

# Max-Min 公平性をみたく効率的な ER スイッチアルゴリズムの提案

大崎 博之 村田 正幸 宮原 秀夫

大阪大学大学院 基礎工学研究科 情報数理系専攻  
〒 560 大阪府豊中市待兼山町 1-3

(Phone) +81-6-850-6588

(Fax) +81-6-850-6589

(E-mail) oosaki@ics.es.osaka-u.ac.jp

あらまし レート制御方式は、網からのフィードバック情報によって、送信側端末のセル送出レートを動的に調節する。レート制御方式は、ABR (Available Bit Rate) サービスクラスに適用される輻輳制御方式として、ATM フォーラムにおいて標準化が行なわれた。スイッチのアルゴリズムは、送信側への輻輳通知方法の違いにより、バイナリスイッチと ER スイッチに分類される。最近、Max-Min 公平性をみたく ER スイッチアルゴリズムとして、Max-Min 方式が提案された。そこで本稿では、Max-Min 公平性をみたくといった Max-Min 方式の利点を残しながらも、スイッチのキュー長を安定させることによって、セルの廃棄を防ぎ、リンクの利用率を高める新たな ER スイッチアルゴリズムを提案する。さらに、提案したスイッチアルゴリズムの有効性をシミュレーション手法によって検証する。

和文キーワード ABR サービスクラス、レート制御方式、ER スイッチアルゴリズム、Max-Min 公平性

## Efficient Explicit-Rate Switch Algorithm with Max-Min Fairness for ABR Service Class in ATM Networks

Hiroyuki Ohsaki Masayuki Murata Hideo Miyahara

Department of Informatics and Mathematical Science  
Graduate School of Engineering Science, Osaka University  
1-3 Machikaneyama, Toyonaka, Osaka 560, Japan

(Phone) +81-6-850-6588

(Fax) +81-6-850-6589

(E-mail) oosaki@ics.es.osaka-u.ac.jp

**Abstract** A rate-based congestion control algorithm regulates cell emission rate of source end systems based on feedback information from the network. It has been standardized by the ATM Forum for application to an ABR (Available Bit Rate) service class. In the standard, two types of congestion notification methods of the switch are specified: EFCI marking and explicit-rate marking. In this paper, we focus on explicit-rate marking switch. We propose our enhancements on a recently proposed switch algorithm called as the max-min scheme. The main objective of our enhancements is to control the queue length of the switch for preventing cell loss and achieving full link-utilization. We show effectiveness of our switch algorithm by simulation experiments.

**英文 key words** ABR Service Class, Rate-Based Congestion Control Algorithm, Explicit-Rate Switch, Max-Min Fairness

## 1 Introduction

A rate-based congestion control algorithm is a closed-loop control method suitable for data transfer applications. In the rate-based congestion control algorithm, cell transmission rates of source end systems are regulated according to congestion information returned by the network. The ATM Forum has adopted it as the congestion control mechanism for the ABR (Available Bit Rate) service class, and has finished its standardization [1]. In the standard, behavior of source and destination end systems (i.e., terminals) are specified in detail. Congestion notification methods from the network (i.e., ATM switches) to source end systems are also specified. The source end system periodically sends a forward RM (Resource Management) cell per  $N_{RM}$  data cells, and the destination end system sends it back to the corresponding source end system as a backward RM cell. The switch notifies its congestion to source end systems by marking an EFCI (Explicit Forward Congestion Indication) bit of data cells or a CI (Congestion Indication) bit of RM cells. Since it uses one-bit information, the switch utilizing the EFCI bit or the CI bit is often referred to as a *binary-mode* switch. In the standard, the switch is allowed to explicitly designate the cell transmission rate by modifying an ER (Explicit Rate) value of the RM cell. This sort of switch is called as *explicit-rate switch*.

One advantage of the binary-mode switch is its implementation simplicity. The binary-mode switch requires less hardware so that it can be implemented at low cost. However, one of its disadvantages is sensitivity to control parameters such as  $RDF$  (Rate Decrease Factor) and  $RIF$  (Rate Increase Factor). If these control parameters are chosen inappropriately, the rate-based congestion control shows poor performance [2]. Although the binary-mode switch with appropriate control parameters works effectively in LAN environments, performance improvement is limited in WAN environments.

While its implementation is rather complex, the explicit-rate switch has a potential to obtain much better performance than the binary-mode switch. A typical operation of the explicit-rate switch is to compute an appropriate bandwidth allocation for every connection based on, for example, the bandwidth available to ABR connections and the degree of congestion. The switch then updates the ER value of forward and/or backward RM cells as

$$ER \leftarrow \min(FS, ER).$$

In the above equation,  $ER$  is the ER value in the RM cell written by some other switch, and  $FS$  is a computed bandwidth allocation for the connection. When the source end system receives the backward RM cell, it updates its  $ACR$  (Allowed Cell Rate) as

$$ACR \leftarrow \min(ACR + PCR \times RIF, PCR, ER) \quad (1)$$

Thus, bandwidth allocation for all connections can be finished within one round-trip time only if  $RIF$  is set to be a large value; that is, if  $RIF$  is small, the source end system needs more RM cells to increase its  $ACR$  to  $ER$ . The brightness of the above equation is that the source end system does not necessarily know the switch type (i.e., binary-mode or explicit-rate switch). In other words, an effectiveness of explicit-rate switches is fully dependent on the determination method of the ER value.

In the ATM Forum, several switch algorithms with explicit-rate marking have been proposed through standardization process of the rate-based congestion control algorithm [1, 3]. These include EPRCA (Enhanced Proportional Rate Control Algorithm) [4], CAPC (Congestion Avoidance using Proportional Rate Control) [5], APRC2 (Adaptive Proportional Rate Control) [6] and ERICA (Explicit Rate Indication for Congestion Avoidance) [7]. Each algorithm has its own advantages and disadvantages in terms of, for example, effectiveness, robustness, fairness and configuration simplicity. We first summarize a recently proposed switch algorithm called as the max-min scheme [8]. A strong point of this algorithm compared with others is that it can satisfy *max-min fairness* for any network configuration; that is, total throughput of the network is maximized and fairness among connections is maintained [9]. However, its defect is in lack of adaptability to changes in the network (e.g., connection addition/disconnection) as will be demonstrated in Section 3. Thus, we propose our enhancements to the max-min scheme to improve its stability and efficiency. We also evaluate its performance by comparing with other explicit-rate switch algorithms.

The rest of this paper is organized as follows. In Section 2, we introduce the max-min scheme and propose our enhancements. Section 3 is devoted to performance evaluation of explicit-rate switch algorithms. Finally, in Section 4, we conclude our paper with a few remarks.

## 2 Designing Explicit Rate Switch Algorithm

We start this section with an introduction of the max-min scheme proposed by Tsang *et. al* in [8] with reviewing its advantages and disadvantages. We next propose our enhancements to the max-min scheme, and explain how the defects of the original max-min scheme are resolved.

### 2.1 Max-Min Scheme

The max-min scheme always maintains an information table at the switch. Two entries are maintained for each connection. An entry of the table is listed in Table 1. In this table,  $VCI$  corresponds to the VC identifier of the connection.  $ER_F$  and  $ER_B$  remember ER values written in the latest forward and backward RM cells, respectively.  $CA$  is the current bandwidth allocation to this connection, and a *constrained* flag indicates whether this connection is con-

Name	$VCI$	$ER_F$	$ER_B$	$CA$	constrained
Type	integer	float	float	float	boolean

Table 1: Information table at the switch.

strained or not by other switches; if this flag is true, it means that this connection cannot achieve its fair share of the bandwidth at the switch. The constrained flag is used to allocate bandwidth according to the max-min fairness. At every receipt of forward and backward RM cells, the switch updates the associated entries and recomputes the bandwidth allocation for the connection as follows.

Suppose that the switch receives a forward RM cell. The switch first checks whether the ER value in the RM cell is different from  $ER_F$ . If different, it implies that the bandwidth allocation for this connection has been changed at other switches, and that the bandwidth allocation should be recomputed. Hence, the switch replaces  $ER_F$  with the ER value in the RM cell, and updates the constrained flag by comparing  $ER_F$  with the allocated bandwidth  $CA$ . Then, the following calculation of the bandwidth allocation is performed.

Let  $FS$  be the fair share of the bandwidth for unconstrained connections (i.e., the *constrained* flag is false).  $FS$  is computed as

$$FS = \frac{ABW - \sum_{n \in G_C} CA_n}{|G_U|}, \quad (2)$$

where  $ABW$  is the available bandwidth to the ABR service class, and  $CA_n$  is  $CA$  of the  $n$ th connection.  $G_C$  and  $G_U$  are sets of constrained and unconstrained connections, respectively.  $|G_U|$  represents the number of unconstrained connections. The switch updates the *constrained* flag of each connection for  $FS$ , and assigns  $FS$  to  $CA$  of unconstrained connections. Namely, the *constrained* flag and  $CA$  are determined as

$$\text{constrained} = \begin{cases} \text{true,} & FS \geq \min(ER_F, ER_B) \\ \text{false,} & FS < \min(ER_F, ER_B) \end{cases}, \quad (3)$$

and

$$CA = \begin{cases} \min(ER_F, ER_B), & \text{if constrained} \\ FS, & \text{otherwise} \end{cases}. \quad (4)$$

The above process is repeated until there is no change in *constrained* flags. Finally, the ER value of the RM cell is updated as

$$ER \leftarrow CA. \quad (5)$$

Refer to [8] for more detail.

## 2.2 Our Enhancements to Max-Min Scheme

In this subsection, we propose enhancements to the max-min scheme. The objective of our enhancements is to eliminate

defects of the max-min scheme without losing its advantages. Advantages of our enhanced max-min scheme over the original max-min scheme are: (1) controllability of the queue length, (2) an effective *TBE* (Transient Buffer Exposure) [1] allocation mechanism, (3) robustness against background traffic, (4) fairness achievement incorporating *PCR* and *MCR*, and (5) interoperability. Details of our enhancements are described below.

The first enhancement is to control the queue length to a desired level. This mechanism is intended to prevent cell loss and to achieve full link-utilization as well as small cell delay. In our enhanced max-min scheme, the switch allocates the bandwidth to connections according to the current queue length. More strictly, the allocation of the ER value in Eq. (5) is changed as

$$ER \leftarrow CA \times z(Q(t)),$$

where  $z(x)$  is a bandwidth adjustment function, and  $Q(t)$  is a current queue length. The bandwidth adjustment function,  $z(x)$ , is a monotonically decreasing function having the following characteristics.

$$z(x) = \begin{cases} 1 + \Delta_1, & x = 0 \\ 1, & x = Q_T \end{cases}, \quad (6)$$

and

$$1 - \Delta_2 \leq z(x) \leq 1 + \Delta_1$$

$Q_T$  is a threshold value at the switch used to control the queue length.  $\Delta_1$  and  $\Delta_2$  are upper and lower bandwidth adjustment factors. For example, when the queue length is zero, the switch allocates  $(1 + \Delta_1)$  times larger bandwidth than the available bandwidth of the ABR service class. On the other hand, when the queue length is greater than  $Q_T$ , the switch reduces the bandwidth allocation. By introducing this mechanism, the queue length is managed to be kept at  $Q_T$ . Namely, if the queue length is below  $Q_T$ , the switch tries to increase its queue length by allocating more bandwidth. If the queue length is over  $Q_T$ , the switch tries to decrease its queue length. Hence, the queue length is restored at  $Q_T$  even when the switch gets overloaded or underloaded.

The second enhancement for the max-min scheme is to support various fairness definitions with *PCR* and *MCR*. To take account of *PCR* and *MCR*, the equation for computing the fair share, Eq. (2), is further extended as

$$FS_n = \alpha \times MCR_n + \beta \times \left\{ (ABW - \sum_{n \in G_C} CA_n) - \alpha \times \sum_{n \in G_U} MCR_n \right\},$$

where  $\alpha$  and  $\beta$  are given as follows.

- Scheme 1: Max-Min Share

$$\alpha = 0 \text{ (or 1)} \quad (7a)$$

$$\beta = \frac{1}{N_{VC}} \quad (7b)$$

This scheme is similar to the max-min fairness criterion (equivalent if  $\alpha = 0$ ). Namely, the bandwidth is allocated equally to all connections regardless of their  $PCRs$  and  $MCRs$ .

- Scheme 2: Weighted Share with  $MCR$

$$\alpha = 0 \text{ (or 1)} \quad (8a)$$

$$\beta = \frac{MCR_n}{\sum_i MCR_i} \quad (8b)$$

This scheme allocates the bandwidth proportional to the connection's  $MCR$ ; that is, the connection with larger  $MCR$  can obtain more bandwidth than other connections. Note that this scheme cannot be applied when there is a connection with  $MCR_n = 0$ .

- Scheme 3: Weighted Share with  $PCR$

$$\alpha = 0 \text{ (or 1)} \quad (9a)$$

$$\beta = \frac{PCR_n}{\sum_i PCR_i} \quad (9b)$$

This scheme allocates bandwidth proportional to the connection's  $PCR$ ; that is, the connection with larger  $PCR$  can obtain more bandwidth than other connections.

- Scheme 4: Weighted Share with  $MCR$  and  $PCR$

$$\alpha = 0 \text{ (or 1)} \quad (10a)$$

$$\beta = \left( \frac{MCR_n}{\sum_i MCR_i} \right)^\gamma \times \left( \frac{PCR_n}{\sum_i PCR_i} \right)^\delta \quad (10b)$$

This scheme is a combination of Schemes 2 and 3; it allocates bandwidth according to both  $MCR$  and  $PCR$ . In the above equation,  $\gamma$  and  $\delta$  are weight ratios ( $0 \leq \gamma \leq 1$  and  $0 \leq \delta \leq 1$ ).

Note that Schemes 3 and 4 require an additional capability at the switch for maintaining  $PCR$  values of all connections although  $MCR$  values of all connections are stored in the RM cell.

In our enhanced max-min scheme, the available bandwidth to the ABR service class is computed at the switch by monitoring the number of arriving CBR and VBR cells within a fixed interval. More specifically, by letting  $I$  be the bandwidth monitoring interval and  $N$  be the number of CBR and VBR cells received during  $I$ , the available bandwidth  $ABW$  is computed as

$$ABW = BW - \frac{N}{I}.$$

We next explain our mechanism to allocate  $TBE$  for a new connection. Let us assume that there are  $N_{VC}$  active connections on the link, and  $(N_{VC} + 1)$ th connection starts cell emission at  $t = t_0$ . At the connection setup time, the switch determines  $TBE$  for this connection as

$$TBE = \min(RTT \times PCR, \\ BL - \max(Q(t), Q_T) - \sum_{n=1}^{N_{VC}} R_n),$$

where  $R_n$  is a reserved buffer capacity for  $n$ th connection, and  $BL$  is the buffer size at the switch.  $RTT$  is an estimated round-trip delay of the RM cell including processing delays, which is signaled at connection setup [1]. The buffer reservation,  $R_n$ , is valid until the source end system receives the first backward RM cell from the network; that is,  $R_n$  is canceled at  $t = t_0 + RTT$ . Thus, the buffer reservation for  $(N_{VC} + 1)$ th connection is given by

$$R_{N_{VC}+1} = \begin{cases} TBE, & t_0 \leq t \leq t_0 + RTT \\ 0, & t_0 + RTT < t \end{cases}$$

Given  $TBE$  from the network, the source end system computes  $ICR$  (Initial Cell Rate) as (see [1])

$$ICR \leftarrow \min(ICR, \frac{TBE}{RTT}).$$

By employing the  $ICR$  negotiation mechanism, buffer overflow caused by activation of a new ABR connection can be completely avoided. Another possibility of buffer overflow is when background traffic suddenly increases its bandwidth requirements. In what follows, we investigate an appropriate setting of control parameters satisfying two objectives: preventing cell loss and achieving full link utilization.

From now on, we analyze the maximum and minimum of the queue length by assuming infinite buffer capacity. To analyze the worst case, we assume that all connections are not constrained at other switches, and that all source end systems always have cells to transmit. We further assume that the network is in steady-state; the queue length is equal to  $Q_T$  because of the queue control mechanism of our enhanced max-min scheme. Let  $N_{VC}$  denote the number of active connections. We introduce  $\tau_{sx_n}$  and  $\tau_{xd_n}$  ( $1 \leq n \leq N_{VC}$ ) as the propagation delays between the  $n$ th source end system and the switch, and between the switch and the corresponding destination end system. The bandwidth of the link is denoted by  $BW$ .

When the amount of the background traffic is increased from  $C$  to  $C'$  ( $C' > C$ ) at  $t = t_0$ , the switch immediately recomputes new bandwidth allocations and notifies them to source end systems via the ER values of RM cells. In this case, the bandwidth allocation for each connection is changed from  $(BW - C)/N_{VC}$  to  $(BW - C')/N_{VC}$ . Since the RM cell containing a new explicit-rate arrives at the  $n$ th source end system  $\tau_{sx_n}$  after the arrival rate of the background traffic is changed, cells are excessively injected into the network. Thus, the envelope of the queue length is given by

$$Q(t) = Q_T \\ + \int_{t_0}^t \left( \sum_{n=1}^{N_{VC}} AC R_n(t - \tau_{sx_n}) - (BW - C') \right) dx, \quad (11)$$

where  $AC R_n(t)$  is the bandwidth allocated for the  $n$ th connection. The backward RM cell having the new bandwidth

allocation of  $(BW - C')/N_{RM}$  are received by the  $n$ th source at  $t = t_0 + \tau_{sxn} + t_{RM}$ , where  $\tau_{sxn}$  is the propagation delay from the switch to the source end system and  $t_{RM}$  is a delay for the next RM cell at the switch. Thus,  $ACR_n(t)$  is given by

$$ACR_n(t) = \begin{cases} \frac{BW-C}{N_{VC}}, & t \leq t_0 + \tau_{sxn} + t_{RM} \\ \frac{BW-C'}{N_{VC}}, & t > t_0 + \tau_{sxn} + t_{RM} \end{cases},$$

and

$$\begin{aligned} t_{RM} &\leq \frac{N_{RM}}{ACR_n(t - \tau_{sxn} - 2\tau_{xdn})} \\ &= \frac{N_{RM} \times N_{VC}}{BW - C}. \end{aligned}$$

The maximum queue length,  $Q_{max}$ , is obtained as

$$\begin{aligned} Q_{max} &= Q_T + \sum_{n=1}^{N_{VC}} \left( \frac{C' - C}{N_{VC}} \times (2\tau_{sxn} + t_{RM}) \right) \\ &\leq Q_T + (C' - C) \\ &\quad \times \left( 2 \times \max_n(\tau_{sxn}) + \frac{N_{RM} \times N_{VC}}{BW - C} \right) \quad (12) \end{aligned}$$

Hence, to prevent buffer overflow,  $Q_T$  should be chosen to satisfy the following relation.

$$Q_{max} \leq BL$$

The queue decreases when the amount of background traffic is decreased. When the amount of background traffic is changed from  $C$  to  $C''$  ( $C'' < C$ ) at  $t = t_0$ , the envelope of the queue length is simply given by replacing  $C'$  in Eq. (11) with  $C''$ . As with the previous case, the minimum queue length is given by

$$\begin{aligned} Q_{min} &= Q_T + \sum_{n=1}^{N_{VC}} \left( \frac{C'' - C}{N_{VC}} \times (2\tau_{sxn} + t_{RM}) \right) \\ &\leq Q_T + (C'' - C) \\ &\quad \times \left( 2 \times \max_n(\tau_{sxn}) + \frac{N_{RM} \times N_{VC}}{BW - C} \right) \quad (13) \end{aligned}$$

Thus, full link utilization can be achieved by setting  $Q_T$  to satisfy the following relation.

$$Q_{min} \geq 0$$

In our enhanced max-min scheme, three control parameters —  $Q_T$ ,  $\Delta_1$ ,  $\Delta_2$  and  $I$  — are newly adopted for fulfilling high performance in exchange for configuration simplicity. However, the threshold value,  $Q_T$ , can be configured according to the above analysis.

In the original max-min scheme, the destination end system must reset the ER value in the RM cell to  $PCR$ . It requires an additional hardware to maintain  $PCR$  values of active connections at the destination end system, and does not follow the ATM Forum standard. In our enhanced max-min scheme, such a mechanism is not required; the destination end system simply sends back the RM cell.

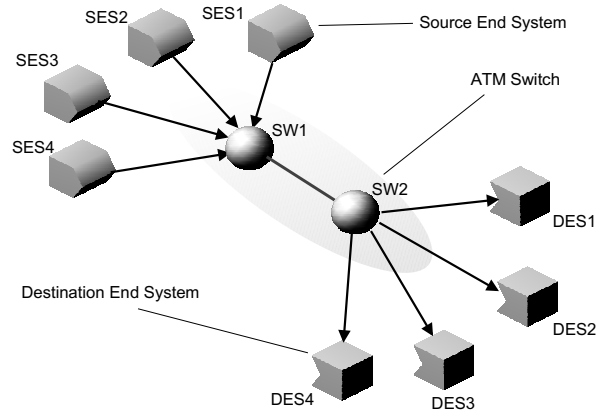


Figure 1: Our Simulation Model.

### 3 Performance Evaluation

#### 3.1 Simulation Model

Figure 1 shows our simulation model, which consists of two inter-connected explicit-rate switches and four ABR connections with identical propagation delays. In the following simulation, the link bandwidth,  $BW$ , is fixed at 353.7 cell/ms assuming a 150 Mbit/s link. The propagation delay of each link (source–switch, switch–switch and switch–destination) is fixed at an identical value denoted by  $\tau$ . A round-trip delay between source and destination end systems is, therefore,  $6 \times \tau$ . We use two values of  $\tau$ : 0.01 ms (about 2 km) as LAN environments and 1.00 ms (about 200 km) as WAN environments. Thus, the round-trip delay is 0.06 ms for LAN environments or 6.00 ms for WAN environments.

At each switch, its buffer size,  $BL$ , is set to 300 Kbyte (5,796 cells). We assume persistent sources; all source end systems always have cells to transmit. In other words, we assume that  $CCR$  (Current Cell Rate) of the source end system is always equivalent to  $ACR$ . We summarize values of control parameters at the source end system used in our simulation in Table 2. See [1] for complete description of control parameters.

#### 3.2 Addition and Departure of ABR Connections

In this subsection, we compare three explicit-rate switch algorithms: ERICA, the max-min scheme and our enhanced max-min scheme. The main objective of this section is to evaluate the influence of connection addition and departure. So we add four connections to the network at different starting points,  $t = 0, 20, 40$  and  $60$  ms, and remove them from the network at  $t = 300, 280, 260$  and  $240$  ms, respectively. For comparison purposes, the  $TBE$  determination algorithm in Subsection 2.2 is not used. Instead, we set the initial cell rate,  $ICR$ , to be  $PCR$ .

We first show simulation results for ERICA in Figs. 2 and 3 for different propagation delays,  $\tau = 0.01$  and  $1.00$  ms, respectively. A target utilization and a load averaging inter-

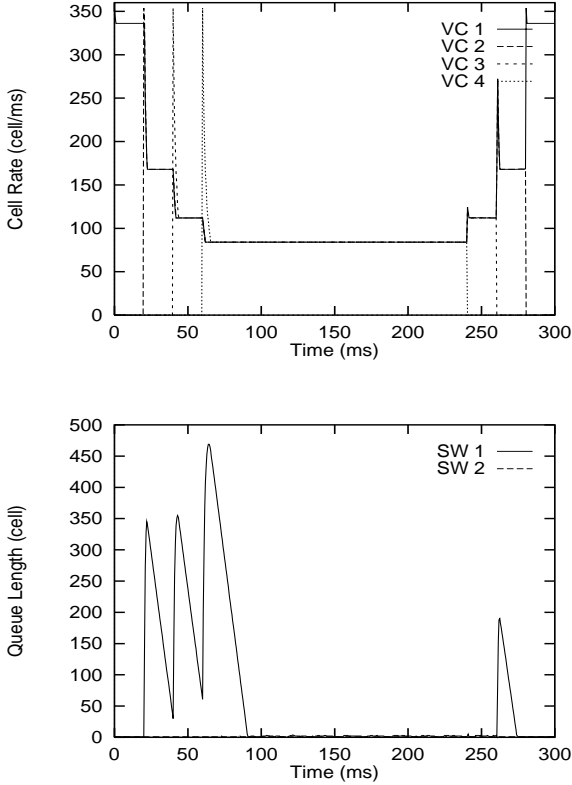


Figure 2: Effect of connection addition/disconnection in ERICA for  $\tau = 0.01$  ms and target utilization of 0.95.

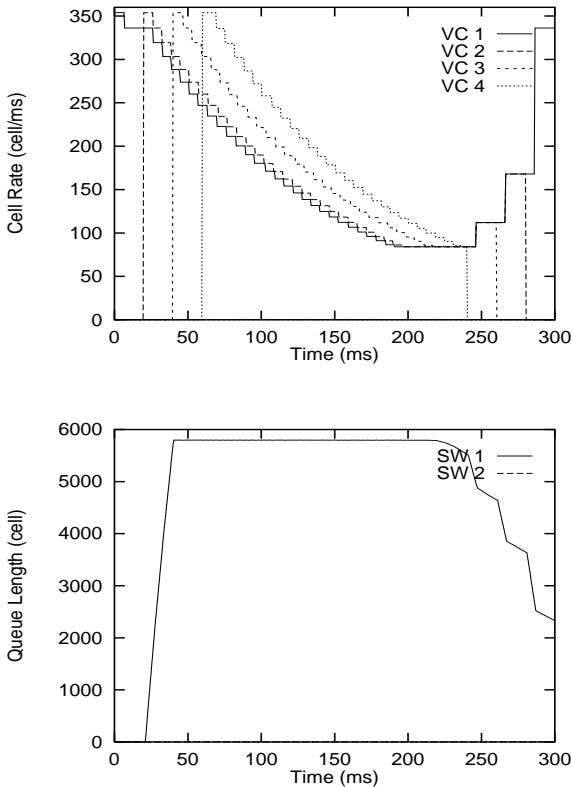


Figure 3: Effect of connection addition/disconnection in ERICA for  $\tau = 1.00$  ms and target utilization of 0.95.

Parameter Name	Assigned Value
$PCR$ (Peak Cell Rate)	$BW$
$MCR$ (Minimum Cell Rate)	$PCR/1000$
$ICR$ (Initial Cell Rate)	$PCR$
$TCR$ (minimum rate for data cells)	0.01
$RIF$ (Rate Increase Factor)	1
$RDF$ (Rate Decrease Factor)	1
$N_{RM}$ (RM cell opportunity)	32
$M_{rm}$ (control cell allocation)	2
$T_{rm}$ (minimum interval of RM cells)	100
$TBE$ (Transient Buffer Exposure)	$2^{24}$
$C_{rm}$ (# of RM cells without control)	32000
$CDF$ (Cutoff Decrease Factor)	1/2
$TOF$ (Time Out Factor)	2
$TDF$ (Time out Decrease Factor)	$ICR / 2^{14}$

Table 2: Control parameters at the source end system.

val are set to be 0.95 and 100 cell time. In ERICA, the target utilization is used to limit the bandwidth allocation for ABR connections; that is,  $(\text{target utilization} \times BW)$  of the bandwidth is shared by ABR connections, and the rest of the bandwidth is not allocated to absorb the rate fluctuation. The load averaging interval is an interval for monitoring the current traffic load at the switch. Readers should refer to [7] for details of ERICA.

Each graph shows  $ACRs$  of source end systems and queue lengths of switches. As can be found from these figures, the queue length grows when the new connection is activated (around  $t = 20, 40$  and  $60$  ms), and the maximum queue length is about 470 cells in the LAN environment. Since the target utilization is less than 1.0, the buffered cells are gradually processed and the queue length diminishes. In simulation, the queue length is decreased in about 30 ms, and the maximum queue length is limited even with several new connections. In the WAN environment, however, many cells are lost due to buffer overflow as can be found from Fig. 3. The number of lost cells was 59,927 cells during the simulation run. It can also be found that fairness among connections is not fulfilled. This problem also occurs in EPRCA++, which is the previous version of ERICA [10]. Buffer overflow can be avoided by setting the target utilization to be a much smaller value. However, it should be noted that setting a small value of the target utilization causes lower utilization of the bandwidth.

In Figs. 4 and 5, we next show simulation results of the original max-min scheme for  $\tau = 0.01$  and 1.00 ms. From the figures, it can be found that cell loss can be prevented even in the WAN environment, and that the maximum queue length is much smaller than the one obtained by ERICA. It is because the max-min scheme can adjust  $ACR$  of the new connection to the correct value in one round-trip time. How-

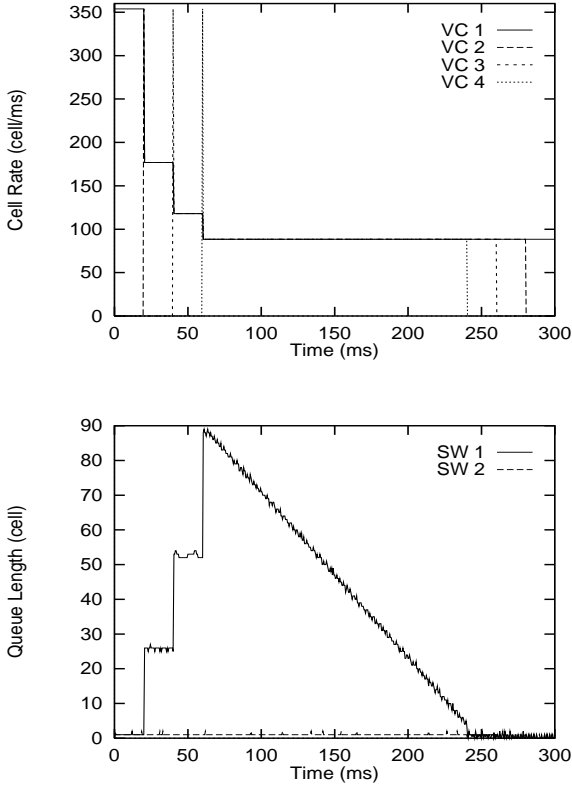


Figure 4: Effect of ABR connection arrival/departure in max-min scheme for  $\tau = 0.01$  ms.

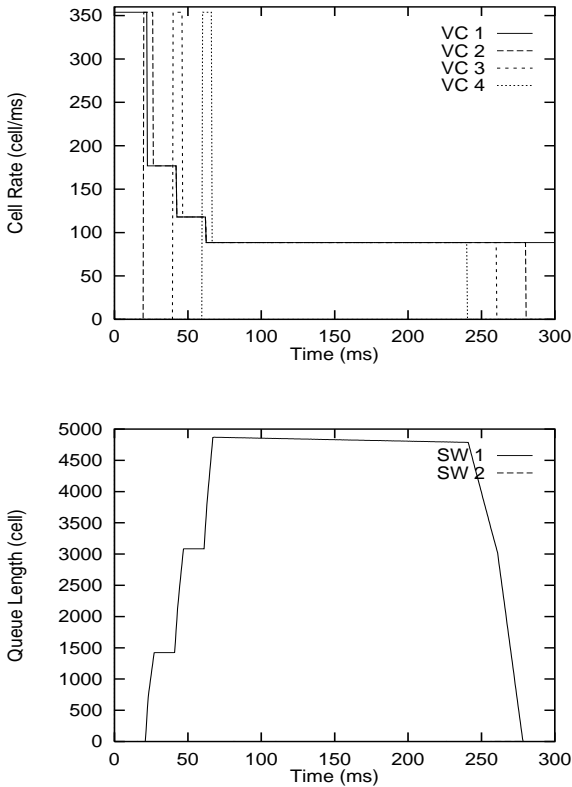


Figure 5: Effect of ABR connection arrival/departure in max-min scheme for  $\tau = 1.00$  ms.

$VCI$	$ER_F$	$ER_B$	$CA$	constrained
1 ~ 4	353.7	88.4	88.4	true

Table 3: Information table at SW1 before VC4 terminates.

$VCI$	$ER_F$	$ER_B$	$CA$	constrained
1 ~ 4	88.4	353.7	88.4	true

Table 4: Information table at SW2 before VC4 terminates.

ever, the serious problem of the max-min scheme is that each connection cannot increase its  $ACR$  even when some connections are terminated. Namely, max-min fairness is not satisfied after  $t = 240$  ms. This is due to a deadlock problem of the max-min scheme explained as follows. Tables 3 and 4 show information tables maintained at SW1 and SW2 before VC4 terminates at  $t = 240$  ms. Note that all connections have the same entry. When VC4 terminates, the switch tries to reallocate the available bandwidth. Since there are three active connections, the switch computes the fair share,  $FS$ , as  $BW/3 (= 117.9$  cell/ms) according to Eq. (2). However, the minimum of  $ER_F$  and  $ER_B$  is 88.4 cell/ms at both SW1 and SW2, all connections are regarded as constrained. Consequently, the bandwidth allocation for each connection is still limited to 88.4 cell/ms (see Eqs. (3) and (4)).

Another problem of the max-min scheme is that the queue length is settled at a high level. It becomes more apparent in the WAN environment as shown in Fig. 5. In the figure, the maximum queue length is about 4,700 cells, and cells would be lost if one more connection is added to the network. In other words, it takes long time for the queue length to be decreased because the max-min scheme tries to fully utilize the available bandwidth even though the queue length is almost full.

We next show simulation results of our enhanced max-min scheme in Figs 6 and 7 for  $\tau = 0.01$  and 1.00 ms, respectively. In these figures,  $Q_T$  is chosen according to our analysis presented in Subsection 2.2: in these cases,  $Q_T = 138$  in the LAN environment and  $Q_T = 1,189$  in the WAN environment. Bandwidth adjustment factors,  $\Delta_1$  and  $\Delta_2$ , are set to be 0.2 and 0.5, respectively. It can be found from these figures that the maximum queue length is small, and that the queue length is stabilized at  $Q_T$ . It can also be found that the queue length is decreased quickly once the queue length exceeds  $Q_T$ . It is owing to the mechanism of our enhanced max-min scheme to control the queue length. Our enhanced max-min scheme frequently updates the bandwidth allocation when compared with the original one. However, frequent computation of the bandwidth allocation would be indispensable when the background traffic coexists in the network.

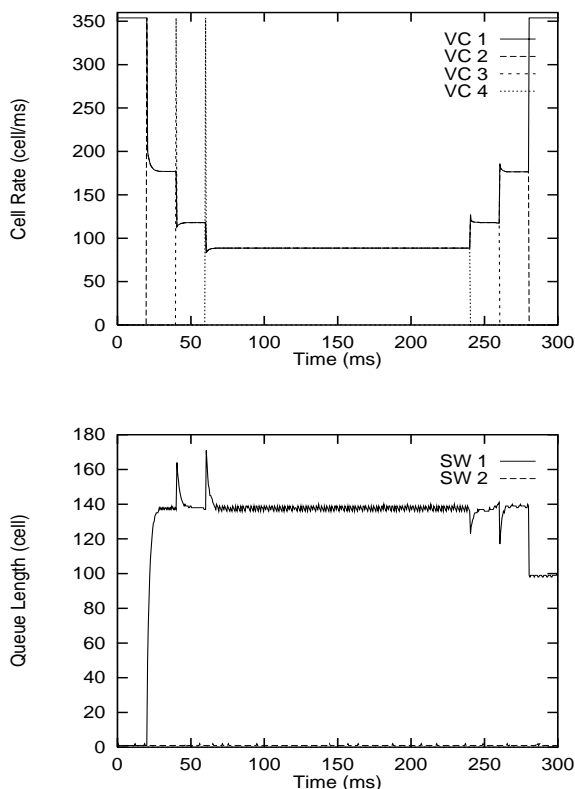


Figure 6: Effect of ABR connection arrival/departure in enhanced max-min scheme for  $\tau = 0.01$  ms.

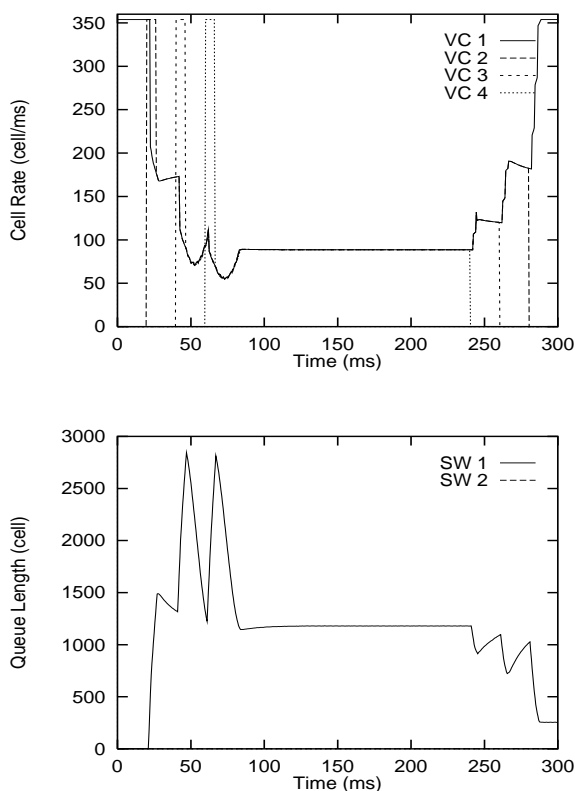


Figure 7: Effect of ABR connection arrival/departure in enhanced max-min scheme for  $\tau = 1.00$  ms.

## 4 Conclusion

In this paper, we have focused on explicit-rate marking switch, which utilizes the ER value in the RM cell for allocating bandwidth to each connection. We have proposed our explicit-rate switch algorithm, which is an enhanced version of the max-min scheme. Through simulation experiments, we have evaluated the performance of our switch algorithm, and have shown that our switch algorithm can achieve better efficiency and stability compared with other switch algorithms.

## References

- [1] The ATM Forum Technical Committee, "Traffic management specification version 4.0 (draft version)," *ATM Forum Contribution 95-0013R10*, February 1996.
- [2] H. Ohsaki, M. Murata, H. Suzuki, C. Ikeda, and H. Miyahara, "Rate-based congestion control for ATM networks," *ACM SIGCOMM Computer Communication Review*, vol. 25, pp. 60–72, April 1995.
- [3] R. Jain, "Congestion control and traffic management in ATM networks: recent advances and a survey," *ATM Forum Contribution 95-0177*, 1995.
- [4] L. Roberts, "Enhanced PRCA (proportional rate-control algorithm)," *ATM Forum Contribution 94-0735R1*, August 1994.
- [5] A. W. Barnhart, "Explicit rate performance evaluations," *ATM Forum Contribution 94-0983R1*, October 1994.
- [6] K.-Y. Siu and H.-Y. Tzeng, "Limits of performance in rate-based control schemes," *ATM Forum Contribution 94-1077*, November 1994.
- [7] R. Jain, S. Kalyanaraman, R. Viswanathan, and R. Goyal, "A sample switch algorithm," *ATM Forum Contribution 95-0178*, February 1995.
- [8] D. H. K. Tsang, W. K. F. Wong, S. M. Jiang, and E. Y. S. Liu, "A fast switch algorithm for ABR traffic to achieve max-min fairness," in *1996 International Zurich Seminar on Digital Communications* (B. Plattner, ed.), pp. 161–172, Springer, February 1996.
- [9] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, New Jersey: Prentice-Hall, 1987.
- [10] H. Ohsaki, M. Murata, H. Suzuki, C. Ikeda, and H. Miyahara, "Performance evaluation of rate-based congestion control algorithms in multimedia ATM networks," *IEEE GLOBECOM '95*, pp. 1243–1248, November 1995.