論文 98-23-6-17

Block Loss Analysis of Queuing Strategy with 2-level Overload Control

Min-Kon Kwag* *Regular Member*

과부하 제어를 위한 2-단계 Queuing 전략의 블록 손실에 대한 분석

정회원 곽 민 곤*

ABSTRACT

When the superposition of realtime traffic and non-realtime traffic is applied to the ATM network, the successive cell loss(block loss) is more influential on the quality of service (QoS) of two traffic streams than the single loss in case of bursty traffic. Block loss can be identified as an important performance measure because of delay-oriented policy for realtime traffic. In this paper, we consider the system with the two-level overload control reducing of the recurrence of shut-down periods and develop a recursive algorithm to obtain both block loss and cell loss probabilities of both traffic. We can see that it gives the more precise and diverse investigations on performance analysis of queuing strategy

요 약

ATM 망에서 실시간 트래픽과 비실시간 트래픽의 중첩이 일어날 때, 군집성이 강한 트래픽의 서비스 품질은 단일 셀 손실보다는 연속적인 셀 손실에 영향을 받게 된다. 따라서, 여러 개의 셀로 구성되는 블록 손실에 대한 분석은 지연에 민감한 실시간 트래픽 군에 대해서는 성능 척도로서의 중요한 의미를 지닌다. 본 논문에서는 과부하 제어를 위한 2-단계 임계치 제어를 통한 대기 시스템을 소개하고, 각 트래픽의 단일 셀 손실과 블록 손실 분포를 규명하기 위해 기존의 분석 방법과는 상이한 재귀적 알고리즘을 소개한다. 그것은 기존의 분석 방법과는 달리 시스템의 제어 전략에 따른 성능분석에 좀 더 정확한 계산과 다양한 탐색을 가능하게 한다.

I. Introduction

Network resources in the ATM link must accommodate diverse traffic and satisfy their the quality of service (QoS) requirements. Once the buffer room is filled up, all arriving messages must be lost or delayed. In that case, QoS of all traffic streams present in the system is significantly deteriorated.

Therefore, overload control, set of all actions re-

* 현대전자 통신연구소
論文番號:98027-0115
接受日字:1998年 1月 15日

ducing the recurrence of shut-down periods, is an important factor in design of the ATM network.

Many researchers studied the overload control based on the various cell discarding schemes such as Push-out and PBS(Partial Buffer Sharing). Refer to [1], [2], [3], [4], [6] and [7]. However, those works are mainly limited to a single cell loss and made it difficult to view the block loss composed of cells. In fact, the quality of realtime traffic, such as video and voice is more sensitive to the block loss than the single cell loss. On top of the unified ATM layer, AAL provides enhanced services with block level. Recently, Cidon *et. al.* [5] investigated the block loss probability of the system without the overload control and developed the recursive algorithm for deriving it.

In this paper, we analyze the performance of the two-level overload control in a finite buffer. This model is applied to the overload control of diverse message composed of cells in an ATM link. We will develop the recursive formula to obtain the block loss and cell loss probabilities considering correlation of successive cell losses.

II. Modeling and Notations

2.1 Traffic modeling

We assume that realtime and non-realtime cells arrive according to a Poisson process with rates λ^r and λ^n, respectively. All incoming cells are stored in a commom buffer of size K. The transmission (service) time is exponentially distributed with a common rate μ for both traffic.

Under the above assumption, we denote two level thresholds for overload control by L_1 and L_2 ($L_1 < L_2 < K$) and define an overload period to be equal to the time period where the buffer contents exceed level L_2 until it drops to level L_1. The interval between two adjacent overload periods is also defined to be underload period. See Figure 1 for the operation modes. For the prevention of the congestion, the main fuction of the two-level overload control is to deter realtime cells from entering the buffer during the overload periods.

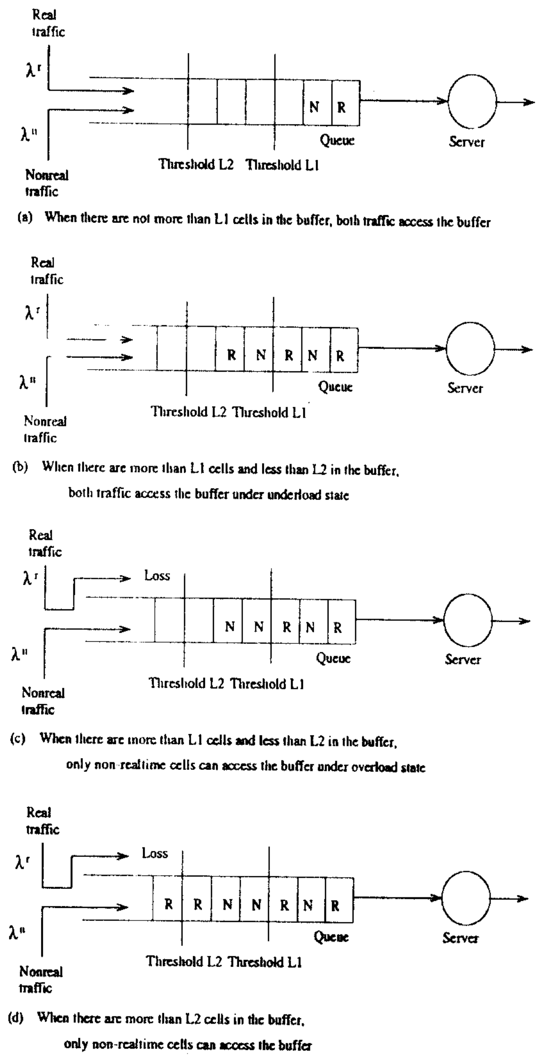

Fig. 1 Operation modes for overload control with two level thresholds

Before the development of recursion formula, we define several notations for the remainder of this paper.

- $\Pi(i)$, $i = 0, 1, ..., K$: The probabilities of having i cells in the system when a new cell arrives.

- Π_i^L, i = 0, 1, ..., L_2-1 : The probabilities of having i cells in the system when the system is in the underload period.
- Π_i^U, i = L_1+1, L_1+2, ..., K : The probabilities of having i cells in the system when the system is in the overload period.
- $Q_i(k)$, i = 0, 1, ..., K, $0 \le k \le i$: The probabilities that k cells out of i cells in the system leave the system (are transmitted) during an inter-arrival time.
- $P^r(j, n)$, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n realtime cells.
- $P^n(j, n)$, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n non-realtime cells.
- $P_i^r(j, n)$, i = 0, 1, ..., K, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n realtime cells given that there are i cells in the system just before the arrival of the first cell of the block.
- $P_i^{\bar{r}}(j, n)$, i = 0, 1, ..., K, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n realtime cells which will be arrived first afterwards given that there are i cells in the system just before the arrival of a non-realtime cell.
- $P_i^n(j, n)$, i = 0, 1, ..., K, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n non-realtime cells given that there are i cells in the system just before the arrival of the first cell of the block.
- $P_i^{\bar{n}}(j, n)$, i = 0, 1, ..., K, $n \ge 1$, $0 \le j \le n$: The probabilities of j losses in a block of n non-realtime cells which will be arrived first afterwards given that there are i cells in the system just before the arrival of a realtime cell.
- $p(r)$: The probability that an arriving cell is a realtime cell, $p(r) = \lambda^r/(\lambda^n + \lambda^n)$ and $p(\bar{r}) = 1 - p(r)$.

III. Recursive Algorithm

We first derive the equilibrium probability of the system. To this end, we devide the original controlled process into two uncontrolled subprocesses:the underload subprocess and the overload subprocess. See Figures 2. It is clear that both subprocesses are phase-type renewal processes having a single absorbing state. We define $\Omega^L = \{0, 1, 2, \cdots, L_2 - 1\}$ and $\Omega^U = \{L_1+1, L_1+2, \cdots, K\}$ for sets of transient states of two subprocesses.

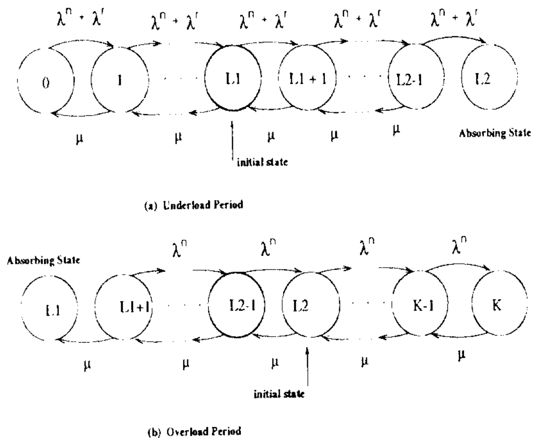

Fig. 2 State Underload/Overload Subprocesses

Let x_i^L be the mean sojourn time at state i in the underload subprocess and $x^L = \{x_0^L, x_1^L, \cdots, x_{L_2-1}^L\}$. By the arguments in [1], we get

$$x_i^L = \sum_{j=\max(L_1, i)}^{L_2-1} \frac{1}{\lambda_j} \prod_{k=i}^{j-1} \frac{\mu_{k+1}}{\lambda_k}, \qquad i \in \Omega_L. \tag{1}$$

Similarly, we can find the expected sojourn time x_i^U in overload period, which is given by

$$x_i^U = \sum_{j=L_1+1}^{\min(L_2, i)} \frac{1}{\mu_j} \prod_{k=i}^{j-1} \frac{\lambda_k}{\mu_{k+1}}, \qquad i \in \Omega_U. \tag{2}$$

Since the overall chain is positive recurrent, it may be seen as an alternating renewal process with its random epochs set at the transitions associated with

μ_{L_1+1} and λ_{L_2-1}. Both alternating renewal variables are independent and characterized by the underload and overload periods. Then, the equilibrium probability must be equal to the long-run proportion of time within an underload and overload cycle. Therefore, we have

$$\Pi_i^L = cx_i^L \qquad i \in \Omega^L, \tag{3}$$

$$\Pi_i^U = cx_i^U \qquad i \in \Omega^U, \tag{4}$$

where

$$c = \left\{ \sum_{j=0}^{L_2-1} x_j^L + \sum_{j=L_1+1}^{K} x_j^U \right\}^{-1}. \tag{5}$$

The inverse of c is the expected length of the underload/overload cycle. Therefore, the distribution of the queue length is given by

$$\Pi(i) = \begin{cases} \Pi_i^L & , 0 \le i \le L_1, \\ \Pi_i^L + \Pi_i^U & , L_1 \le i \le L_2, \\ \Pi_i^U & , L_2 \le i \le K. \end{cases} \tag{6}$$

If an arrival sees $i(L_1 < i < L_2)$ cells in the buffer, the state of the system is underload period with probability $\phi^L = (\Sigma_{j=L_1+1}^{L_2-1} \Pi_j^L)/(\Sigma_{j=L_1+1}^{L_2-1} \Pi(j))$ or overload period with probability $\phi^U = 1 - \phi^L$.

Now, consider the computation of the probabilities $P^r(j, n)$ and $P^n(j, n)$. By conditioning on the number of cells seen by the first cell in the block, we can obtain following relation.

$$P^r(j, n) = \sum_{i=0}^{K} \Pi(i)\, P_i^r(j, n), \tag{7}$$

$$P^n(j, n) = \sum_{i=0}^{K} \Pi(i)\, P_i^n(j, n). \tag{8}$$

3.1 Realtime Traffic

The initial values of $P_i^r(j, n)$ can be written down as follows.

$$P_i^r(j, 1) = \begin{cases} 1, & j = 0 \\ 0, & \text{otherwise} \end{cases}, \quad 0 \le i \le L_1, \tag{9}$$

$$P_i^r(j, 1) = \begin{cases} 1, & j = 1, \\ 0, & \text{otherwise} \end{cases}, \quad L_2 \le i \le K, \tag{10}$$

$$P_i^r(0, k) = 0, \quad L_2 \le i \le K,\ 1 \le k \le n, \tag{11}$$

$$P_i^r(n, n) = 0, \quad 0 \le i \le L_1. \tag{12}$$

When an arriving realtime cell sees $(L_1 < i < L_2)$ cells in the buffer, the cell can be entered only in the underload period. Hence,

$$P_i^r(j, 1) = \begin{cases} \phi_L, & j = 0 \\ \phi_U, & j = 1 \\ 0, & \text{otherwise} \end{cases}, \quad L_1 < i < L_2. \tag{13}$$

Let us denote $Q_i(k)$ to be the probability that $k(\le i)$ cells among i cells are transmitted during an inter-arrival time. From the assumption on the traffic, we have

$$Q_i(k) = \begin{cases} \left(\dfrac{\mu}{\mu+\lambda^n+\lambda^r}\right)^k \left(\dfrac{\lambda^n+\lambda^r}{\mu+\lambda^n+\lambda^r}\right), & 0 \le k < i \\ \left(\dfrac{\mu}{\mu+\lambda^n+\lambda^r}\right)^i, & k = i \end{cases} \tag{13}$$

Again, consider a general case for $n \ge 2$. When the first cell in a realtime block arrives and sees $i(0 \le i \le L_1)$ cells in the system, it enters the system. Then, j cells out of the next $n-1$ cells in a realtime block must be lost. If $k(0 \le k \le i+1)$ cells are transmitted during an inter-arrival time after the arrival epoch of the first cell, a new arriving cell sees $i+1-k$ cells in the system. Since a new arriving cell is a realtime cell with probability $p(r)$, we have

$$P_i^r(j, n) = \sum_{k=0}^{i+1} Q_{i+1}(k)[p(r) P_{i+1-k}^r(j, n-1) + p(\bar{r}) P_{i+1-k}^{\bar{r}}(j, n-1)], \quad 0 \le i \le L_1. \tag{15}$$

When the buffer is occupied by $i(L_1 < i < L_2)$ cells at arrival epoch of the first cell in the realtime block, the buffer accepts the first cell in underload periods but not in overload periods. Thus, we have

$$\begin{aligned} P_i^r(j,n) &= \phi_L \Big\{ \sum_{k=0}^{i+1} Q_{i+1}(k)[p(r)P^r_{i+1-k}(j,n-1) \\ &\quad + p(\bar{r})P^{\bar{r}}_{i+1-k}(j,n-1)] \Big\} \\ &= \phi_U \Big\{ \sum_{k=0}^{i} Q_i(k)[p(r)P^r_{i-k}(j-1,n-1) \\ &\quad + p(\bar{r})P^{\bar{r}}_{i-k}(j-1,n-1)] \Big\} \\ &\quad , \quad L_1 < i < L_2 . \end{aligned} \tag{16}$$

Since more than L_2-1 cells in the buffer make the realtime cell lost, we have

$$\begin{aligned} P_i^r(j,n) &= \sum_{k=0}^{i} Q_i(k)[p(r)P^r_{i-k}(j-1,n-1) \\ &\quad + p(\bar{r})P^{\bar{r}}_{i-k}(j-1,n-1)], \quad L_2 < i < K. \end{aligned} \tag{17}$$

On the other hand, the arrival of the non-realtime cell does not affect the realtime cell block size. Therefore, the equations for $P_i^{\bar{r}}(j,n)$ are

$$\begin{aligned} P_i^{\bar{r}}(j,n) &= \sum_{k=0}^{i+1} Q_{i+1}(k)[p(r)P^r_{i+1-k}(j,n) \\ &\quad + p(\bar{r})P^{\bar{r}}_{i+1-k}(j,n)], \quad 0 < i < K, \end{aligned} \tag{18}$$

and when $i=K$,

$$P_K^{\bar{r}}(j,n) = P^{\bar{r}}_{K-1}(j,n). \tag{19}$$

Using the above equations (9)--(19), we can calculate the probabilities $P_i^r(j,n)$ recursively. First, the probabilities $P_i^r(j,1)$, i=0, 1, ..., K, are computed from the initial conditions from (9) to (13). In step k, k=1, 2, ..., $n-1$, the probabilities $P_i^{\bar{r}}(j,k)$, $0 \le i \le K$, are computed from (18)--(19) and the probabilities $P_i^r(j,k)$, $0 \le i \le K$, which have been computed in step $k-1$. Then, the probabilities $P_i^r(j,k+1)$ are computed recursively from (14) to (17). For the computation of the $P_i^{\bar{r}}(j,k)$, $0 \le i \le K$, we use the general solutions for Hessenburg matrix. Therefore, we obtain the probabilities $P^r(j,n)$ from the relation (7).

3.2 Non-realtime Traffic

The initial values of $P_i^n(j,n)$ are as follows.

$$P_i^n(j,1) = \begin{cases} 1, & j=0 \\ 0, & \text{otherwise} \end{cases}, \quad 0 \le i \le K, \tag{20}$$

$$P_K^n(j,1) = \begin{cases} 1, & j=1 \\ 0, & \text{otherwise} \end{cases}, \quad i = K. \tag{21}$$

By similar arguments for realtime traffic, we have for $n \ge 2$.

$$\begin{aligned} P_i^n(j,n) &= \sum_{k=0}^{i+1} Q_{i+1}(k)[p(\bar{r})P^n_{i+1-k}(j,n-1) \\ &\quad + p(r)P^{\bar{n}}_{i+1-k}(j,n-1)], \quad 0 \le i \le K, \end{aligned} \tag{22}$$

$$\begin{aligned} P_i^n(j,n) &= \sum_{k=0}^{i} Q_i(k)[p(r)P^{\bar{n}}_{i-k}(j-1,n-1) \\ &\quad + p(\bar{r})P^n_{i-k}(j-1,n-1)], \quad i = K. \end{aligned} \tag{23}$$

Since the arrival of the realtime cell does not affect the non-realtime cell block size and the realtime cell loss depends on state i, we have

$$\begin{aligned} P_i^{\bar{n}}(j,n) &= \sum_{k=0}^{i+1} Q_{i+1}(k)[p(r)P^{\bar{n}}_{i+1-k}(j,n) \\ &\quad + p(\bar{r})P^n_{i+1-k}(j,n)], \quad 0 \le i \le L_1 , \end{aligned} \tag{24}$$

$$\begin{aligned} P_i^{\bar{n}}(j,n) &= \phi_L \Big\{ \sum_{k=0}^{i+1} Q_{i+1}(k)[p(r)P^{\bar{n}}_{i+1-k}(j,n) \\ &\quad + p(\bar{r})P^n_{i+1-k}(j,n)] \Big\} \\ &= \phi_U \Big\{ \sum_{k=0}^{i} Q_i(k)[p(r)P^{\bar{n}}_{i-k}(j,n) \\ &\quad + p(\bar{r})P^n_{i-k}(j,n)] \Big\}, \\ &\quad , \quad L_1 < i < L_2 , \end{aligned} \tag{25}$$

$$\begin{aligned} P_i^{\bar{n}}(j,n) &= \sum_{k=0}^{i} Q_i(k)[p(\bar{r})P^n_{i-k}(j,n) \\ &\quad + p(r)P^{\bar{n}}_{i-k}(j,n)], \quad L_2 \le i \le K. \end{aligned} \tag{26}$$

From the above equations (20)-(26) and the relation (7), we can calculate the probabilities $P^n(j,n)$

recursively.

IV. Numerical Analysis

We compute the distribution of the number of lost cells in a fixed size block by assuming that realtime and non-realtime streams follow the Poisson process. We consider block loss and cell loss probabilities for two traffic. Unless otherwise stated, we assume that the buffer size is 20 and the threshold values, L_1 and L_2, are 12 and 16, respectively and block size is 10. The QoS requirement for each traffic will be assumed as follows : A block more than 3 lost cells is unrecoverable for realtime traffic and a block at least 1 lost cell for non-realtime traffic. In this case, block loss probability will be $\Sigma_{i=3}^{n} P^r(i, n)$ and mean of cell loss $\frac{\Sigma_{i=0}^{n} iP^r(i, n)}{n}$ for realtime traffic. Alternative approach will be get using independence assumption. That is, block loss probability is $\Sigma_{j=3}^{n} \left[\begin{matrix} n \\ j \end{matrix} \right] p_b^{\,j}(1-p_b)^{n-j}$ where $p_b = \phi^{U} \Sigma_{i=L_1+1}^{L_2-1} \Pi(i) + \Sigma_{i=L_2}^{K} \Pi(i)$.

Independence assumption can ignore the correlation between adjacent cell losses due to bufer overflows and may lead to incorrect consecutive cell loss probabilities. Figure 3 shows the block loss and cell loss probabilities for realtime traffic by using the recursive algorithm and the independence assumption. There exists a considerable discrepancy between block losses calculated by the two methods for realtime traffic. This pheomenon results from the assumption of independence among the losses of arriving cells, which is more optimistic for realtime traffic with strong burstness.

Admissible load can be different based on the criterion of block loss requirement. Figure 4 shows the loss distributions for realtime traffic and non-realtime traffic according to the total load. As the realtime traffic ratio increases, block loss for realtime traffic decreases but not for non-realtime traffic under the fixed total load. That is, the decrease of non-realtime traffic means the relative increase of service for real-

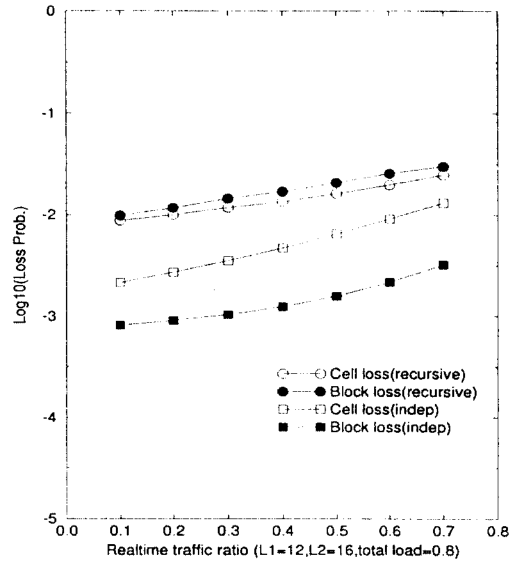

Fig. 3 Block loss and cell loss probabilities for realtime traffic

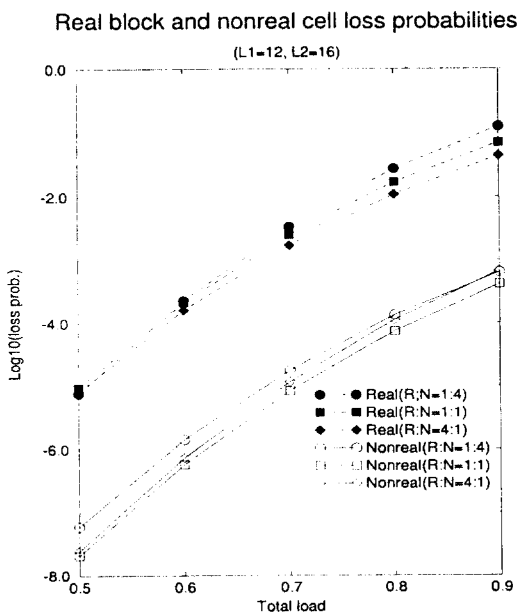

Fig. 4 Loss probability with total load

time traffic. As the total load increases largely, block loss probability increases proportionally. It means that block loss is affected by total load but not related to the specific portion of traffic type under the buffer sharing mechanism.

Priority schemes are to be compared in terms of the mean cell loss for non-realtime traffic because of loss-oriented policy. Figure 5 shows that 2-level buffer sharing scheme is little different from the PBS scheme in [3], which is shown to be efficient for space priority control of buffer management. It is the reason that all functions except thresholding are same in two methods and the effect of the threshold L_1 is poor for cell loss of non-realtime traffic.

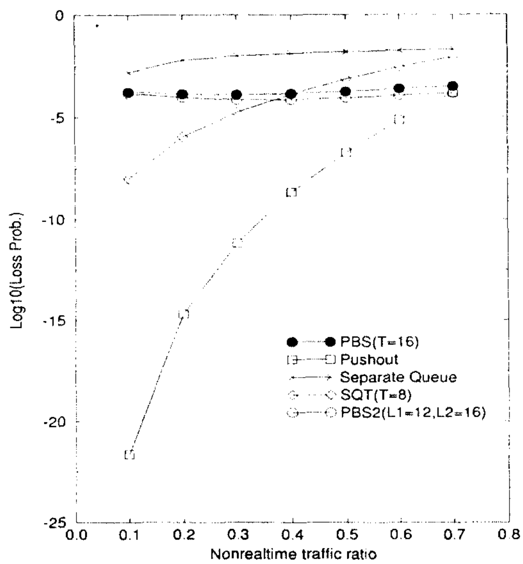

Fig. 5 Block loss probabilities for nonrealtime traffic

It is known that 2-level thresholding can provide the smoothing effect from the abrupt change of input rate on arrival process. Although research on 2-level thresholding has been developed, it is limited to single cell loss [1]. Therefore, the analytic approach with recursive algorithm is recommendable for precise computation and diversity of analysis in the distribution of lost cells whthin a block.

V. Conclusion

In this paper we analyzed the cell loss probabilities within a fixed size block with two level thresholds using a recursive algorithm. Block loss analysis by recursive algorithm can be eligible for bursty traffic because it includes the information of correlation between adjacent cell losses due to buffer overflows. Considering the performance measure in terms of admissible load, overload control through buffer sharing can be proposed to optimized the allocation of resources according to the different quality requirements of realtime and non-realtime traffic. Therefore, we can achieve a high degree of flexibility for resource utilization through two threshold values. Block loss and cell loss can be more affected by total load than by traffic ratio under the buffer sharing scheme. It means that traffic ratio can not be main factor on the admissible load in this scheme. Two level thresholding includes the characteristics of the PBS mechanism. Therefore, this scheme is more flexible than PBS mechanism since it provides the smoothing effect of input rate under congestion control. Finally, we can infer that performance analysis on 2-level thresholding is investigated more precisely by using recursive algorithm.

References

1. S. Q. Li, Overload Control in a Finite Message Storage Buffer, IEEE Trans. on Commun 37(1989) 1330-1338.
2. ITU-T Recommendation I.371, "Traffic Control and Congestion Control in B-ISDN", Geneva, May 1996.
3. H. Kröner, G. Hebuterne, and P. Boyer, Priority management in ATM switching node, IEEE J. Select. Areas in Commun. 9(1991) 418-427.
4. H. Saito, Queuing Analysis of Cell Loss Priority control in ATM Networks, ITC-13(1991) 19-24.
5. I. Cidon, A. Khamish, and M. Sidi, Analysis of packet loss processes in high-speed networks, IEEE Trans. on Information Theory 39(1993) 98-108.

6. M.K. Kwag and C.K. Kim, "An Analysis of Cell Loss Process in an ATM Network under Partial Buffer Sharing Policy", *TENCON Conference*, 1993.
7. M.K. Kwag and C.K. Kim, "Space Priority Queuing Strategy", *ICCCN Conference*, 1993 pp. 77-84.

곽 민 곤(Min-Kon Kwag) 정회원
1986년 2월:서울대학교 계산통계학과(학사)
1993년 2월:서울대학교 계산통계학과(석사)
1997년 2월:서울대학교 계산통계학과(박사)
1986년 1월~1993년 3월:엘지(LG) 정보통신 연구소
1994년 10월~현재:현대전자 통신 연구소 책임연구원
※주관심분야:ATM Traffic, 이동 통신, 통신시스템 성능분석