A Simple ABR Switch Algorithm for the Weighted Max-Min Fairness Policy *

Yiwei Thomas Hou,[†] **Henry H.-Y. Tzeng,**[‡] **Shivendra S. Panwar**[§]

Abstract

An important concept in the ABR service model is the minimum cell rate (MCR) guarantee as well as the peak cell rate (PCR) constraint for each ABR virtual connection. Due to the MCR and PCR requirements, the well-known max-min fairness policy is not sufficient to determine the fair rate allocation in the ABR service model. We present the weighted max-min (WMM) fairness policy, which supports both the MCR and PCR requirements for each ABR virtual connection. A centralized algorithm is presented to compute network-wide bandwidth allocation to achieve this policy. Furthermore, a simple ABR algorithm based on the Intelligent Marking technique is developed with the aim of achieving the WMM fairness policy in the distributed ABR environment. The effectiveness of our ABR algorithm is demonstrated by simulation results based on the benchmark network configurations suggested by the ATM Forum.

Key Words: Max-Min Fairness, Minimum Cell Rate, Peak Cell Rate, ABR Service, Centralized and Distributed Algorithms, Traffic Management, Congestion/Flow Control, ATM Networks.

*Part of this work was performed while Y. T. Hou spent the summer of 1996 at Bell Labs, Lucent Technologies, Holmdel, NJ. This work is supported in part by the New York State Center for Advanced Technology in Telecommunications (CATT), Polytechnic University, Brooklyn, NY.

[†]Y. T. Hou is a Ph.D. candidate under the National Science Foundation (NSF) Graduate Research Traineeship Program at Polytechnic University, Brooklyn, NY. Tel. (718) 260-3218, Fax (718) 260-3906, Email: hou@endeavor.poly.edu.

[‡]H. H.-Y. Tzeng is with Bell Labs, Lucent Technologies, Holmdel, NJ.

[§]S. S. Panwar is on the faculty of the Dept. of Electrical Engineering, Polytechnic University, Brooklyn, NY.

1 Introduction

The available bit rate (ABR) service defined by the ATM Forum [1] supports applications that allow the ATM source end system to adjust the information transfer rate based on bandwidth availability in the network. Such applications include LAN interconnect, file transfer and Frame Relay. By the specifications in [1], on the establishment of an ABR connection, the user shall specify to the network both a maximum bandwidth and a minimum required bandwidth, designated as peak cell rate (PCR) and minimum cell rate (MCR), respectively, for the requested connection. The source starts to transmit at an initial cell rate (ICR), which is greater than or equal to MCR, and may adjust its rate up to PCR based on congestion and bandwidth information from the network.

A key performance issue associated with ABR service is fair allocation of network bandwidth for each virtual connection (VC). An intuitive notion of fairness in ABR service is that every VC is entitled to as much network bandwidth as any other VC. In particular, the ATM Forum has adopted the max-min fairness criterion to allocate network bandwidth for ABR connections [2].

Prior efforts to design ABR algorithms to achieve the max-min fair rate allocation, such as [3, 7, 8, 9, 10] did not address the fairness issue in the context of each individual connection's MCR requirement and PCR constraint. Only very simple cases where the MCR is assumed to be negligible and the PCR is assumed to be the link rate are considered. For connections with MCR/PCR constraints, a new fair rate allocation policy is required.

In this paper, we present the weighted max-min (WMM) fairness policy with MCR/PCR support for each individual connection. Here, we assign the weight of each session to be its MCR,[1] which is likely to be used as the billing criterion for ABR service by network providers. This policy was informally described in [5, 14] for the simple single node case without the PCR constraint. In this paper, we further extend this policy to include the PCR constraint. We also present a centralized bandwidth assignment algorithm to achieve the WMM fairness policy in any network topology with an arbitrary number of virtual connections and prove its correctness.

Even though a formal definition and a centralized algorithm for WMM fairness policy are essential for our understanding of how this policy works in allocating network bandwidth for each VC, the practical significance of this policy would be limited if we cannot come up with a distributed ABR algorithm to achieve this policy. Therefore, we develop a distributed ABR algorithm consistent with the ATM Forum ABR Traffic Management framework to achieve the WMM fairness policy. Our ABR algorithm is based on the *Intelligent Marking* technique by Siu and Tzeng [11, 12], which achieves the max-min fair rate allocation policy with no MCR or PCR constraints. We extend this technique to design an ABR algorithm to achieve the WMM policy. The effectiveness of our ABR algorithm is demonstrated by simulation results based on benchmark network configurations suggested by the ATM Forum.

[1] Strictly speaking, this policy does not work for a session with zero MCR. But in high speed ATM network environment, the assignment of a small MCR value to a VC shall not pose any fundamental technical difficulty. Therefore, we assume a nonzero MCR for each session throughout our paper.

2 The Weighted Max-Min Fairness Policy

In our model, a network $\mathcal{N}$ is characterized by a set of links $\mathcal{L}$ and sessions $\mathcal{S}$.[2] Each session $s \in \mathcal{S}$ traverses one or more links in $\mathcal{L}$ and is allocated a specific rate r_s. The (aggregate) allocated rate F_ℓ on link $\ell \in \mathcal{L}$ of the network is

$$F_\ell = \sum_{s \in \mathcal{S} \text{ traversing link } \ell} r_s \,.$$

Let C_ℓ be the capacity of link ℓ. A link ℓ is *saturated* or *fully utilized* if $F_\ell = C_\ell$.

Let MCR_s and PCR_s be the MCR requirement and PCR constraint for session $s \in \mathcal{S}$. For the sake of feasibility, we assume throughout our paper that the sum of the MCRs of connections traversing any link does not exceed that link's capacity. That is,

$$\sum_{\text{all } s \in \mathcal{S} \text{ traversing } \ell} \text{MCR}_s \le C_\ell \quad \text{for every } \ell \in \mathcal{L}.$$

This condition is used by admission control at call setup time to determine whether or not to accept a new ABR connection.

We say that a rate vector $r = (\cdots, r_s, \cdots)$ is *ABR-feasible* if the following two constraints are satisfied:

$$\text{MCR}_s \le r_s \le \text{PCR}_s \quad \text{for all } s \in \mathcal{S},$$
$$F_\ell \le C_\ell \quad \text{for all } \ell \in \mathcal{L}.$$

Informally, the weighted max-min fairness policy achieves max-min for the normalized (with respect to each individual connection's MCR) rate vector r. Formally, this policy is defined as follows.

Definition 1 A rate vector r is *weighted max-min (WMM) fair* if it is ABR-feasible, and for each $s \in \mathcal{S}$ and every ABR-feasible rate vector $\hat{r}$ in which $\hat{r}_s > r_s$, there exists some session $t \in \mathcal{S}$ such that $\frac{r_s}{\text{MCR}_s} \ge \frac{r_t}{\text{MCR}_t}$ and $r_t > \hat{r}_t$. □

Since there is a minimum cell rate requirement in WMM fairness policy, we define a new notion of bottleneck link as follows.

Definition 2 Given an ABR-feasible rate vector r, a link $\ell \in \mathcal{L}$ is a *WMM-bottleneck link* with respect to r for a session s traversing ℓ if $F_\ell = C_\ell$ and $\frac{r_s}{\text{MCR}_s} \ge \frac{r_t}{\text{MCR}_t}$ for all sessions t traversing link ℓ. □

It can be shown that the following theorems hold for the WMM fairness policy [4].

Theorem 1 An ABR-feasible rate vector r is WMM fair if and only if each session has either a WMM-bottleneck link with respect to r or a rate assignment equal to its PCR. □

[2] From now on, we shall use the terms "session", "virtual connection", and "connection" interchangeably throughout our paper.

Theorem 2 There exists a unique rate vector that satisfies the WMM fairness policy. □

Based on Theorem 1, we construct the following centralized algorithm to compute the rate allocation for each session in any network $\mathcal{N}$ such that the WMM fairness policy is satisfied.

Algorithm 1 This algorithm describes the iterative steps of rate allocation for each session to achieve the WMM fairness policy.

1. Start the rate allocation of each session with its MCR.
2. Increase the rate of each session at a rate proportional to its MCR until either some link becomes saturated or some session reaches its PCR, whichever comes first.
3. For the sessions that either traverse saturated links or have reached their PCRs, remove such sessions and their associated bandwidth from the network.
4. If there is no session left, the algorithm terminates; otherwise, go back to Step 2 for the remaining sessions and remaining network capacity. □

To show the practical merit of implementing the WMM rate allocation policy for ABR service, we will develop an explicit rate (ER)-based ABR algorithm conforming to the ATM Forum ABR traffic management specifications [1] in the next section.

3 A Simple ABR Implementation

A generic closed-loop rate-based congestion control mechanism for ABR service is shown in Fig. 1. Resource Management (RM) cells are inserted periodically among ATM data cells to convey network congestion and available bandwidth information to the source. RM cells contain important information such as the source's allowed cell rate (ACR) (also called the current cell rate (CCR) in the RM cell's field), minimum cell rate (MCR) requirement, explicit rate (ER), congestion indication (CI) bit and no increase (NI) bit. A transit node and destination end system (DES) may set the ER field, CI and NI bits in RM cells. All RM cells of an ABR virtual connection are turned back towards its source after arriving at the destination. Upon receiving backward RM cells, the source adjusts its cell generation rate accordingly.

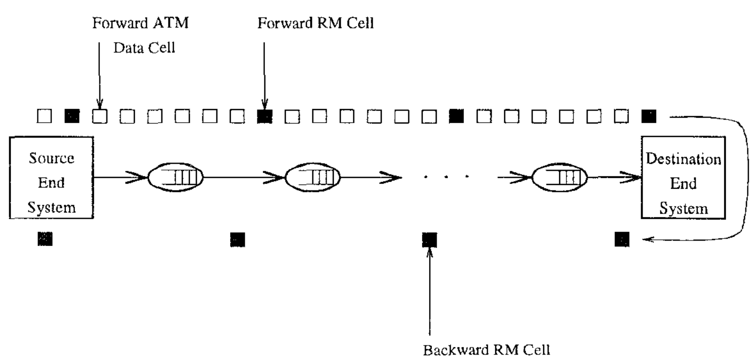

Figure 1: Closed-loop rate-based flow control for an ABR virtual connection.

The ATM Forum Traffic Management Group has specified source and destination behavior. However, the specific ABR switch algorithm is left to the vendors. Switch algorithms can be classified into two categories, namely, binary mode and explicit rate (ER) mode. Binary schemes (e.g. explicit forward congestion indication (EFCI) [13]) rely on a single bit feedback to indicate congestion. Due to limited feedback information about network congestion status, the source only knows that either congestion in the network is present or absent, but doesn't know *how much* to increase or decrease its transmission rate. Therefore, the source's cell rate experiences oscillations. On the other hand, ER schemes employ rate calculation at a switch to estimate available bandwidth and convey this information through the ER field in the returning RM cells. Hence, an ER scheme promises higher efficiency and stability than a binary scheme.

We first specify the source behavior of our ABR algorithm [1].

Algorithm 2 ABR Source Behavior

- Start to transmit at ACR := ICR, which is greater than or equal to MCR;
- An RM(CCR, MCR, ER) cell is inserted for every N_{rm} ATM data cells with CCR := ACR; MCR := MCR; ER := PCR;
- Upon the receipt of a backward RM(CCR, MCR, ER) cell, the ACR at source is adjusted to: ACR := $\max\{\min\{(\text{ACR} + \text{AIR}), \text{ER}, \text{PCR}\}, \text{MCR}\}$. □

The destination end system simply returns every RM cell back towards the source upon receiving it.

Our ABR switch algorithm for WMM fairness policy employs ER calculation and is based on the Intelligent Marking technique for max-min fair rate allocation, originally proposed in [10] and further refined in [11, 12]. The key idea of this technique is to let each congested switch estimate the max-min bottleneck link rate at a link of a switch with a small number of computations and without keeping track of each VC's state information (so called per-VC accounting). Using feedback mechanisms, the ER field of a returning RM cell is set to the minimum of all the estimated bottleneck link rates on all its traversing links to achieve max-min fair share. The details of the Intelligent Marking technique are given in [11].

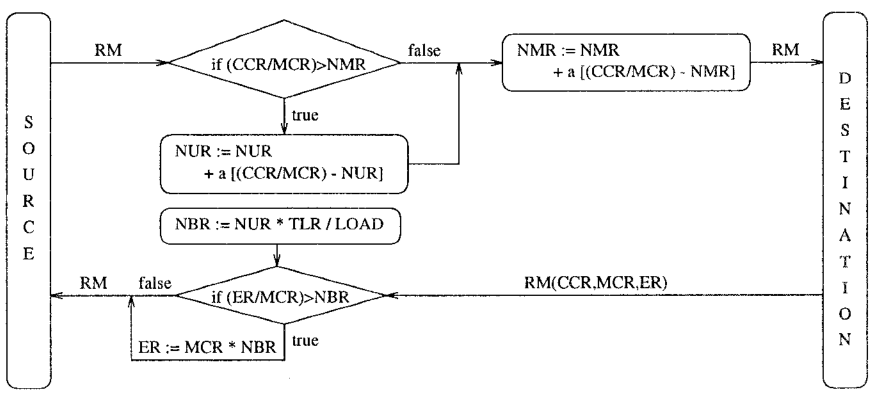

Figure 2: Switch behavior for the WMM fairness policy.

The WMM fairness policy can be regarded as max-min fairness for the *normalized cell rate* (with respect to MCR requirement) for each VC. This motivates us to design our ABR algorithm for WMM fairness policy by letting the normalized cell rates (e.g. CCR/MCR, ER/MCR) from an RM cell to participate in the Intelligent Marking. Fig. 2 illustrates the switch behavior of our ABR algorithm. Four variables named LOAD, NMR (Normalized Mean Rate), NUR (Normalized Upper Rate) and NBR (Normalized Bottleneck Rate) are defined for each output port of an ATM switch. The value of LOAD corresponds to the aggregated cell rate entering the queue normalized with respect to the link rate and is measured by the switch over a period of time. The value of NMR contains an estimated normalized (with respect to MCR_s, $s \in \mathcal{S}$) average rate for all VCs traversing this link; the value of NUR contains an estimated normalized upper cell rate; and NBR contains an estimated normalized WMM-bottleneck link rate. Here, NMR, NUR and NBR are all dimensionless. TLR is the targeted load ratio and $0 < \alpha < 1$.

Algorithm 3 ABR Switch Behavior

Upon the receipt of RM(CCR, MCR, ER) from the source of a VC
    if (CCR/MCR) > NMR, then
        NUR := NUR + α (CCR/MCR - NUR);
    NMR := NMR + α (CCR/MCR - NMR);
    Forward RM(CCR, MCR, ER) to its destination;

Upon the receipt of RM(CCR, MCR, ER) from the destination of a VC
    NBR := NUR * TLR / LOAD;
    if (QS > QT),[3] then
        NBR := (QT / QS) * NBR;
    if (ER/MCR) > NBR, then
        ER := MCR $\times$ NBR;
        Forward RM(CCR, MCR, ER) to its source;
    else
        Forward RM(CCR, MCR, ER) to its source. □

4 Simulation Results

Here we present a simulation study demonstrating the effectiveness of our ABR algorithm in achieving the WMM fairness policy.

The ATM switches in all the simulations are assumed to have output buffers with a speedup equal to the number of their ports. The buffer of each output port of a switch employs the simple FIFO queuing discipline and is shared by all VCs going through that port. At each output port of an ATM switch, we implement the switch algorithm for the WMM fairness policy. In all of our simulations, we assume persistent sources, i.e., all sources will attempt to transmit cells at their maximum allowable cell rates.

Table 1 lists the parameters used in our simulation. The distance from source/destination to the switch is 100 m and the link distance between ATM switches is 10 km (this corresponds to a LAN environment).

The Peer-to-Peer Network Configuration

In this network configuration (Fig. 3), the output port link of SW1 (Link 12) is the only bottleneck link for all VC sessions. Assume that all links are of unit capacity. The MCR

[3] QS is the queue size at a link and QT is a predefined queue threshold.

End System	PCR	PCR
	MCR	MCR
	ICR	MCR
	Nrm	32
	AIR	3.39 Mbps
Link	Speed	150 Mbps
Switch	Cell Switching Delay	4 μSec
	α	0.125
	Queue Threshold (QT)	50 cells
	Output Buffer Size	2000 cells

Table 1: Simulation parameters.

requirement, the PCR constraint and the WMM rate allocation for each session are listed in Table 2.

Fig. 4 shows the ACR at source for sessions s1, s2 and s3, respectively. The cell rates shown in the plot are normalized with respect to the link rate (150 Mbps) for easy comparison with those values obtained with our centralized algorithm under unit link capacity (Table 2). After the initial transient period, we see that the cell rates of each VC match fairly well with the rates listed in Table 2. To study the network utilization of our ABR algorithm, we also show the inter-switch link utilization (Link 12) and the queue size of the congested switch (SW1) in Fig. 5, we find that the link is 100% utilized with reasonably small buffer requirements.

The Parking Lot Network Configuration

The name of this configuration is derived from theater parking lots, which consist of several parking areas connected via a single exit path [6]. The specific parking lot configuration that we use is shown in Fig. 6 where VC sessions $s1$ and $s2$ start from the first switch and go to the last switch. $s3$ and $s4$ start from SW2 and SW3, respectively, and terminate at the last switch.

Table 3 lists the MCR requirement, the PCR constraint, and the WMM rate allocation for each session.

Fig. 7 shows the normalized cell rate of each VC session under our ABR algorithm. We see that they match well with the rates listed in Table 3, which are obtained through the WMM centralized algorithm. Fig. 8 shows the link utilizations of Link34 and the output port buffer occupancy of SW3 for the same simulation run. Again, the congested link is 100% utilized with low buffer occupancy.

Our simulation results show that the rate allocation by our simple ABR algorithm matches closely with the centralized WMM fairness policy in a LAN environment. For a wide area network, a heuristic algorithm such as ours usually requires careful system parameter tuning, and a more sophisticated ABR switch algorithm requiring per-VC accounting such as [4] may be necessary. But in a LAN environment, where implementation cost may well be the most important criterion in the choice of an ABR switch algorithm, our algorithm offers satisfactory performance with minimum implementation complexity.

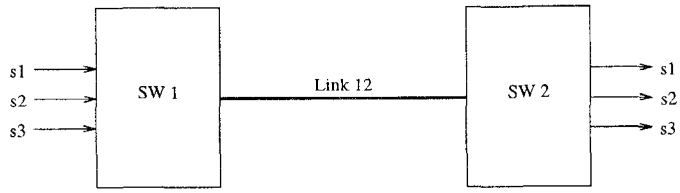

Figure 3: The peer-to-peer network configuration.

Session	MCR	PCR	WMM Rate Allocation
$s1$	0.15	1.00	0.525
$s2$	0.10	0.30	0.30
$s3$	0.05	0.50	0.175

Table 2: MCR requirement, PCR constraint, and the WMM rate allocation for each session for the peer-to-peer network configuration.

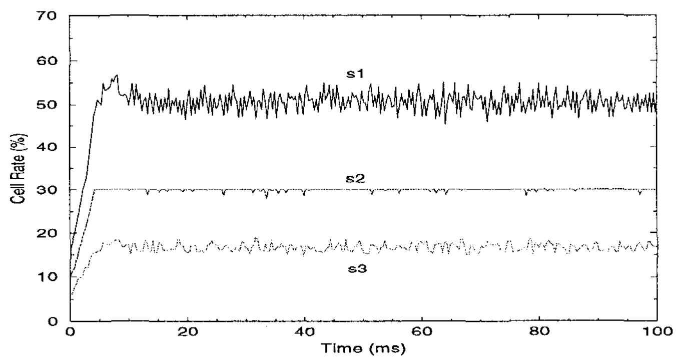

Figure 4: The cell rates of all connections for the WMM fairness policy in the peer-to-peer network configuration.

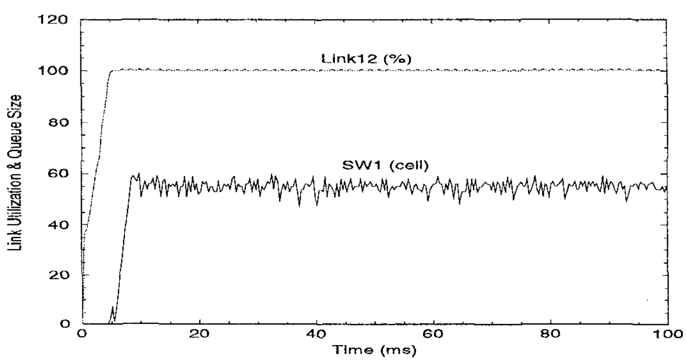

Figure 5: The link utilization and the queue size of the congested switch for the WMM fairness policy in the peer-to-peer network configuration.

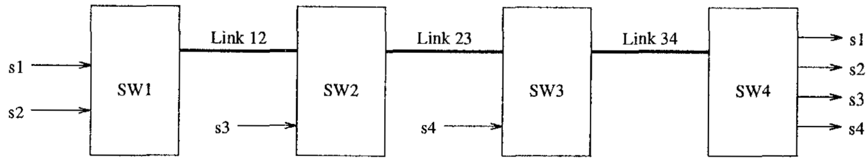

Figure 6: The parking lot network configuration.

Session	MCR	PCR	WMM Rate Allocation
$s1$	0.15	0.35	0.35
$s2$	0.10	0.20	0.20
$s3$	0.10	0.50	0.30
$s4$	0.05	0.50	0.15

Table 3: MCR requirement, PCR constraint and the WMM rate allocation for each session for the parking lot network configuration.

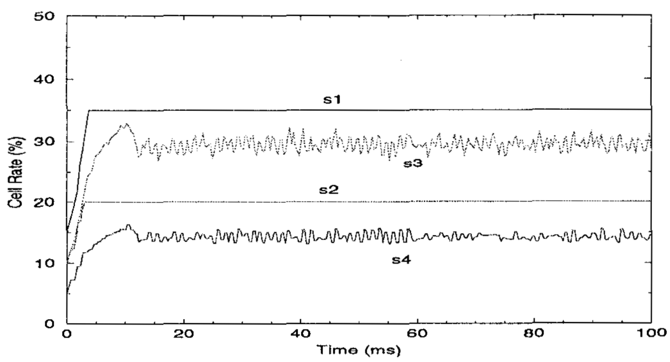

Figure 7: The cell rates of all connections for the WMM fairness policy in the parking lot network configuration.

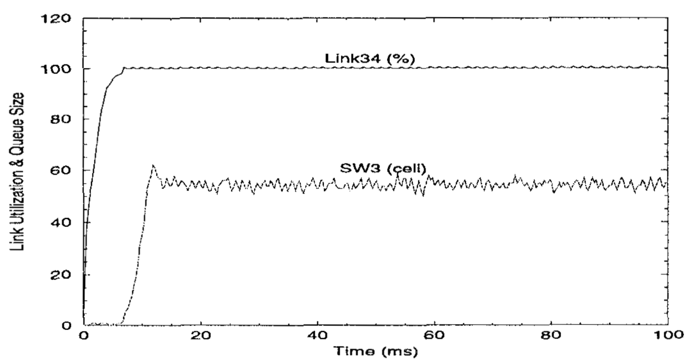

Figure 8: The link utilization and the queue size of the congested switch (SW3) for the WMM policy in the parking lot network configuration.

References

[1] ATM Forum Technical Committee, "Traffic Management Specification - Version 4.0," *ATM Forum/95-0013R13,* February 1996.

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

[3] A. Charny, D. Clark, and R. Jain, "Congestion Control with Explicit Rate Indication," *Proc. IEEE ICC'95,* pp. 1954-1963.

[4] Y. T. Hou, H. Tzeng, and S. S. Panwar, "The Weighted Max-Min Fair Rate Allocation for ABR Service: Theory and Implementation," *Tech. Report CATT 97-110,* Center for Advanced Technology in Telecommunications, Polytechnic University, Brooklyn, NY, March 17, 1997.

[5] D. Hughes, "Fair Share in the Context of MCR," *ATM Forum Contribution 94-0977,* October 1994.

[6] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," *Computer Networks and ISDN Systems,* February 1995.

[7] R. Jain, *et al.*, "ERICA Switch Algorithm: A Complete Description," *ATM Forum Contribution, 96-1172,* August 1996.

[8] L. Kalampoukas, A. Varma, and K. K. Ramakrishnan, "Dynamics of an Explicit Rate Allocation Algorithm for Available Bit Rate (ABR) Service in ATM Networks," *Proc. 6th IFIP Int. Conf. High Performance Networking (HPN '95),* Sept. 1995, pp.143-154.

[9] L. Roberts, "Enhanced PRCA (Proportional Rate Control Algorithm)," *ATM Forum Contribution 94-0735R1,* August 1994.

[10] K.-Y. Siu and H.-Y. Tzeng, "Intelligent Congestion Control for ABR Service in ATM Networks," *ACM SIGCOMM Computer Communication Review*, 24(5):81–106, October 1994.

[11] K.-Y. Siu and H.-Y. Tzeng, "Limits of Performance in Rate-based Control Schemes," *ATM Forum Contribution 94-1077,* November 1994.

[12] H.-Y. Tzeng and K.-Y. Siu, "Comparison of Performance Among Existing Rate Control Schemes," *ATM Forum Contribution 94-1078,* November 1994.

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

[14] N. Yin, "Max-Min Fairness vs. MCR Guarantee on Bandwidth Allocation for ABR," *Proc. IEEE ATM'96 Workshop,* San Francisco, CA, August 25-27, 1996.