EPRCA を用いたレート制御方式の性能評価

— EFCI スイッチの初期状態解析 —

大崎 博之, 村田 正幸, †鈴木 洋, †池田 千夏, 宮原 秀夫

大阪大学 基礎工学部 情報工学科
〒 560 大阪府豊中市待兼山町 1-3

(Phone) +81-6-850-6587
(Fax) +81-6-850-6589
(E-mail) oosaki@ics.es.osaka-u.ac.jp

†日本電気 C&C システム研究所
〒 216 神奈川県川崎市宮前区宮崎 4-1-1

あらまし　レート制御方式は、その実現の容易性などから、 ATM 網において有効な輻輳制御方式であると考えられている．また、その実現方式の 1 つとして EPRCA (Enhanced Proportional Rate Control Algorithm) が ATM フォーラムにおいて提案されている．本稿では、 EPRCA の解析モデルを提案し、これを用いて、初期状態における EPRCA の性能を定量的に評価する． EPRCA で提案されている 3 種類の ATM スイッチおよび制御用セルに優先権を持たせるスイッチを評価の対象とする．また、数値例によって、伝播遅延時間やコネクション数が、 EPRCA の性能に与える影響を明らかにする．

和文キーワード　レート制御方式、 ERPCA、 EFCI スイッチ、初期状態解析

Performance of Rate-Based Control Method Using EPRCA

– Initial Transient State Analysis of EFCI Switch –

Hiroyuki OHSAKI, Masayuki MURATA, †Hiroshi SUZUKI, †Chinatsu IKEDA, Hideo MIYAHARA

Faculty of Engineering Science, Osaka University
Toyonaka, Osaka 560, Japan

(Phone) +81-6-850-6587
(Fax) +81-6-850-6589
(E-mail) oosaki@ics.es.osaka-u.ac.jp

†C&C System Research Labs. NEC Corporation
Kawasaki, Kanagawa 216, Japan

Abstract　Rate-based congestion control is effective and still simple for traffic management in ATM networks. As one of practical realization schemes, Enhanced Proportional Rate Control Algorithm (EPRCA) has been discussed in the ATM Forum. The main purpose of this paper is to analyze the performance of EFCI switch, which is an essential one suggested in EPRCA, in an initial transient state. By providing an analytic model for EPRCA with homogeneous traffic sources and a single bottleneck ATM link, we obtain performance measures in terms of the maximum queue length at the switch and its throughput by utilizing the first order fluid approximation method. We provide suggestive equations for parameter tunings of EPRCA and show some numerical examples.

英文 key words　Rate-Based Congestion Control, EPRCA, EFCI Switch, Initial Transient State Analysis

1 Introduction

Congestion control mechanism plays an essential part for efficient traffic management in ATM networks. Recently, rate-based congestion control has been adopted by the ATM Forum because of its simplicity for implementation and scalability from local area to wide area networks. In rate-based congestion control, cell transmission rate at a source end system is controlled by feedback information from the network. The source end system increases cell transmission rate if the network is not congested. Once the network is congested, the source end system decreases its cell transmission rate after it receives congestion indication.

Our main objective of the current paper is to evaluate the rate-based flow control scheme by an analytic approach. Several analytic studies for the rate control scheme have already been made [1, 2, 3]. For example, Yin et al. analyzed a dynamical behavior based on the timer-based approach [3]. In their model, the source end system changes its cell transmission rate regularly at every fixed time interval. However, it has already been found that this approach causes problems in some situations [4]. Henceforth, an improved rate control algorithm called Enhanced Proportional Rate Control Algorithm (EPRCA) is recently proposed in [5], which has been adopted by the ATM Forum as a standard for the rate-based congestion control scheme in ATM LANs.

The EPRCA suggests three types of switches, EFCI bit setting switch (EFCI), Binary Enhanced Switch (BES) and Explicit Down Switch (EDS). Each switch has different functionalities against network congestion. While we have analyzed the effectiveness of the EFCI switch in steady state in [6], a careful treatment is necessary when the large number of VC's share the link. Since the aggregate rate is increased rapidly in such a circumstance, the queue length tends to be unacceptably large. This tendency becomes remarkable when multiple VC's start cell transmission at same time. This is known as a "large VC's problem", and one of serious problems for effective rate-based congestion control, which is our main subject of the current paper.

In this paper, we analyze an initial transient behavior of the EFCI switch by a similar approach taken in [6]. By assuming that all connections begin their cell transmission simultaneously, we will show a dynamical behavior of EPRCA and a maximum buffer requirement for the switch.

The rest of this paper is organized as follows. In Section 2, the mechanism of EPRCA is introduced and our analytic model is presented. Sections 3 is devoted to the analysis of the EFCI switch. Our concluding remarks are presented in Section 4.

2 Analytic Model

The EPRCA is based on a positive feedback mechanism [5]. SES periodically sends a RM (Resource Management) cell every N_{RM} data cells to check congestion status of the network. The congested switch marks an EFCI (Congestion Indication) bit in the header of on-going data cells. Then, a destination end system can recognize the congestion by EFCI bits. The received RM cell at the destination is returned to the source along the backward path if the previous data cell experience no congestion.

As presented in [3], in a basic operation of the rate-based congestion control, each source end system normally increases its allowable cell transmission rate, which will be called ACR (Allowed Cell Rate). ACR is decreased when the network falls into congestion. However, a notable feature of EPRCA is that each source always decreases its cell transmission rate until it receives an RM cell from the network. The source end system can increase ACR only when it receives the RM cell. If the RM cell is discarded due to congestion, it results in that the source end system continues to decrease its ACR. By this mechanism, a fast congestion recovery can be achieved. In the EPRCA, three types of switch architectures are suggested in the form of pseudo codes with different functionalities. The first one is an EFCI bit setting switch (EFCI), which is same as the original PRCA [7], and can be expected as a least expensive one. In EFCI switches, forward RM cells are not necessary because the EFCI bit contained in the header of data cells can be used for the congestion indication. However, the RM cells are required when there exist other switches such as BES or EDS switches.

As in the previous paper [6], we provide analytic results for the EFCI switch for a rather simple network model which consists of homogeneous traffic sources and a single bottleneck ATM link, which is shared by the number N_{VC} of VC's (Fig. 1). We assume that those VC's behave identically by using the identical control parameters. The service rate of the switch, i.e., the bandwidth of the bottleneck link, is denoted by BW. For example, when the link capacity is 150Mbps, BW is equal to 353.208 cells/msec. Propagation delays from the source end system to the switch and from the switch to the destination end system are represented by τ_{sx} and τ_{xd}. Propagation delays τ_{sx} and τ_{xd} may differ according to the network configuration (e.g., LAN, WAN and the location of UNI). Defining τ as a round-trip propagation delay between the source and destination end systems: $\tau = 2(\tau_{sx} + \tau_{xd})$. We further introduce $\tau_{xds} = 2\tau_{xd} + \tau_{sx}$ which is a propagation delay of the congestion indication from the switch to the source via the destination end system.

The congestion is detected by queue length threshold values. The EFCI switch has high and low threshold values denoted by Q_H and Q_L for congestion occurrence and relief. When the queue length at the switch exceeds Q_H, the switch detects congestion, and the action against congestion is taken. On the other hand, it is regarded as termination of congestion when the queue length goes under Q_L.

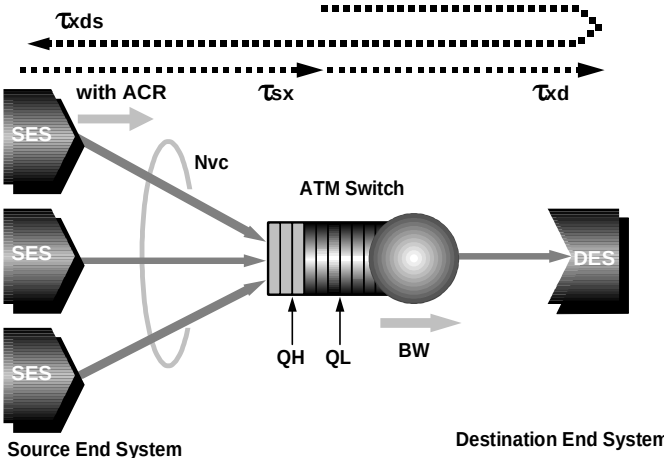

Fig. 1: Analytic Model.

Let us introduce $ACR(t)$ and $Q(t)$ which represent ACR for each source and the queue length at the switch observed at time t. In [6], we have analyzed evolution of $ACR(t)$ and $Q(t)$ in steady state. As shown in [6], as the number of VC's becomes large, the queue length tends to be extremely large. This tendency becomes unacceptable in the case where the large number of VC's starts cell transmission at same time. Of course, it depends on ICR, which is an initial value of ACR.

Since ICR can be chosen freely, one may consider that ICR should be set to be small enough so that the queue length does not become large. As shown in the numerical examples, it is true if one can know the number of active VC's (N_{VC}) in advance. However, in the actual network environment, it may be difficult. Further, a smaller value of fixed ICR cannot achieve high utilization of the link when N_{VC} is small. In this paper, we provide the analysis of an initial transient behavior of rate-based congestion control for given N_{VC} and ICR to exploit the suggestive control parameters by assuming that (1) the switch has infinite capacity of the buffer, and that (2) SES always has cells to transmit. Therefore, $ACR(t)$ is equivalent to the actual cell transmission rate.

3 Analysis

In this section, we analyze an initial transient behavior of the allowed cell rate $ACR(t)$ and the queue length $Q(t)$ for a given ICR to show that an appropriate choice of ICR plays an important role for achieving effective control while keeping the maximum queue length to an appropriate value. The latter is important for buffer-dimensioning in designing ATM switches.

3.1 Derivation of $ACR(t)$

Figs. 2 and 3 show pictorial views of $ACR(t)$ and $Q(t)$ in the initial transient state for different ICR's. As illustrated, the evolution of $ACR(t)$ and $Q(t)$ is classified into two categories according to the following relations: $N_{VC}\,ICR > BW$ or $N_{VC}\,ICR < BW$.

To see this, we will derive $ACR(t)$ and $Q(t)$ from now on. Here, we should note that a more rigorous treatment is required regarding the above classification because it should also be affected by other parameters as will be shown in the below.

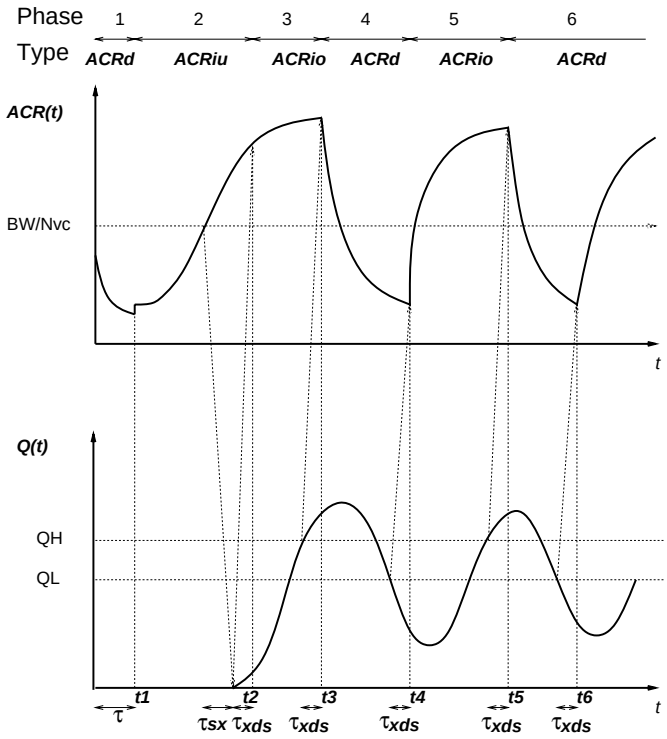

Fig. 2: Pictorial View ($ICR < BW/N_{VC}$).

We first divide $ACR(t)$ into phases, each of which has a different form dependent on the congestion status of the switch. There are three types of $ACR(t)$.

Type I: $ACR_d(t)$

$ACR(t)$ is decreased exponentially.

Type II: $ACR_{iu}(t)$

$ACR(t)$ is increased and the offered load to the link is below its capacity BW.

Type III: $ACR_{io}(t)$

$ACR(t)$ is increased and the offered load to the link is beyond its capacity BW.

In the case of $N_{VC}\,ICR > BW$, $ACR_d(t)$, $ACR_{iu}(t)$, $ACR_{io}(t)$ appears in that order (Fig. 2). Then, $ACR_d(t)$ and $ACR_{io}(t)$ are repeated. On the other hand, a cycle consisting of $ACR_d(t)$ and ACR_{io} is repeated in the case of $N_{VC}\,ICR < BW$ (Fig. 3). Let us denote $ACR_i(t)$ and the corresponding $Q_i(t)$ as $ACR(t)$ and $Q(t)$ of Phase i, respectively, that is,

$$\begin{aligned} ACR_i(t) &= ACR(t - t_{i-1}), \quad 0 \le t < t_i, \\ Q_i(t) &= Q(t - t_{i-1}), \quad 0 \le t < t_i, \end{aligned}$$

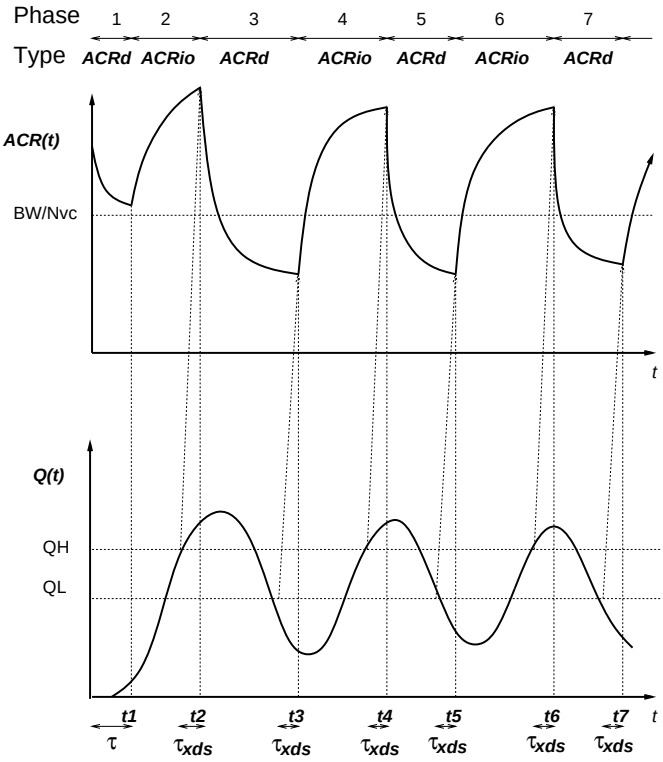

Fig. 3: Pictorial View ($ICR > BW/N_{VC}$).

where t_i is the time when Phase i terminates. Further, the length of Phase i is defined by

$$t_{i-1,i} = t_i - t_{i-1}.$$

Each source end system transmits the RM cell at time 0 followed by data cells. Until the source end system receives the first RM cell, its ACR is decreased exponentially. At time τ_{sx}, the RM cells arrive at the switch. At this time, the switch is not congested, and the RM cells will be returned to the source. In actual, it takes one cell time for the switch to handle each RM cell. Then, it may be queued up during processing RM cells when the number of VC's is large. However, we assume that all RM cells from source end systems are transferred simultaneously by the switch. Since the queue length is zero at time τ_{sx}, all of source end systems will increase ACR at time $\tau_{sx} + \tau_{xds} = \tau$ by receiving the first backward RM cell. The next phase is determined by the relation between ICR and BW/N_{VC}. Type II ($ACR_{io}(t)$) appears when $N_{VC}ICR < BW$ (Fig. 2), and Type III if $N_{VC}ICR < BW$ (Fig. 3). In what follows, we first derive $ACR_d(t)$, $ACR_{iu}(t)$ and $ACR_{io}(t)$. Then, evolution of $ACR(t)$ and $Q(t)$ is shown by determining the length and the initial value of each phase.

Since the derivation of $ACR(t)$ and $Q(t)$ has already been shown in [6], only results are presented. Refer to [6] for more details.

$ACR_d(t)$ takes an exponential form given by

$$ACR_d(t) = ACR_d(0)e^{\frac{-ACR_d(0)}{MD}t}.$$

$ACR_{io}(t)$ is represented as

$$ACR_{io}(t) = \frac{a_1 e^{-a_1 t} + a_2 r e^{-a_2 t}}{c_1(e^{-a_1 t} + re^{-a_2 t})} \quad (1)$$

where a_1 and a_2 are roots of the equation

$$a^2 + c_2 a + c_1 c_3 = 0,$$

and c_1, c_2 and c_3 are given as

$$c_1 = -\frac{1}{MD}; c_2 = \frac{BW}{MD\,N_{VC}}; c_3 = \frac{BWAIR}{N_{VC}},$$

The initial value $ACR_{io}(0)$ determines r as

$$r = \frac{a_1 - ACR_{io}(0)c_1}{ACR_{io}(0)c_1 - a_2}.$$

Finally, $ACR_{iu}(t)$ is approximately obtained as

$$ACR_{iu}(t) \cong ACR_{iu}(0)e^{\beta t}, \quad (2)$$

where β is given as a root of the equation

$$\beta = \frac{N_{RM}\,AIR}{MD\,\log(\frac{MD}{MD - N_{RM}})}e^{-\tau\beta}.$$

3.2 Evolution of $ACR(t)$ and $Q(t)$

In this subsection, $ACR(t)$ and $Q(t)$ are derived by taking the same approach presented in [6].

Case 1: $ICR < BW/N_{VC}$

In this case, ACR is decreased until the first RM cell is returned to the source at $t = \tau$. Therefore,

$$\begin{aligned} t_1 &= \tau \\ ACR_1(t) &= ACR_d(t), \quad 0 \le t \le t_1, \\ Q_1(t) &= 0, \quad 0 \le t \le t_1 + \tau_{sx}, \end{aligned}$$

where the initial value of $ACR_1(t)$ is ICR. ACR is then increased until reaching at the value BW/N_{VC}.

$$\begin{aligned} ACR_2(t) &= ACR_{iu}(t), \quad 0 \le t \le t_{12}, \\ Q_2(t) &= 0, \quad \tau_{sx} \le t \le t_{12} + \tau_{sx}. \end{aligned}$$

The time t_{12} is given by as

$$t_{12} = ACR_2^{-1}(BW/N_{VC}) + \tau$$

During Phase 3, the queue length grows and RM cells are returned with a fixed interval $N_{VC}N_{RM}/BW$ and ACR is increased according to $ACR_{io}(t)$.

$$\begin{aligned} ACR_3(t) &= ACR_{io}(t), \quad \le t \le t_{23} \\ Q_3(t) &= \int_{x=\tau_{sx}}^{t} (N_{VC}\,ACR_3(x - \tau_{sx}) - BW)dx, \\ &\quad \tau_{sx} \le t \le t_{23} + \tau_{sx}. \end{aligned}$$

where t_{23} is a solution of $Q_3(t_{23}) = Q_H$.

In what follows, we will use the convention $t_{23} = Q_3^{-1}(Q_H)$ for brevity. After the queue length reaches the threshold value Q_H, ACR is again decreased. ACR is then increased when Q becomes the lower threshold value Q_L.

Case 2: $ICR > BW/N_{VC}$

In this case, ACR is decreased until the first RM cell is returned to the source at $t = \tau$ as in the above case, i.e.,

$$\begin{aligned} t_1 &= \tau \\ ACR_1(t) &= ACR_d(t), \quad 0 \le t \le t_1, \\ Q_1(t) &= \int_{x=tsx}^{t} (N_{VC}\, ACR_1(x - \tau_{sx}) - BW)dx, \\ & \quad \tau_{sx} \le t \le t_1 + \tau_{sx}. \end{aligned}$$

If $ACR_1(\tau)$ is still beyond BW/N_{VC}, ACR is increased and the RM cells are returned with a fixed interval $N_{VC}\, N_{RM}/BW$ during Phase 2.

$$\begin{aligned} ACR_2(t) &= ACR_{io}(t), \quad 0 \le t \le t_{12}, \\ Q_2(t) &= Q_1(t_1 + \tau_{sx}) \\ &\quad + \int_{x=tsx}^{t} (N_{VC}\, ACR_2(x - \tau_{sx}) - BW)dx, \\ & \quad \tau_{sx} \le t \le t_{12} + \tau_{sx}. \end{aligned}$$

where t_{12} is a solution of

$$t_{12} = Q_2^{-1}(Q_H) + \tau_{xds}$$

Otherwise, Phase 2 according to $ACR_{iu}(t)$ begins as in the above case. When the queue length reaches the high threshold value Q_H, ACR is again decreased.

Last, we note that for later phases, the steady state analysis presented in [6] can be applied.

3.3 Maximum Queue Length

In this subsection, we show the maximum queue length for two cases.

Case 1: $ICR < BW/N_{VC}$

As shown in Fig. 2, $Q(t)$ reaches at its maximum value during Phase 4. Let the queue length take its maximum value Q_{max} at t_{max}. Then, we have a relation

$$\begin{aligned} Q_{max} &= Q^{-1}(t_{max}) \\ t_{max} &= ACR^{-1}(BW/N_{VC}) + \tau_{sx} \end{aligned}$$

Case 2: $ICR > BW/N_{VC}$

As shown in Fig. 3, $Q(t)$ reaches at its maximum value during Phase 3 as

$$\begin{aligned} Q_{max} &= Q^{-1}(t_{max}), \\ t_{max} &= ACR^{-1}(BW/N_{VC}) + \tau_{sx}. \end{aligned}$$

3.4 Numerical Examples

In this subsection, some numerical examples for the EFCI switch are provided. In these examples, both Q_H and Q_L are identically set to 500, and other control parameters are set to the suggested values shown in [5].

Effects of the propagation delays on $ACR(t)$ and $Q(t)$ are displayed in Figs. 4 and 5 for $N_{VC} = 10$. Since a larger value of τ causes slower congestion notification, the queue length is built up initially. Therefore, it is hard to directly apply the EFCI switch in the case where we interconnect LANs located in the long distance. After then, the queue length is cyclically fluctuated. For example, the maximum queue length is 2,000 in the case of $\tau = 1.0$ in steady state [6] while the queue length becomes about 1,240 initially.

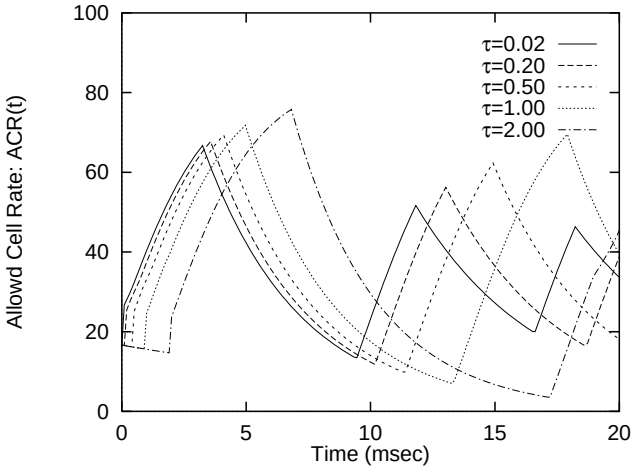

Fig. 4: Effect of Propagation Delay on $ACR(t)$ ($N_{VC} = 10$).

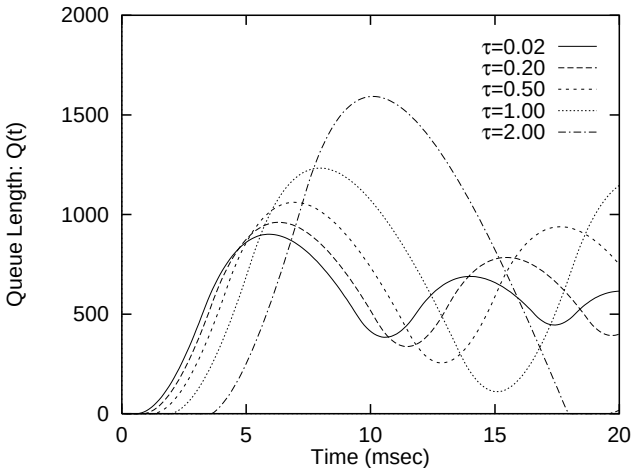

Fig. 5: Effect of Propagation Delay on $Q(t)$ ($N_{VC} = 10$).

Figures. 6 and 7 show $ACR(t)$ and $Q(t)$ for different values of N_{VC}. The propagation delay between the source and destination end systems is set to be 0.05 msec (around 2km) as a typical value for a LAN environment. It is obvious that the large N_{VC} causes an increase of the maximum queue length even in the case of short propagation delays. In Fig. 6, $ACR(t)$ is growing in spite of the switch congestion when $N_{VC} = 50$. It is due to the form of $ACR(t)$ given by Eq. (1). Therefore, the queue length becomes large instead of congestion relief.

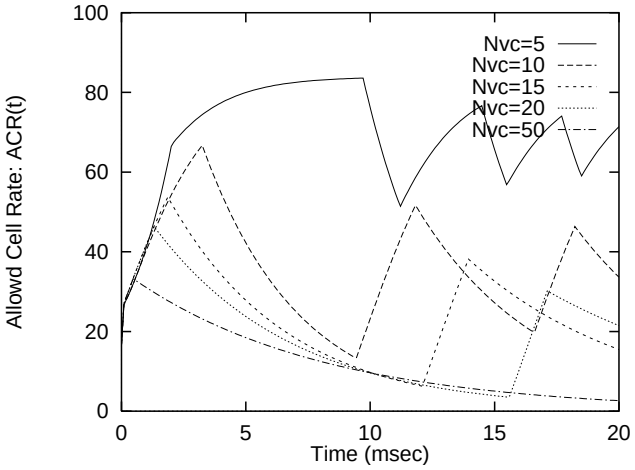

Fig. 6: Effect of N_{VC} on $ACR(t)$ ($\tau = 0.02$ msec).

A possible solution for decreasing the maximum queue length is to set ICR properly. In Fig 8, the different values of ICR are used in the case where $N_{VC} = 50$ and $\tau = 0.02$. As can be seen in the figure, appropriate ICR can decrease the maximum queue length to some extent. However, it requires to know the active number of connections, N_{VC}, in advance.

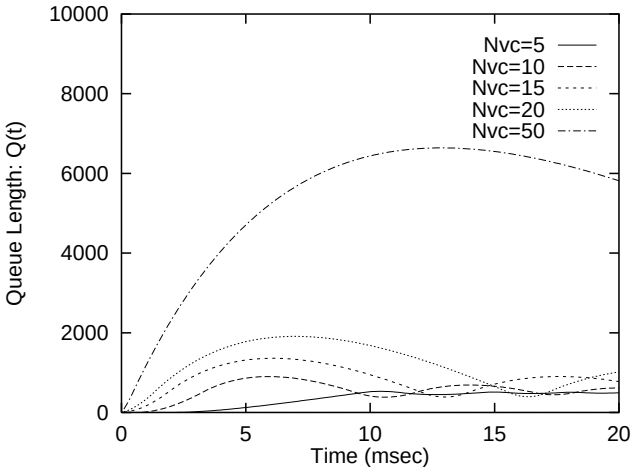

Fig. 7: Effect of N_{VC} on $Q(t)$ ($\tau = 0.02$ msec).

4 Conclusion

In this paper, Enhanced Proportional Rate Control Algorithm (EPRCA), which has been adopted as a standard by the ATM Forum, has been analyzed under the assumption that all VC's begin their cell transmission simultaneously. Through numerical results, we have quantitatively shown the effect of the number of VC's, the propagation delays and the initial cell rate on the maximum queue length at the switch.

For further works, some mechanism to estimate the initial cell rate appropriately according to the number of active connections should be investigated. In general, it is not easy to know the number of active VC's in advance. However, an appropriate signalling protocol may realize such a mechanism.

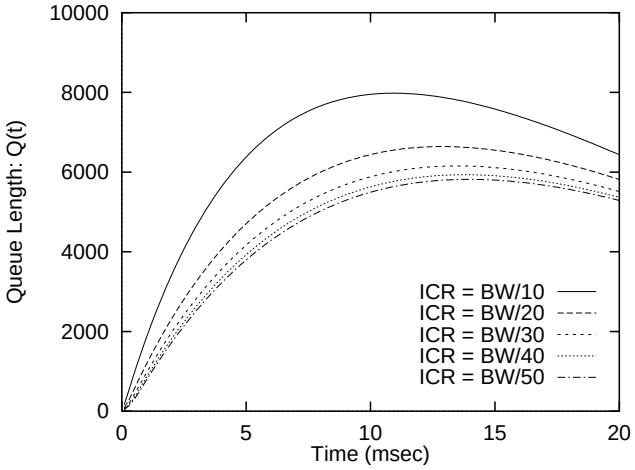

Fig. 8: Effect of Initial Transmission Rate on $Q(t)$ in EFCI Switch ($N_{VC} = 50$).

References

[1] J.-C. Bolot and A. U. Shankar, "Dynamical behavior of rate-based flow control mechanisms," *Computer Communication Review*, vol. 20, pp. 35–49, 4 1990.

[2] K.K.Ramakrishnan and R.Jain, "A binary feedback scheme for congestion avoidance in computer networks," *ACM Transactions on Computer Systems*, vol. 8, no. 2, pp. 158–181, 1990.

[3] N. Yin and M. G. Hluchyj, "On closed-loop rate control for ATM cell relay networks," *Proceedings of INFOCOM '94*, pp. 99–109, 1994.

[4] J. C. R. Bennett and G. T. D. Jardins, "Failure modes of the baseline rate based congestion control plan," *ATM Forum/94-0512*, July 1994.

[5] L. Roberts, "Enhanced PRCA (Proportional Rate-Control Algorithm)," *ATM Forum/94-0735*, September 1994.

[6] H. Ohsaki, M. Murata, H. Suzuki, C. Ikeda, and H. Miyahara, "Performance of rate-based control method using ERPCA — steady state analysis of efci switch —," *IEICE Technical Report*, pp. 69–74, January 1995.

[7] A. W. Barnhart, "Baseline performance using PRCA rate-control," *ATM Forum/94-0597*, July 1994.