明確速率指示之交換器的置放方法

The Placement of ER Switches in the Heterogeneous ATM Environments

賴源正　　林盈達
Yuan-Cheng Lai　　Ying-Dar Lin

交通大學資訊科學研究所
Department of Computer and Information Science
National Chiao Tung University

摘要

隨著交換器技術的進步，明確速率指示(ER)之交換器將會比明確擁塞指示(EFCI)之交換器更受歡迎，原因在於它擁有較佳的性能，然而在此過渡時期，兩種交換器將會共同存在於 ATM 網路中，在此架構下明確速率指示之交換器放置的位置將對網路效率產生重大影響，故本篇論文提出一些放置的規則，以提供網路管理者的參考，使其產生更好的網路效能。

關鍵字：明確速率指示，明確擁塞指示，可利用位元速率，速率流速控制。

Abstract

With the advances in switching technologies, ER switches are becoming popular since they perform better than EFCI switches. In the transitional period, the EFCI and ER switches may coexist in the same ATM network. Because the location of an ER switch in this topology is critical to its performance, some placement rules, which describe how to place the ER switches for the network operators, are developed in order to achieve a better performance of the network.

Keywords: ER (Explicit Rate), EFCI (Explicit Forward Congestion Indication), ABR (Available Bit Rate), Rate-based flow control

1. Introduction

The rate-based flow control for available bit rate (ABR) traffic is defined in the ATM Forum Traffic Management Specification Version 4.0 to provide a wide range of non-real time applications [1]. The behavior of traffic sources and destinations is clearly defined in order to provide the base line for the vendors to follow. However, the methods that the switches should use to control the source rate are up to the vendors. Currently most ATM vendors have already provided ATM switches equipped with Explicit Forward Congestion Indication (EFCI) functions recommend by the ITU-T [2].

These switches, which use EFCI marking, are called the first-generation switches [3]. With the advances in switching technologies, the second-generation switches, which own the Explicit Rate (ER) setting capacity, is becoming popular. In the transitional period from first to second-generation switches, the interoperating EFCI and ER switches becomes unavoidable. In the mixed EFCI-ER environment, the location of ER is a critical issue for performance. When a customer buys a switch equipped with ER setting, which EFCI switch should be replaced first to obtain the best performance? We performed some simulations with various network configurations.

The simulation results provide important implications on which EFCI switches to be replaced. Therefore, from our results, the network managers can have better insight to help them purchasing proper ER switches to fit their needs. Also when they are about to install their ER switches, they can choose the appropriate EFCI switches to be replaced.

The remainder of this paper is organized as follows. In section 2, we describe the operations of EFCI switches and some ER switches. Simulation results in the environment that all switches use the same scheme are presented in section 3. Some guidelines to place ER switches are given in section 4. Finally, we give a conclusion in section 5.

2. ABR Flow Control

First we briefly introduce the basic operation of the closed-loop rate-based control mechanism [1]. When a virtual circuit (VC) is established, the source end system

(SES) sends the cells at the allowed cell rate (ACR) which is set as initial cell rate (ICR). In order to probe the congestion status of the network, the SES sends a forward Resource Management (RM) cell every Nrm data cells. Each switch may set certain fields of the RM cell to indicate its own congestion status, or the bandwidth the VC source should use. The destination end system (DES) returns the forward RM cell as a backward RM cell to the SES. According to the received backward RM cell, the SES adjusts its allowed cell rate, which is bounded between Peak cell rate (PCR) and Minimum cell rate (MCR).

The RM cell contains a 1-bit congestion indication (CI) which is set to zero, and an explicit rate (ER) field which is set to PCR initially by the SES. When the SES receives a backward RM cell, it modifies its ACR using additive increase and multiplicative decrease. The new ACR is computed as follows, depending on CI and ER fields in RM cells:

ACR=max(min(ACR+RIF*PCR,ER),MCR), if CI=0,
ACR=max(min(ACR*(1-RDF),ER),MCR), if CI=1,

where RIF is the rate increase factor and RDF is the rate decrease factor. According to the way of congestion monitoring and feedback mechanism, various switch mechanisms can be classified into two types. One is the Explicit Forward Congestion Indication (EFCI) switch, the other is the Explicit Rate (ER) switch.

2.1 EFCI Scheme

In this scheme [8], when congestion occurs, the switch sets the EFCI bit to one (EFCI=1) in the header of each passing data cells. The DES, if a cell with EFCI=1 has been received, marks the CI bit (CI=1) to indicate congestion in each backward RM cells. In most cases, the queue length is used to decide whether congestion occurs or not. As the queue length exceeds a threshold, denoted by Q_t, congestion is claimed. When the queue length falls below the threshold, congestion is relieved.

2.2 Explicit Rate Feedback Schemes

In the ER schemes, the switch computes the fair share of bandwidth that a VC can be supported with, determines the load and determines the actual explicit rate. When each RM cell passes, the switch sets the ER field to the determined explicit rate. Note each switch is not allowed to increase the ER field. Thus a source shall receive the minimum allowed cell rate of all the switches along the path. Examples of ER switch mechanisms are the EPRCA, ERICA, CAPC, Charny Max-Min, Tsang Max-Min schemes [9-13].

- Enhanced Proportional Control Algorithm (EPRCA) [9]

Each switch maintains a mean allowed cell rate (MACR) using a running exponential weighted average. When a switch receives a forward RM cell during the congestion period, MACR is updated as

$$MACR=(1-\alpha)MACR + \alpha CCR,$$

where α is the exponential averaging factor generally set to be 1/16 and CCR is the current cell rate of the VC recorded in the RM cell. The fair bandwidth share is computed as a fraction of the MACR:

$$\text{Fair share} = DPF*MACR,$$

where DPF is a switch down pressure factor set close to but below 1. When a switch receives a backward RM cell, it reduces the ER field to the fair share if its queue length is larger than Q_t.

- Explicit Rate Indication for Congestion Avoidance (ERICA) [10]

The ERICA uses a load factor, z, to indicate the overload or underload state of the switch. The load factor is defined as

$$z=\frac{\text{Input rate}}{\text{Target rate}}.$$

The input rate is measured over a fixed averaging interval and the target rate is usually set slightly below the link bandwidth. Because the goal of this algorithm is to maintain the load factor close to one, the sources ought to change their current sending rates inversely proportional to the calculated load factor. The VC share and fair share are as follows:

$$VCShare=\frac{CCR}{z},$$

Fair share=Target rate/Number of active connections.

A switch updates the ER field, in the backward RM cell it received, to be the maximum value of the fair share and VCshare.

- Congestion Avoidance using Proportional Control (CAPC) [11]

Again as in the ERICA scheme, the switches set a target utilization slightly below 1 and compute the load factor. The main difference lies in the way the fair share is computed, which depends on whether z<1 or z>1. Thus, we have

Fair share=Fair share*min(ERU, 1+(1-z)*Rup), if z<1,
Fair share=Fair share*max(ERF, 1+(z-1)*Rdn), if z>1,

where Rup is a slope parameter between 0.025 and 0.1, and Rdn is between 0.2 and 0.8. ERU and ERF determine the maximum allowed increase and minimum allowed decrease, respectively. Usually ERU is set to 1.5 and ERF is set to 0.5. When a returning RM cell arrives at the switch, the ER field is updated to be the fair share.

- Charny Max-Min Scheme [12]

The fair share is computed using an iterative

procedure in this scheme. Initially, the fair share is set to the link bandwidth divided by the number of active VCs. Some VCs can not achieve the fair share at a switch because of the constraints imposed by the limited amount of bandwidth available at other switches along its path. For this switch, these VCs are called "constrained VCs". The switch can determine whether a VC is constrained or not by comparing the fair share with the CCR field in the received forward RM cell. If the CCR field is less than the fair share, the VC is a constrained VC. Otherwise, it is an unconstrained VC. For high throughput, the available bandwidth which the constrained VCs can not use should be utilized by the unconstrained VCs. Hence the fair share is computed as follows:

$$\text{Fair share}=\frac{\text{Link bandwidth - Bandwidth of constrained VCs}}{\text{Number of VCs - Number of constrained VCs}}.$$

As a forward RM cell traverses the network, the switch determines whether the VC is constrained or not, recomputes the fair share and reduces the ER field and CCR field of the RM cell down to their fair share. The ER field and CCR field of a backward RM cell may be further reduced down to the most current fair share on the forward path.

- Tsang Max-Min Scheme [13]

This scheme is similar to Charny Max-Min method, except for three differences:

1. The switch does not update the CCR field of the RM cell.
2. The switch determines a VC state depending on ER field, instead of CCR field, of the RM cell.
3. The switch determines the VC state and computes the fair share on forward and backward RM cells, not just the forward RM cell.

The following parameters are set for the above schemes in our simulation: Nrm=32, PCR=155Mbps, MCR=0bps, RIF=PCR/256, and RDF=15/16.For EFCI, we used Q_t=1000 cells. For EPRCA, we set Q_t=1000 cells, a=1/16, and DPF=7/8. Target rate is set to be 95% of the link bandwidth for EFICA and CAPC. Also in the CAPC scheme, we use the following parameters: ERU=1.5, Rup=0.05, ERF=0.5, and Rdn=0.5.

3. Homogeneous Environment

To understand the effect that the placement of ER switches in the heterogeneous environments, we first show the performance of these schemes in the homogeneous environment in which all switches utilize the same control scheme. We use a simple 3-switch configuration as shown in Figure 1 to be our network topology. It is sufficient to exhibit the characteristics of various switches [14,15]. Depending on the switches they pass, three groups of VCs are distinguished. Group 1 (G1) is the VCs traveling through link 1 only and Group 2 (G2) is the VCs traveling through link 2 only. Group 3 (G3) is the VCs passing both links 1 and 2. There are N1, N2, and N3 connections in the G1, G2, and G3, respectively. In our simulation of the homogeneous environment, there are two VCs in each group, i.e. N1=N2=N3=2.

The link between two neighboring switches is 100 km in length and 155 Mbps in capacity. For G3, the distance between a source and a switch is 1km. On the other hand, the distance between a source and a switch is 51km for G1 and G2. Hence the propagation delay is the same for all groups. The reason is because we want to eliminate any unfairness caused by the different propagation delays. Also all sources are persistent in order to investigate the performance in the most stressful situations.

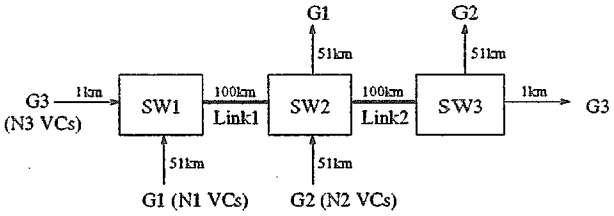

Figure 1. Simulation model.

Three performance issues we concern are the maximum queue length (MQL), utilization (U), and fairness (F). The maximum queue length is directly related to the cell loss probability when the buffer at the switch is finite. Also we can observe the situation about the oscillation of ACR with this value. When this value is high, the ACR usually has larger oscillation. From the value of utilization, we can know how much bandwidth is wasted at the switch. Unfair behavior is observed from the value of the fairness, which is defined as

$$F=\max(1-\max(|x_i-1|), 0)$$

where x_i is the ratio of the actual throughput to the fair throughput for source i. This definition is the maximum radio difference between ideal and achieved rate. We do not use the fairness definition, $(\Sigma x_i)^2 / n\Sigma x_i^2$, because only a few differences are observed. This is the major drawback of this metric.

Scheme	G1	G2	G3	MQL1	MQL2	U1	U2	F
Ideal	38.75	38.75	38.75	0	0	100%	100%	100%
EFCI	39.48	42.46	27.29	1580	1443	85.8%	89.7%	70.4%
EPRCA	40.85	41.10	36.41	1470	1393	99.4%	99.7%	93.9%
ERICA	38.53	38.87	36.85	3	3	96.9%	97.4%	95.1%
CAPC	37.49	36.87	35.89	172	131	94.4%	93.6%	92.6%
CMM	38.57	38.57	38.57	3	3	99.2%	99.2%	99.5%
TMM	38.57	38.57	38.57	3	3	99.2%	99.2%	99.5%

Table 1. Comparison of the various switch schemes in the homogeneous environments.

Simulation results in the homogeneous environment are shown in Table 1. MQL1 and U1 represent the maximum queue length and utilization at the switch 1. Similarly MQL2 and U2 represent the maximum queue length and utilization at the switch 2. Note the results about switch 3 are not shown because it do not become a bottleneck at any time. From the table, it is more unfair in the homogeneous EFCI environment than in the homogeneous ER environments. The main reason is that the beat down problem occurs when we use the pure EFCI switches. The beat down problem is that VCs passing through a larger number of switches get less bandwidth than VCs passing through a smaller number of switches [4,16].

Also we observe that the maximum queue length of EFCI and EPRCA is large. The is caused by the large oscillation of ACR. Actually, in the homogeneous environments, the ACR of EFCI and EPRCA has large oscillation, the ACR of ERICA and CAPC has small oscillation, and the ACR of CMM and TMM is oscillation free [5,7].

Regarding utilization, link capacity is not fully utilized because Q_t is set too low in the homogeneous EFCI environment. However, if Q_t is set high, the maximum queue length will be raised dramatically [17]. Hence we sacrifice some bandwidth to keep the maximum queue length in the reasonable range. On the other hand, high utilization is achieved in the homogeneous ER environments. EPRCA has high utilization. ERICA and CPAC achieve the target rate. For TMM and CMM, they utilize almost the complete link bandwidth.

We summary these results which are observed in the homogeneous environments as Table 2.

Type	Scheme	Setting	Advantage	Drawback
EFCI	EFCI	using queue length	simplicity	high oscillation of ACR high maximum queue length seriously beat down problem
ER	EPRCA	using MACR	simplicity	high oscillation of ACR high maximum queue length beat down problem parameter tuned problem
	ERICA	using load factor and CCR	do not need per-connection information	achieve target load sensitive to CCR errors
	CAPC	using load factor and some parameters	do not need per-connection information	achieve target load parameter tuned problem
	CMM	Max-Min method using CCR	oscillation free high utilization fairness	need per-connection information sensitive to CCR errors
	TMM	Max-Min method using ER	oscillation free high utilization fairness	need per-connection information

Table 2. Comparison of the various switch schemes in the homogeneous environments.

4. Location of ER Switch

In this section, we want to investigate the location that ER switches should be placed in the mixed EFCI-ER environments. To save of space, CMM scheme is used to stand for the various ER switches because its better performance. Some simulations are done on various configurations. From the simulation results, a few rules are given to help us to place the ER switch in the EFCI-ER environments.

- N1=0, N2=2, N3=2:

Two VCs travel through the link 2 only, and the other two VCs travel through both the links 1 and 2. In this case, switch 2 is the bottleneck switch. The simulation results are shown in Figure 2. We observe that the performance is better when ER scheme is placed in the bottleneck switch. This placement can produce the same results as the case that all switches have the CMM schemes.

As we know, the ACR of all VCs are controlled by the CI bit and ER field of the backward RM cells. Meanwhile, these values are generally decided by the bottleneck switch. Hence the performance that the bottleneck switch equipped with the ER scheme is similar to the performance that all switches have ER functions. However, if the ER switches are placed on the non-bottleneck points, their functions are not fully exhibited. We can see that the placement of the ER

switch in a non-bottleneck point has no effect in our simulation.

There is another important reason to do this arrangement. Under this configuration, if the algorithm at the bottleneck switch is substantially more oscillatory than the algorithm at the non-bottleneck switch, the rate mismatch problem occurs, which causes the unfairness [14].

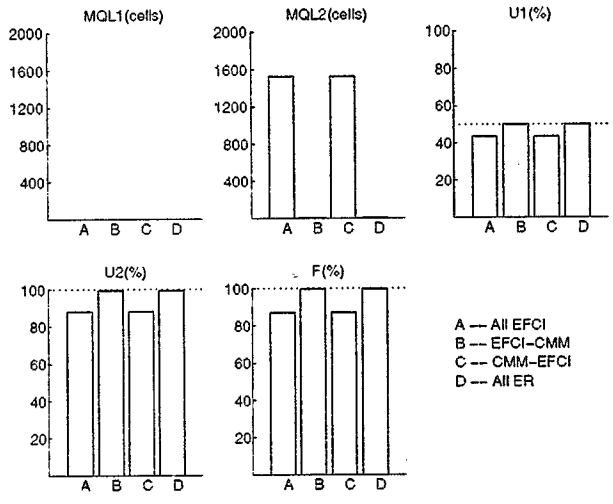

Figure 2. Comparison of the various switch schemes in the configuration (N1=0, N2=2, N3=2).

- N1=2, N2=4, N3=2:

Under this configuration, the number of VCs passing through the link 1 is 4, and the number of VCs passing through the link 2 is 6. Since some VCs only travel link 1 and some VCs travel link 2, the total traffic load is close to the link bandwidth at each switch. We can not divide them into the bottleneck and non-bottleneck switches as the above example. Hence we need to study what place should ER be settle down in this configuration.

First we define the concept of the term 'critical'. The most critical switch for a VC can be described as a switch that gives the source the lowest fair share based on the max-min fairness criteria. Similarly, the least critical switch is the switch giving the highest fair share in the VC's path. In our experiment, switch 1 is the least critical switch and switch 2 is the most critical switch.

The simulation results are shown is Figure 3. The utilization of EFCI switch for the placement of EFCI-CMM and CMM-EFCI are 93.2% and 96.1%, respectively. The utilization of ER switch for the arrangement of EFCI-CMM and CMM-EFCI are 92.0% and 83.2%, respectively. Hence the total utilization of both switch of EFCI-CMM placement is higher than that of CMM-EFCI placement. Also the fairness degree of EFCI-CMM is obviously larger than that of CMM-EFCI placement.

Hence we suggest that an ER switch ought to be placed at the most critical point first. This conclusion is completely opposite with the paper [15]. It suggest that an ER switch should be placed at the least critical point first because that can relieve the severe beat down problem more. The reason is as follows. From Table 3, the throughput of G3 improves when the least (most) critical switch is an ER switch. Also the throughput of G3 is higher when the ER scheme is equipped at the least critical switch comparing with that at the most critical switch.

The is because more bandwidth is released when the least critical switch is replaced an ER switch. Hence the VCs of G3 get more released bandwidth to lessen the beat down problem. Therefore, we should place the ER switch at the least critical point from the view of the beat down problem [15]. However, we find that the extra bandwidth G3 got is not completely equal to the bandwidth released by the other VCs at the same ER switch. Much bandwidth is wasted and just a little bandwidth is used by G3. This is the reason that more bandwidth is wasted when the ER switch is placed at the least critical point.

Therefore, we think that an ER switch ought to be placed at the most critical point first. This placement can produce the less maximum queue length, the larger fairness degree, and the higher throughput, although the beat down problem is slightly serious.

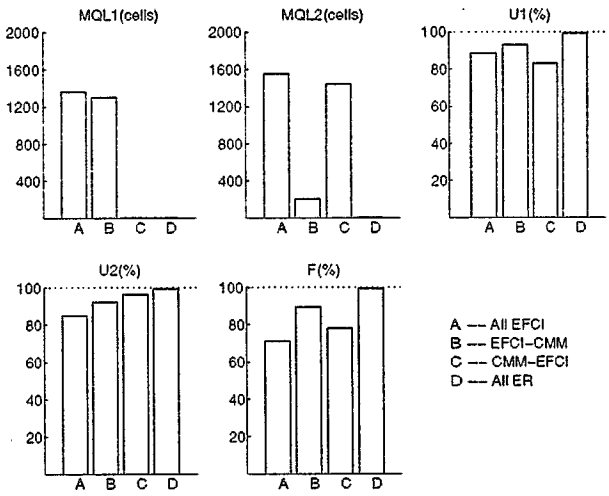

Figure 3. Comparison of the various switch schemes in the configuration (N1=2, N2=4, N3=2).

Scheme	G1	G2	G3
Ideal	51.83	25.92	25.92
all EFCI	50.35	29.10	18.47
EFCI-CMM (most critical switch is ER)	49.39	25.33	23.13
CMM-EFCI (least critical switch is ER)	40.43	26.14	24.25
all CMM	51.63	25.72	25.72

Table 3. Throughput of all groups among the various switch schemes

- N1=2, N2=2, N3=2:

In this case, the bottleneck and critical status among all switches are same. We want to investigate the place ER should locate near the source or destination. From Figure 4, the performance is better when the ER is located near the source. This is because lots of ER (except for CMM and TMM) record the information from the forward RM cell, and modify the ER field on the backward path. Hence the newer information is received at the ER switch and quicker feedback is sent to the source when the ER switch is near the source. On the other hand, the EFCI switch marks each forward data cell when congestion occurs, and does not do anything at the backward path. Hence when it is placed near the source, the propagation delay increases, which causes the slower reflection at the source.

For EPRCA and CMM, which use the CCR value recorded in the forward RM cells, the queue delay of these RM cells can be relieved when we put the ER scheme at the switch which is near the source end.

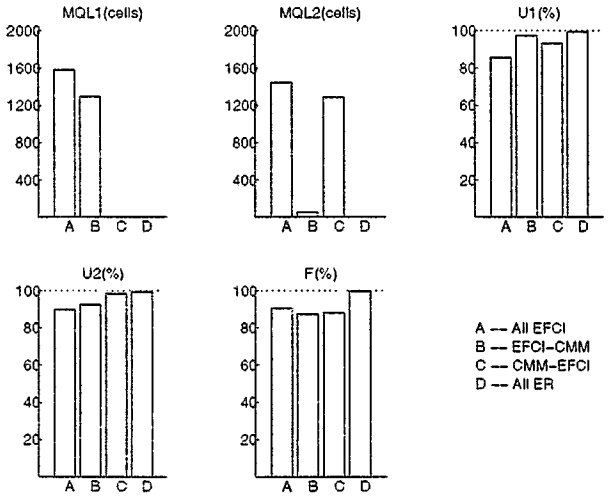

Figure 4. Comparison of the various switch schemes in the configuration (N1=2, N2=2, N3=2).

Under the configurations described as above, we can determine how to place the ER switch. However, there are some conflicts among them. For example, a backbone switch is generally a bottleneck and most critical point, also is far from the source. According to the simulation results, we set the priority: bottleneck > critical > distance. In this concept, we feel that a backbone switch is first considered to has ER capacity.

5. Conclusion

In the EFCI-ER environment, the location of ER switches is critical to the network performance. Hence we present an important rule: the ER switches should be placed at the bottleneck, most critical point and near-source location.

Reference

[1]ATM Forum, "ATM Forum Traffic Forum Traffic Management Specification Version 4.0", ATM Forum af-tm-0056.000 , April 1996.

[2]ITU-T SG 13, "Traffic Control and Congestion Control in B-ISDN," Draft Rec. I.371, Nov. 1995.

[3]H. Saito, K. Kawashima, H. Kitazume, A. Koike, M. Ishizuka, and A. Abe, "Performance Issues in Public ABR Service," IEEE Communication Magazine, pp. 40-48, Nov. 1996.

[4]H. Ohsaki, M. Murata, H. Suzuki, C. Ikeda, and H. Miyahara, "Rate-Based Congestion Control for ATM Networks," Computer Communication Review, April 1995.

[5]D. Sisalem and H. Schulzrinne, "Switch Mechanisms for the ABR Service: A Comparison Study," http://www.lirmm.fr/atm/articles.html, June 1996.

[6]A. Kolarov and G. Ramamurhy, "Comparison of ER and EFCI Flow Control Scheme for ABR Service in Wide Area Networks," IFIP Broadband Communication Proceedings, April 1995.

[7]L. Gerda and O. Casals, "A Simulation Study of Switching Mechanisms for ABR Service in ATM Networks," Lloren Cerd Olga Casals. Interim Report UPC-DAC-1996-21, Feb. 1996.

[8]N. Yin and M. G. Hluchy, "On Closed-Loop Rate Control for ATM Cell Relay Networks," IEEE Infocom'94, pp. 99-108, 1994.

[9]L. Roberts, "Enhanced Proportional Rate Control Algorithm," ATM Forum 94-0735R1, Aug. 1994.

[10]R. Jain, S. Kalyanaraman, R. Goyal, S. Fahmy, and R. Viswanathan, "The ERICA Switch Algorithm: A Complete Description," ATM Forum 96-1172, Aug. 1996.

[11]A. W. Barnhart, "Explicit Rate Performance Evaluations," ATM Forum 94-0983, Oct. 1994.

[12]A. Charny, K. K. Ramakrishnan, and A. Lauck, "Time Scale Analysis and Scalability Issues for Explicit Rate Allocation in ATM Networks," IEEE/ACM Trans. Networking, vol.4, no. 4, pp. 569-581, Aug. 1996.

[13]Danny H. K. Tsang and Wales K. F. Wong, "A New Rate-based Switch Algorithm for ABR Traffic to Achieve Max-Min Fairness with Analytical Approximation and Delay Adjustment," pp. 1174-1181, IEEE Infocom'96.

[14]N. T. Plotkin and J. J. Sydir, "The Rate Mismatch Problem in Hetergeneous ABR Flow Control", IEEE Infocom'97.

[15]Y. Chang, N. Golmie, and David Su, "Study of Interoperability between EFCI and ER Switch Mechanisms for ABR Traffic in an ATM Network," Fourth International Conference on Computer Communications and Networks, pp. 310-315, 1996.

[16]A. Arulambalam, X. Chen, and N. Ansari, "Allocating Fair Rates for Available Bit Rate Service in ATM Networks", IEEE Commun. Mag., pp. 92-100, Nov. 1996,

[17]Y. C. Lai and Y. D. Lin, "Performance Analysis of Rate-based Congestion Control Scheme and Choice of High and Low Thresholds", ICCCN'97.