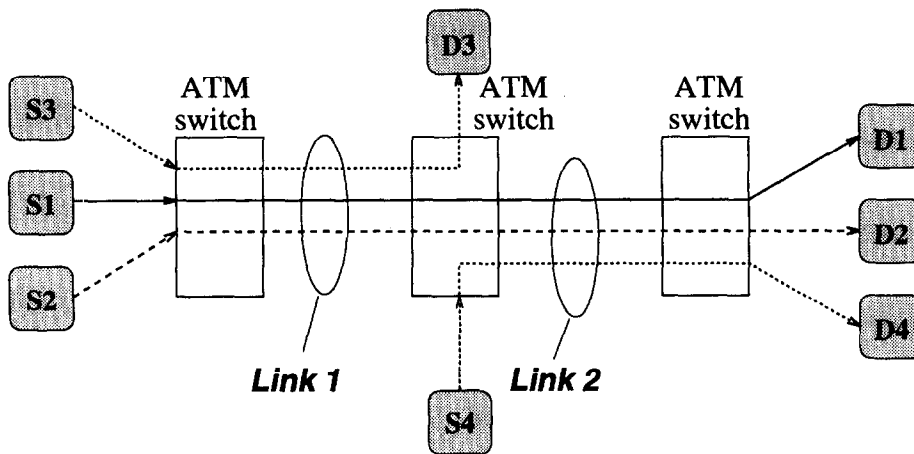


Fig. 3. Simulated network.

## 5. Simulation experiments

To provide some insight into the dynamics of the multi-level flow control scheme, we present simulation experiments of the transient behavior during changes of the network load. The simulations were implemented in a discrete event simulator written in the C programming language. The implementation of the source, destination, and switch behavior of ABR flow control is based on the Traffic Management Specification Version 4.0 from October 1995 [20].

We present four experiments. In each experiment we address one of the following questions:

- *Experiment 1*: How fast does the multi-level flow control protocol converge after load changes? How do connection-level and class-service flow control interact?
- *Experiment 2*: What is the impact of long propagation delays on the convergence of the flow control scheme?
- *Experiment 3*: What is the impact of service level flow control on the stability of the network?
- *Experiment 4*: How does multi-level flow control interact with binary feedback schemes in situations of network congestion?

As shown in Fig. 3, the simulated network consists of three ATM switches connected by two links with a capacity of  $C_l = 155$  Mb/s each. There are four virtual connections (VCs) with source endsystems S1–S4 and destination endsystems D1–D4. 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  $20 \mu\text{s}$  and 2 ms, corresponding to a distance of 6–600 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  $C_l^{\text{min}} = C_l^{\text{ABR}} = 150$  Mb/s. A safety margin of 5 Mb/s is withheld from the physical link capacity of 155 Mb/s to accommodate very short-term rate fluctuations. The four ABR connections are from three different traffic classes: VC1 is from class 1, VC2 and VC3 are from class 2, and VC4 is from class 3. The class guarantees are identical on each link and set to:

class 1	class 2	class 3
$g_1 = 20\%$	$g_2 = 50\%$	$g_3 = 30\%$

Table 1  
Connection parameters

ABR connection	Source system	Destination system	Route	Traffic class	Peak cell rate (Mb/s)	Start time (ms)
VC1	S1	→D1	(Link 1, Link 2)	1	10 70	$t = 0$ $t = 5$
VC2	S2	→D2	(Link 1, Link 2)	2	65	$t = 25$
VC3	S3	→D3	(Link 1)	2	80	$t = 75$
VC4	S4	→D4	(Link 2)	3	50	$t = 125$

Table 2  
Parameters for ABR flow control

Switch parameter	Value	Source parameter	Value
Buffer capacity	2000 cell	N <sub>rm</sub>	32 cells
NI threshold	200 cells	T <sub>rm</sub>	0.1 s
		RTF	1/16
CI threshold			
low	1000	RDF	1/16
high	1500	X <sub>rm</sub>	64

With the default value of  $C_i^{\text{ABR}} = 150 \text{ Mb/s}$ , we obtain the following class rate guarantees.

$$\begin{array}{lll} \text{class 1} & \text{class 2} & \text{class 3} \\ G_1 = 30 \text{ Mb/s} & G_2 = 75 \text{ Mb/s} & G_3 = 45 \text{ Mb/s} \end{array}$$

The default parameters of the four connections in Fig. 3 are summarized in Table 1. All connections are initially idle and start to transmit at the 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. Unless otherwise stated the length of *share intervals* and *surplus intervals* are identical and set to 1 ms. In this case, both surplus values and share values are changed simultaneously.

The parameters for ABR flow control scheme are shown in Table 2. The table contains parameters for the switches and parameters for the sources. Parameters not listed in the table are set to the recommended default values given in [20].

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

$$\text{MCR} = 0 \text{ Mb/s} \quad \text{and} \quad \text{ICR} = 7 \text{ Mb/s}.$$

All RM cells are sent “*in-band*”, i.e., the transmission of RM cells is not discounted to the traffic rate.

### 5.1. Experiment 1: Connection-level and class-level flow control

This experiment will demonstrate that connection-level and class-level flow control with fixed class guarantees quickly converge to the correct values after load changes in the network. All parameters are set to the default values given above; we assume that no CBR or VBR traffic is admitted.

The link latencies are set to 20, 200  $\mu\text{s}$  and 2 ms, corresponding to propagation delays of fiber optic links with approximate lengths 6, 60 and 600 km, respectively. For each link latency, two graphs are presented, one for the throughputs of connections on Link 1 and one for the throughputs of connections on Link 2. The

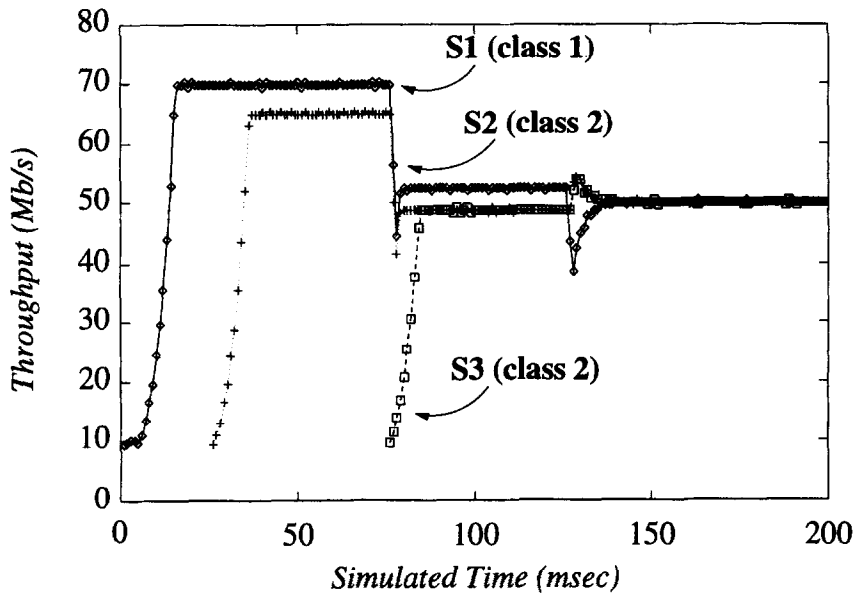
throughput values are given in Mb/s and they are calculated as the data sent from the corresponding source during an interval of 1 ms. Thus, the reported throughput values follow accurately the rate fluctuations resulting from the fairness protocol. The results of Experiment 1 are presented in Fig. 4. We now discuss the outcome of the simulations in detail.

- All connections are initially idle. At  $t = 0$ , connection VC1 from class 1 becomes active with a peak cell rate  $PCR = 10$  Mb/s. This value is increased at  $t = 5$  to  $PCR = 70$  Mb/s. Note that connection VC1 exceeds its bandwidth guarantee of class 1. However, as none of the other classes utilize their bandwidth guarantee, class 1 can ‘borrow’ extra bandwidth from the other classes. This allows VC1 to transmit at its offered load.
- At  $t = 25$ , class-2 connection VC2 begins transmission with  $PCR = 65$  Mb/s. Since sufficient bandwidth is guaranteed to class 2, VC2 can transmit at its peak cell rate.
- At  $t = 75$ , connection VC3 from class 2 starts to transmit on Link 1 with  $PCR = 80$  Mb/s. Then, traffic classes 1 and 2 require all of their respective bandwidth guarantees on Link 1. Since there is no class-3 traffic on Link 1, there is a surplus bandwidth of 45 Mb/s ( $= G_3$ ) on this link. Therefore, class-level flow control takes effect and evenly divides the surplus bandwidth between classes 1 and 2. Since VC1 is the only connection in class 1, 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 2 the available bandwidth on Link 1 after class-level flow control is  $75 + 22.5 = 97.5$  Mb/s. Since there are two class-2 connections on Link 1, connection-level flow control splits the bandwidth between VC2 and VC3. As a result, both connections obtain a throughput of 48.75 Mb/s.
- At time  $t = 125$  ms, connection VC4 from class 3 becomes active on Link 2 with a peak rate of  $PCR = 50$  Mb/s. At this point, the available bandwidth on Link 2, after one considers the current ACR of VC1 and VC2, is 48.75 Mb/s. This available bandwidth is enough to satisfy the minimum guarantee for class 2 and is given to the new connection, but an additional 1.25 Mb/s over the available bandwidth is required by VC4 to satisfy its PCR. Thus, VC4 becomes temporarily bottlenecked at Link 2 until it can claim this additional bandwidth from a recalculation of the surplus. The reduced surplus at Link 2 results in a drop of the share available to VC1 at this link. As a result, the bottleneck of VC1 moves from Link 1 to Link 2 and it causes an increase of the shares available for the class 2 traffic at Link 1. The result is a waste-free equilibrium where the available bandwidth at both links is fully utilized by the ABR traffic and split equally between the three connections present at each link.

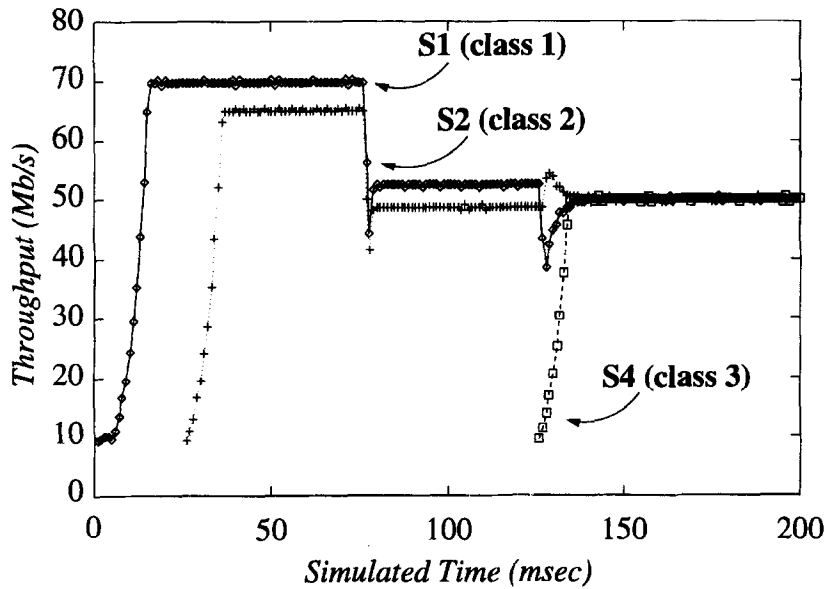
### 5.2. Experiment 2: Impact of link latency

In this experiment, we investigate the impact of the propagation delay on the effectiveness of multi-level flow control. In Fig. 5 we consider the same simulation scenario as in Experiment 1, however, with propagation delays 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. Note that Figs. 4 and 5 are almost identical.

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. Fig. 6 depicts the simulation results if the propagation delay is set to 2000  $\mu$ s per link, corresponding to a length of about 600 km. In this case, the maximum round-trip delay is given by 4 ms. We see in Fig. 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 Fig. 6 that the protocol stabilizes at the correct values.



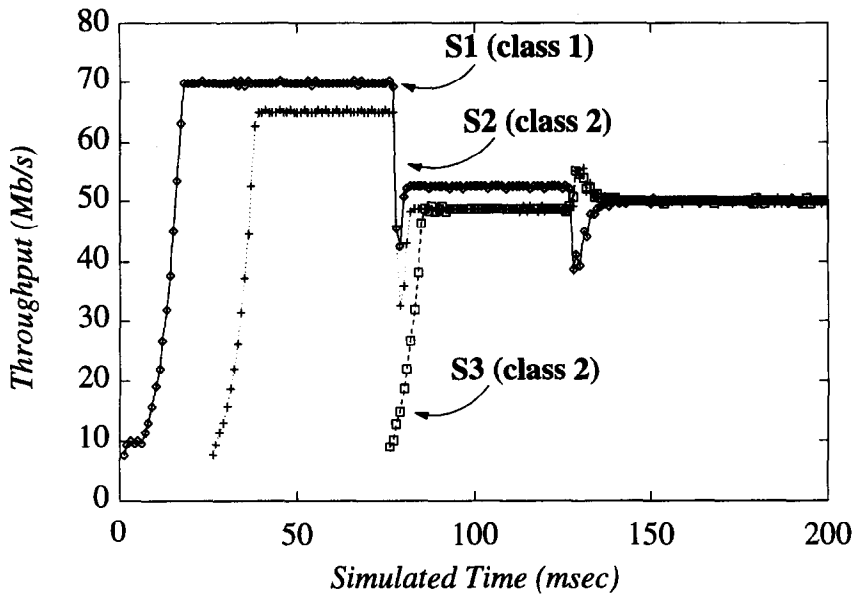
(a)



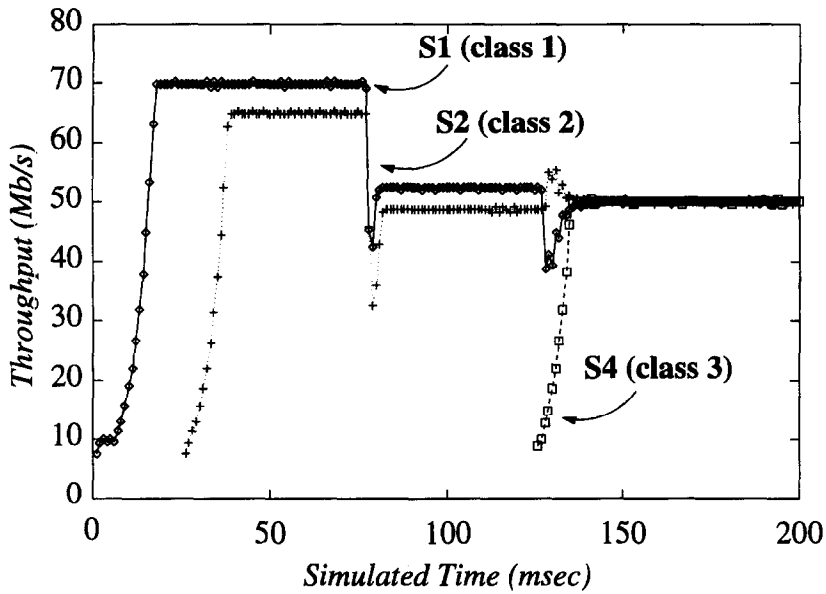
(b)

Fig. 4. Experiment 1:  $20\ \mu\text{s}$  latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$ : (a) throughput on Link 1; (b) throughput on Link 2.



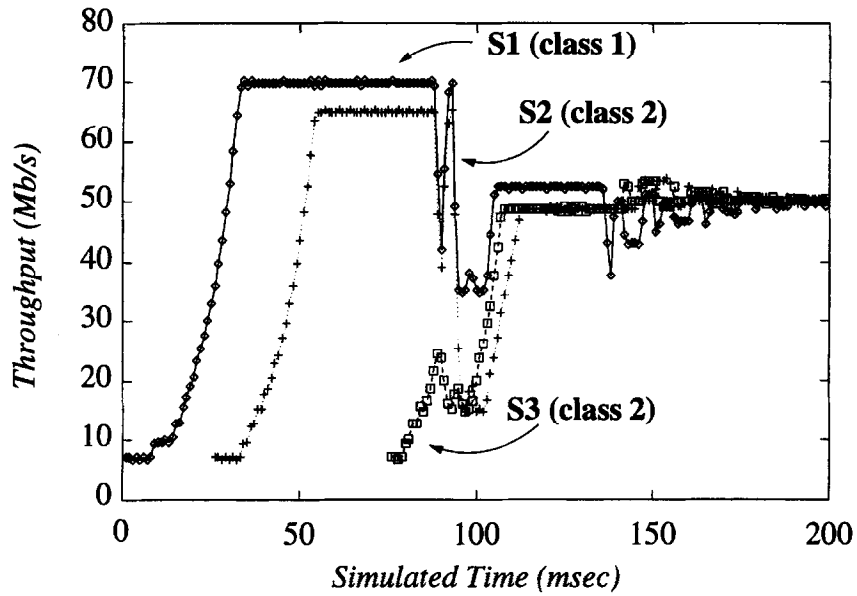


(a)

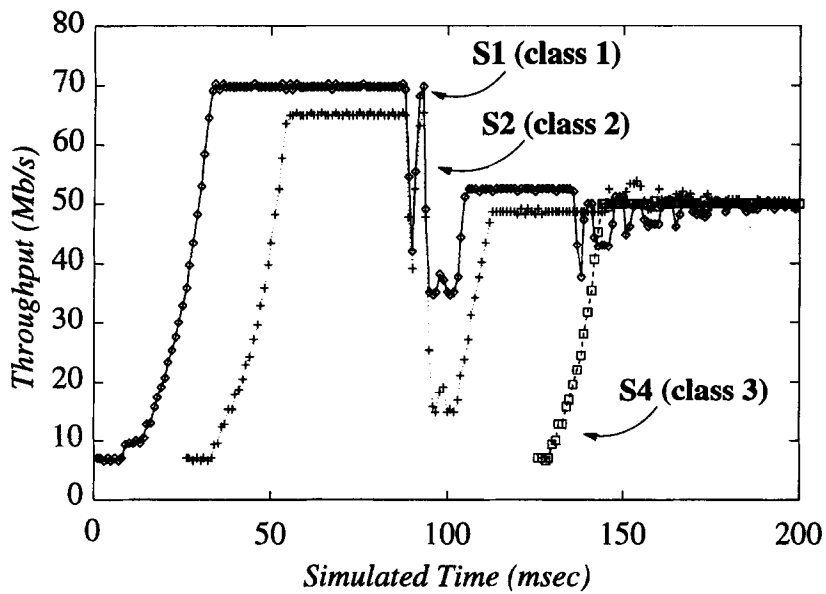


(b)

Fig. 5. Experiment 2: 200  $\mu$ s latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$ : (a) throughput on Link 1; (b) throughput on Link 2.



(a)



(b)

Fig. 6. Experiment 2: 2 ms latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$ : (a) throughput on Link 1; (b) throughput on Link 2.

### 5.3. Experiment 3: Service-level flow control

In the third set of experiments we demonstrate the influence of non-ABR traffic on the equilibrium of the ABR rate control scheme, via service-level flow control. Note from Section 3.3 that CBR and VBR traffic classes are given higher priority than ABR traffic. We consider the same simulation setup as before, however, we assume that from time  $t = 100$  ms until time  $t = 200$  ms, a CBR source is active on Links 1 and 2. The results of the experiment are summarized in Figs. 7 and 8 for a network with  $20 \mu\text{s}$  propagation delay at each link. In Fig. 7, the bandwidth allocation to the CBR traffic source is set to  $\Gamma^{\text{CBR}} = 40$  Mb/s, and in Fig. 8, the bandwidth allocation is set to  $\Gamma^{\text{CBR}} = 100$  Mb/s. As long as the CBR traffic source is active, the class rate guarantees are changed as follows:

	class 1	class 2	class 3
for $\Gamma^{\text{CBR}} = 40$ Mb/s	$G_1 = 22$ Mb/s	$G_2 = 55$ Mb/s	$G_3 = 33$ Mb/s
for $\Gamma^{\text{CBR}} = 100$ Mb/s	$G_1 = 10$ Mb/s	$G_2 = 25$ Mb/s	$G_3 = 15$ Mb/s

Note that the duration of the experiment is extended to 300 ms. We assume that the four ABR sources continue their activity up until time 300 ms, in exactly the same way as they were at time 200 ms. Note that in both experiments, the bandwidth regulation scheme quickly accommodates the demands of the CBR traffic source and brings the ABR rates into a new equilibrium.

### 5.4. Experiment 4: Binary congestion control

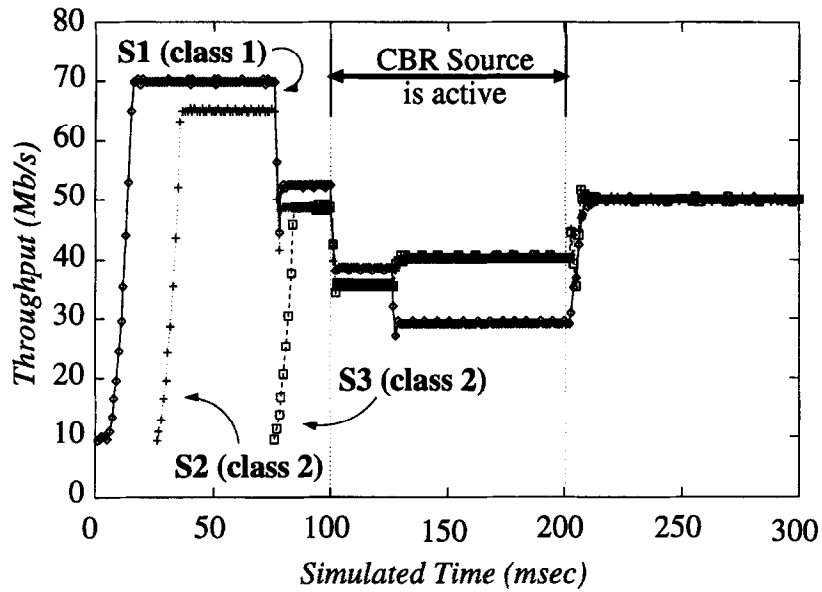
In this experiment we demonstrate that our proposed protocol can coexist with the binary congestion control functions described in the ABR flow control scheme of the ATM Forum specification [21]. More specifically we will demonstrate how the protocol is influenced if the congestion indication mechanisms take effect, i.e., the CI and NI bits are set. We create congestion in the network by adding a bursty VBR traffic source with geometric ON and OFF periods which is active on both Links 1 and 2 for the entire time of the experiment (0–300 ms). The average bit rate of the VBR source is set to 10 Mb/s and its average ON period, the *burst period*, is 1 ms. To intensify the effects of congestion, in this experiment the original buffer sizes from Table 2 are scaled down everything to buffer size = 200, NI threshold = 20, CI threshold (low) = 100 cells and CI threshold (high) 150 cells.

The bandwidth allocation to the VBR source is set equal to its average rate, i.e.,  $\Gamma^{\text{VBR}} = 10$  Mb/s. To be able to compare the plots of the VBR source, we first provide the throughput figures for the case where the VBR connection is replaced by a CBR connection with rate  $\Gamma^{\text{CBR}} = 10$  Mb/s. The results for this case are shown in Fig. 9, and the results for the VBR ON–OFF source are shown in Fig. 10.

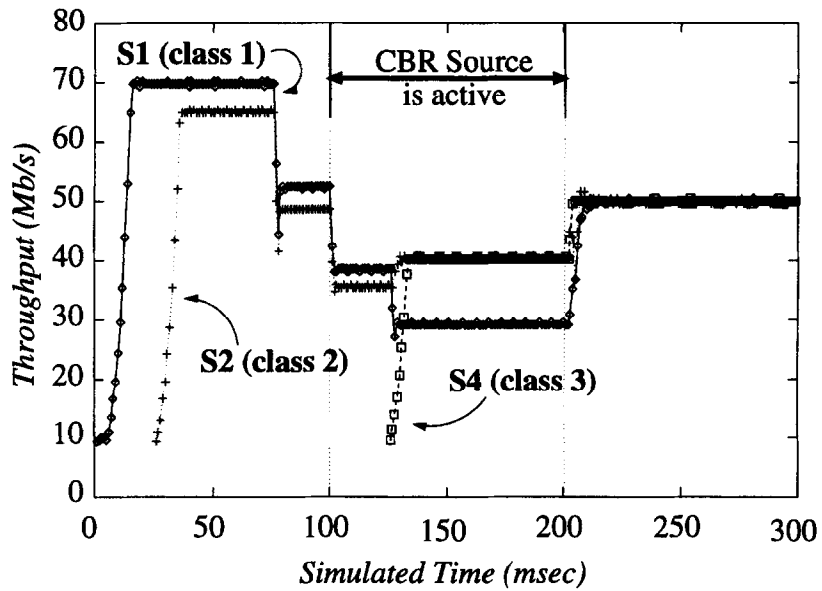
Due to the bandwidth requirements of the VBR (or CBR) source, the ABR capacity on both links is set to  $C_i^{\text{ABR}} = 140$  Mb/s, and the class rate guarantees are set to:

class 1	class 2	class 3
$G_1 = 28$ Mb/s	$G_2 = 70$ Mb/s	$G_3 = 42$ Mb/s

In Fig. 10 we see the effects of the binary congestion control on the ABR connections as “dips” of the throughput curves. These “dips” are a result of the returning RM cells with the CI bit set, which cause the ABR sources to reduce their ACR *without* ever losing their assigned rates. Thus, the response to the returned CI is simply to facilitate the coexistence with bursty sources like this VBR source *without* any impact on the bandwidth allocation scheme which stays immune to congestion phenomena.

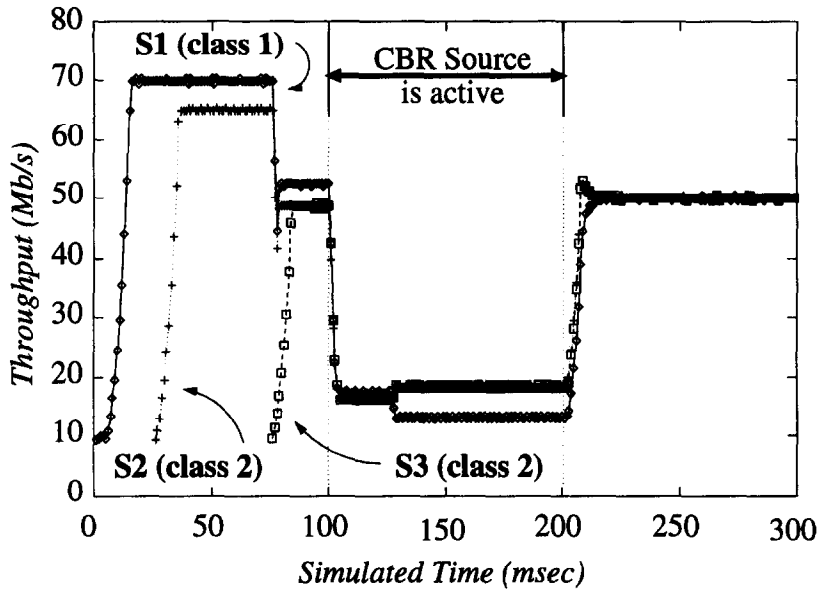


(a)

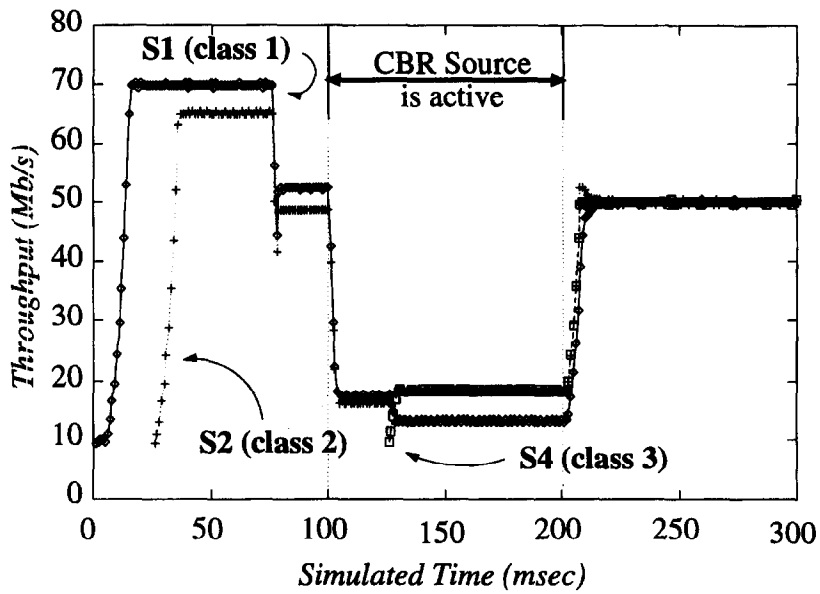


(b)

Fig. 7. Experiment 3:  $20 \mu\text{s}$  latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$  (CBR traffic is active in time interval  $[100, 200]$  ms with rate  $40 \text{ Mb/s}$ ): (a) throughput on Link 1; (b) throughput on Link 2.

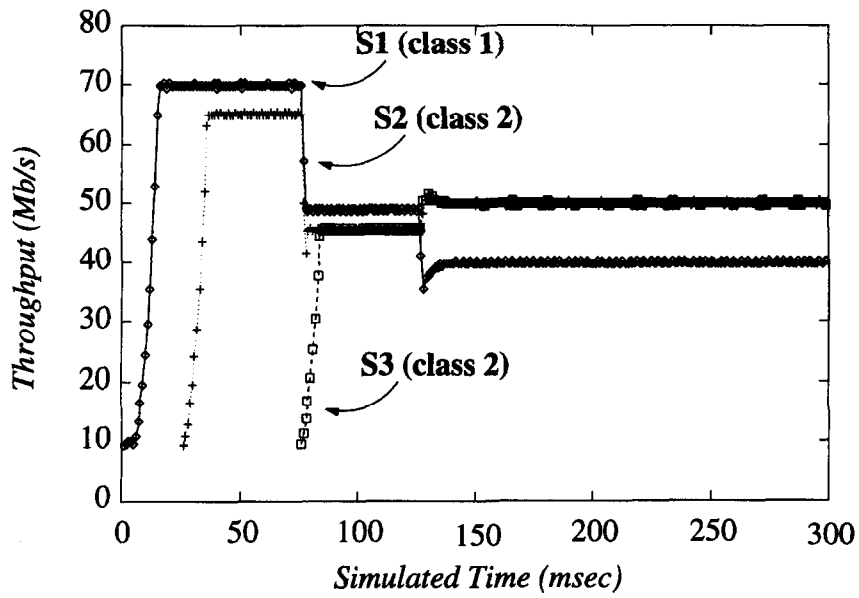


(a)

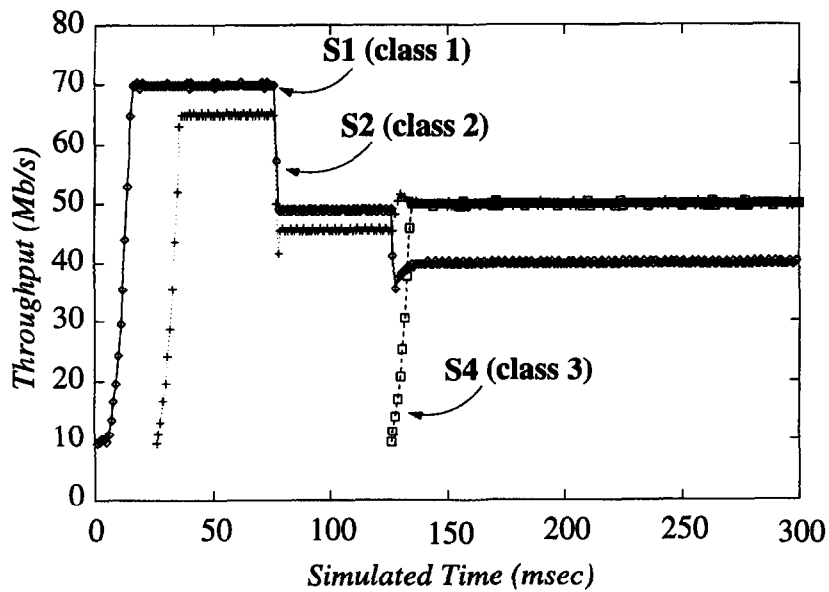


(b)

Fig. 8. Experiment 3: 20  $\mu$ s latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$  (CBR traffic is active in time interval [100, 200] ms with rate 100 Mb/s): (a) throughput on Link 1; (b) throughput on Link 2.

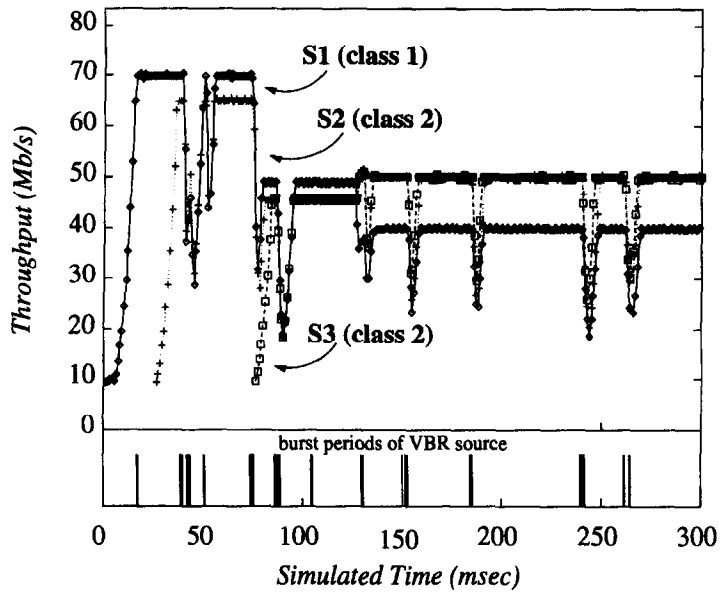


(a)

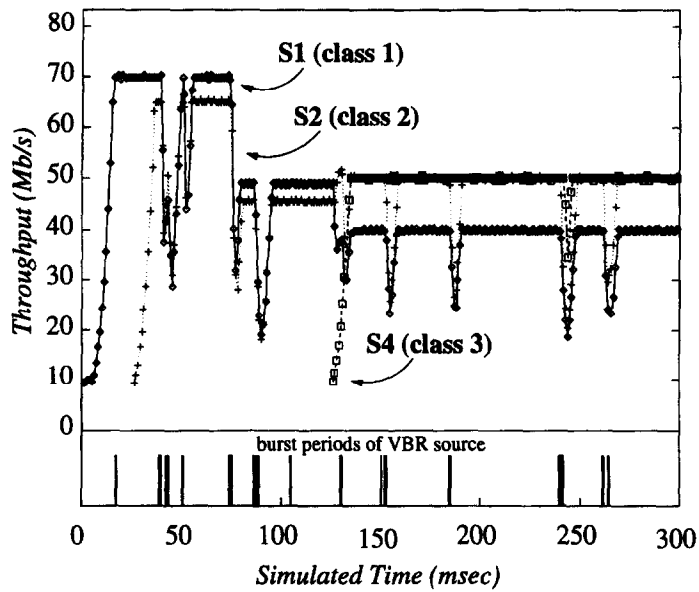


(b)

Fig. 9. Experiment 4:  $20 \mu\text{s}$  latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$  (CBR traffic source is active from 0 to 300 ms with rate 10 Mb/s): (a) throughput on Link 1; (b) throughput on Link 2.



(a)



(b)

Fig. 10. Experiment 4: 20  $\mu$ s latency, with guarantees  $g_1 = 20\%$ ,  $g_2 = 50\%$ ,  $g_3 = 30\%$  (VBR traffic from 0 to 300 ms with rate 10 Mb/s): (a) throughput on Link 1; (b) throughput on Link 2.

## 6. Conclusions

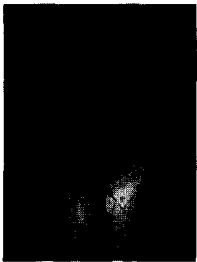
The rate-based scheme for traffic control of ABR connections proposed by the ATM Forum does not consider that ABR connections may have fundamentally different service requirements. In the current traffic management specification, the explicit rate calculated for ABR connections does not distinguish among different types of ABR connections. In this paper we have proposed a solution to this problem by presenting a multi-level flow control scheme for ABR connections in an ATM network. Our proposed scheme uses the notion of *traffic classes* and assigns each 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 flow control scheme. We have discussed how to implement multi-level flow 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 flow control simultaneously at multiple levels and satisfy the control objectives at every stage.

## References

- [1] I.F. Akyildiz, J. Liebeherr and D. Sarkar, Bandwidth regulation of real-time traffic classes in internetworks, *Comput. Networks ISDN Systems* **28** (1996) 855–872.
- [2] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed., Prentice-Hall, Englewood, NJ (1992).
- [3] F. Bonomi and W. Fendick, The rate-based flow control framework for the available bit rate ATM service, *IEEE Network* **9** (2) (1995) 25–39.
- [4] A. Charny, D.D. Clark and R. Jain, Congestion control with explicit rate indication, ATM Forum/94-0692, July 1994.
- [5] A. Charny, K.K. Ramakrishnan and A.G. Lauck, Scalability issues for distribute explicit rate allocation in ATM networks, *Proc. IEEE Infocom'96*, 1996, pp. 1198–1205.
- [6] T.M. Chen, S.S. Liu and V.K. Samalam, The available bit rate service for data in ATM networks, *IEEE Comm. Magazine* **5** (34) (1996) 56–71.
- [7] D.-M. Chiu and R. Jain, Analysis of the increase and decrease algorithms for congestion avoidance in computer networks, *Comput. Networks ISDN Systems* **17** (1989) 1–14.
- [8] M.G. Hluchyj et al., Closed-loop rate-based traffic management, ATM Forum/94-0211R3, April 1994.
- [9] Y. Gong and I.F. Akyildiz, Dynamic traffic control using feedback and traffic prediction in ATM networks, *Proc. IEEE Infocom'94*, 1994, pp. 91–99.
- [10] J.M. Jaffe, Bottleneck flow control, *IEEE Trans. Comm.* **29** (1981) 954–962.
- [11] J. Jain, Congestion control and traffic management in ATM networks: Recent advances and a survey, Manuscript, January 1995.
- [12] R. Jain et al., ERICA +: extensions to ERICA switch algorithm, ATM Forum/95-1346, October 1995.
- [13] R. Jain, S. Kalyanramana and R. Viswanathan, The EPRCA + scheme, ATM Forum/94-0988, September 1994.
- [14] H.T. Kung, T. Blackwell and A. Chapman, Credit-based flow control for ATM networks: Credit update protocol, adaptive credit allocation and statistical multiplexing, *Proc. ACM SIGCOMM'94*, 1994, pp. 101–114.
- [15] H.T. Kung and R. Morris, Credit-based flow control for ATM networks, *IEEE Network* **9** (2) (1995) 40–48.
- [16] J. Liebeherr, I.F. Akyildiz and A. Tai, A multi-level explicit rate control scheme for ABR traffic with heterogeneous service requirements, *Proc. 16th IEEE Int. Conf. on Distributed Computer Systems* (1996).



- [17] P. Newman, Backward explicit congestion notification for ATM local area networks, *Proc. IEEE Globecom'93*, 1993, pp. 719–723.
- [18] K.K. Ramakrishnan and P. Newman, Integration of erate and credit schemes for ATM flow control, *IEEE Network* **9** (2) (1995) 49–56.
- [19] L. Roberts, Enhanced PRCA proportional rate-control algorithm, ATM Forum/94-0735R1, August 1994.
- [20] S.S. Sathaye, Draft ATM Forum traffic management specification version 4.0, ATM Forum/95-0013R1, October 1995.
- [21] S.S. Sathaye, Draft ATM Forum traffic management specification version 4.0, ATM Forum/95-0013R11, March 1996.
- [22] C.-Q. Yang and A.V.S. Reddy, A taxonomy for congestion control algorithms in packet switching networks, *IEEE Network* **9** (4) (1995) 34–45.
- [23] N. Yin and M.G. Hluchyj, On closed-loop rate control for ATM cell relay networks, *Proc. IEEE INFOCOM'94*, 1994, pp. 99–108.
- [24] M. Zukerman and S. Chan, Fairness in ATM networks, *Comput. Networks ISDN Systems* **26** (1) (1993) 109–117.

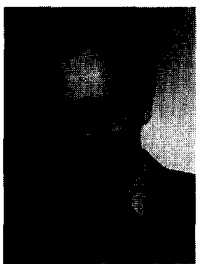


**Ian F. Akyildiz** received his BS, MS, and PhD degrees in Computer Engineering from the University of Erlangen-Nürnberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is a Full Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He has held visiting professorships at the Universidad Tecnica Federico Santa Maria, Chile, Université Pierre et Marie Curie (Paris VI) and Ecole Nationale Supérieure Télécommunications in Paris, France as well as with Universidad Politecnico de Cataluna in Barcelona, Spain. He has published over hundred fifty technical papers in journals and conference proceedings. He is an editor for “IEEE/ACM Transactions on Networking”, “Computer Networks and ISDN Systems Journal”, for “ACM-Baltzer Journal of Wireless Networks”, for “ACM-Springer Journal of Multimedia Systems” and a past editor for “IEEE Transactions on Computers” (1992–1996).

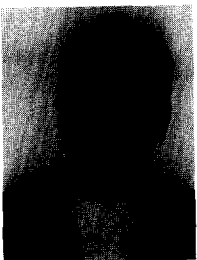
Dr. Akyildiz served as the program chair for the Ninth IEEE Computer Communications Workshop in 1994; and ACM/IEEE Mobicom'96 conference as well as is the program chair for IEEE INFOCOM'98 conference.

Dr. Akyildiz is an IEEE FELLOW, an ACM Fellow, and is a National Lecturer for ACM. He received the “Don Federico Santa Maria Medal” for his services to the Universidad de Federico Santa Maria in Chile. Dr. Akyildiz is listed on “Who's Who in the World (Platinum Edition)”. He received the ACM Outstanding Distinguished Lecturer Award for 1994. Dr. Akyildiz also received the IEEE Leonard G. Abraham Prize Award for best paper published in IEEE Journal of Selected Areas in Communications (JSAC) in 1996.

His current research interests are in ATM Networks, Wireless Networks and Satellite Communication.



**Jörg Liebeherr** received the Diplom-Informatiker degree from the University of Erlangen-Nürnberg, Germany, in 1988 and the Ph.D. degree in Computer Science from the Georgia Institute of Technology in 1991. From 1990 until 1991 he did his dissertation work partly at the IBM T.J. Watson Research Center. In 1992 he was a Postdoc in the Computer Science Division of the University of California, Berkeley. From 1992 until 1997 he was an Assistant Professor in the Department of Computer Science at the University of Virginia. Since September 1997 he is an Associate Professor in the Department of Electrical Engineering at Polytechnic University. He serves on the editorial boards of IEEE Network, Real-Time Systems Journal, and Cluster Computing. His research interests are multimedia networks, broadband networks, real-time systems and performance evaluation. He is a member of IEEE and ACM (Sigcomm and Sigmetrics).



**Ioanis Nikolaidis** is an Assistant Professor at the Computing Science Department of the University of Alberta, Edmonton. He was born in 1967 in Serres, Greece. He received his B.S. in Computer Engineering and Informatics in 1989 from the University of Patras, Greece and his M.S. and Ph.D. in Computer Science from the College of Computing of the Georgia Institute of Technology in 1991 and 1994 respectively. He worked as a co-Op for IBM, Research Triangle Park, North Carolina in the summer of 1991. He was a research scientist with the European Computer Industry Research Center (ECRC) in Munich, Germany, from November 1994 to September 1996. His research interests include performance and modeling of computer and communication systems, parallel simulation techniques, high-speed network protocols and multimedia computing. He is a member of the IEEE and the ACM.