4

A new method for assessing the performances in ATM networks

Hoon Lee
Korea Telecom, Telecommunication Network Research Labs.
Jeonmin-dong, Yusong-gu, Daejeon, Korea
Tel:+82-2-526-6423; Fax:+82-2-526-6781
E-mail:hlee@prima.kotel.co.kr

Abstract

The authors present a new approach to evaluating the performances in ATM networks. First, we present a typical model for the ATM output multiplexer with a threshold in the queue and the priority in the bursty incoming cells. Secondly, we describe a basic concept for defining the penalty functions for the expected performance based on the status of the buffer occupancy. Finally, we present the results of numerical experiments for the proposed method, and discuss the implication.

Keywords

ATM networks, QoS, performance assessment, penalty function

1 INTRODUCTION

The real time services such as voice and video conferencing will occupy an important portion in future ATM networks. As a result, the problem of guaranteeing the time and loss related Quality of Service (QoS) to the real time services has been recognised to be significant.

One of the characteristics inherent in real time service is that it requires a finite cell delay and it can tolerate a very small portion of cell loss. Thus, the major QoS measures for it in high speed networks are the cell delay time and the cell loss rate [Ferrari(1990),Li(1989),Towsley(1993),Yuan(1989)]. If there exists a priority in a cell, the partial loss rate of high priority (HP) and low priority (LP) cell can be another QoS measure [Awater(1991)].

Performance of Information and Communication Systems U. Körner & A. Nilsson (Eds.)

The maximum cell delay can be guaranteed by providing a finite queue. In this case, the cell loss due to queue overflow follows inevitably. To guarantee the cell loss, in particular, for the HP cell of real time traffic, the partial buffer sharing (PBS) scheme has been recognised to be effective [Kröner(1993)].

In PBS scheme, it is assumed that there are two types of cell based on the loss tolerance: high-priority (HP) and low-priority (LP). A threshold is assumed in a queue. At the beginning of every time slot, the queue occupancy is observed, and if it is greater than the threshold, the LP cell is rejected to enter the queue, otherwise the LP cells can enter the queue so far as there is a vacant space in the queue. The HP cells can enter the queue so far as there is a vacant space in the queue irrespective of the state of the queue occupancy.

In this paper, we propose a new method for assessing the performance of the ATM switch which adopts the PBS scheme. The basic philosophy of the proposed method is as follows: We impose a penalty to the system based on the degree of heaviness of the buffer and the cell priority. The heavier the buffer, the more expensive the cost function, and the cost of penalty for HP cell is higher than that of the LP cell. The detailed discussion is given in section 3.

This paper is organised as follows : In section 2, we describe the system model and the queue behavior, and present a procedure for obtaining the steady state queue occupancy. In section 3, we present a method to assessing a penalty on the performance in the described system model considering the delay and loss related QoS measures. In section 4, we will present numerical results. Finally, in section 5, we summarise the paper.

2 MODEL AND ANALYSIS

2.1 System model

Consider an output multiplexer of ATM switch. ATM switch operates in discrete time basis called a time slot. A time slot is a duration to serve a fixed size cell. Assume that cells generated from multiple connections are routed uniformly to output multiplexer via a non-blocking hardware switch fabric. Since a number of cells can be routed to a specific queue in a time slot, the arrival process to a multiplexer is bursty and the cell arrival process in aggregation may be assumed to be independent on the time slot [Yegenoglu(1994)]. So, we can assume that the arrival process of to the queue has a general VBR (variable bit rate) batch distribution [Marafih(1994)].

The queue capacity is finite with size B. A threshold T is assumed to a queue, and cell input regulation of PBS scheme is based on this threshold. At the beginning of every time slot, the queue state x is observed. If $x > T$, the LP cells are discarded and only the HP cells are admitted to the queue within the available space. If $x \leq T$, both HP and LP cells are admitted to the queue within the available space.

2.2 System analysis

Consider an arbitrary time slot i, and assume that, during that time slot, cells arrive to a queue from N independent and identically distributed (i.i.d.) sources. The cell departure from the queue occurs just after the beginning of a time slot, and it is served during that time slot. Thus, the input-output principle seen from the queue is departure first.

Let X_i be the number of cells waiting in the queue just before the beginning of time slot i. Let a_i be the number of aggregated HP and LP cells which arrive during time slot i, and let b_i be the number of HP cells which arrive during time slot i. The service rule is FIFO (First-In-First-Out) and the service order for the simultaneously arrived cells in a batch is random. Since we assumed that a cell is served in a time slot, the state transition equation for the queue length between the consecutive time slots i and $i+1$ is given as follows:

$$X_{i+1} = \begin{cases} \min[\max(X_i-1,0)+\alpha_i, B], & 0 \le X_i \le T, \\ \min[\max(X_i-1,0)+\beta_i, B], & T < X_i \le B, \end{cases} \tag{1}$$

where α_i and β_i are given as follows :

$$\alpha_i = \begin{cases} a_i, & 0 \le a_i \le B-X_i+1, \\ B-X_i+1, & a_i > B-X_i+1, \end{cases} \tag{2}$$

and

$$\beta_i = \begin{cases} b_i, & 0 \le b_i \le B-X_i+1, \\ B-X_i+1, & b_i > B-X_i+1. \end{cases} \tag{3}$$

The sequence $(X_i), i > 0$, constitutes a Markov chain [Kemeny(1976)], and its state transition probability is defined by

$$p(k,l) = Pr\{X_{i+1} = l | X_i = k\}, \quad 0 \le k \le B, \ 0 \le l \le B. \tag{4}$$

If we rewrite $p(k,l)$, we have

$$\begin{aligned} p(k,l) &= Pr\{\min[\max(X_i-1,0)+\gamma_i, B] = l \mid X_i = k\} \\ &= Pr\{\min[\max(k-1,0)+\gamma, B] = l\} \end{aligned} \tag{5}$$

where $\gamma_i = \alpha_i$ when $0 \le X_i \le T$ and $\gamma_i = \beta_i$ when $T < X_i \le B$, and γ is the time independent value for γ_i since the cell arrival is i.i.d.. Similarly, we can represent α_i and β_i without the subscript i.

Then, we have, for $k=0$,

$$p(0,l) = \begin{cases} Pr\{\alpha = l\} \equiv p_l, & 0 \le l \le B-1, \\ Pr\{\alpha \ge B\} \equiv P_B, & l = B, \end{cases} \tag{6}$$

and for $1 \leq k \leq T$,

$$p(k,l) = \begin{cases} p_{l-k+1}, & k-1 \leq l \leq B-1, \\ P_{B-k+1}, & l = B, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

and for $T+1 \leq k \leq B$,

$$p(k,l) = \begin{cases} Pr\{\beta = l\} \equiv \tilde{p}_{l-k+1}, & k-1 \leq l \leq B-1, \\ Pr\{\beta \geq B\} \equiv \tilde{P}_{B-k+1}, & l = B, \\ 0, & \text{otherwise}. \end{cases} \tag{8}$$

The state transition matrix $P = \big(p(k,l)\big)$, $0 \leq k \leq B$ and $0 \leq l \leq B$, is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ T \\ T+1 \\ \vdots \\ B-1 \\ B \end{array} \begin{array}{c} \begin{array}{cccccccccc} 0 & 1 & \cdots & \cdots & T & T+1 & \cdots & B-1 & B \end{array} \\ \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & \cdots & \cdots & \cdots & p_{B-1} & P_B \\ p_0 & p_1 & p_2 & \cdots & \cdots & \cdots & \cdots & p_{B-1} & P_B \\ 0 & p_0 & p_1 & \cdots & \cdots & \cdots & \cdots & p_{B-2} & P_{B-1} \\ \vdots & \cdots & \ddots & \ddots & \cdots & \cdots & \cdots & \vdots & \vdots \\ 0 & \cdots & \cdots & p_0 & p_1 & \cdots & \cdots & p_{B-T} & P_{B-T+1} \\ 0 & \cdots & \cdots & 0 & \tilde{p}_0 & \tilde{p}_1 & \cdots & \tilde{p}_{B-T-1} & \tilde{P}_{B-T} \\ \vdots & \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \cdots & \tilde{p}_0 & \tilde{p}_1 & \tilde{P}_2 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \tilde{p}_0 & \tilde{P}_1 \end{pmatrix} \end{array}. \tag{9}$$

Let π_n be the probability that the queue length equals n in equilibrium and we denote the stationary probability vector of the Markov chain by π, $\pi = (\pi_0, \pi_1, \ldots, \pi_B)$, then π is the solution of the matrix equation given by $\pi\mathbf{P} = \pi$, $\pi e = 1$, where e is the $(B+1)\times 1$ column matrix with all elements equal to one. The equilibrium probability π can be computed by employing the standard numerical method for the matrix equation [Neuts(1981)].

3 ASSESSING PENALTIES AND PERFORMANCE MEASURES

First, let us describe a method to assessing penalties to the possible performance degradation. Next, we will describe performance measures.

3.1 Assessment of penalty

First, let us define a basic philosophy for assessing a penalty to the arriving cells. The penalty is given to the system according to the class of the cell and the position of the cell in the queue. In particular, as to the queue position, let us impose penalty value differently for HP and LP cells. For HP cells, we assume three classes: fatal, h-dangerous, and h-cautious. We denote that the system is in fatal, h-dangerous or h-cautious state when the arriving HP cell finds the queue in overflow, in between [T+1,B-1] or in between [0,T], respectively. For LP cells, we assume three classes: l-dangerous, l-cautious, and normal. We denote that the system is in l-dangerous, l-cautious, and normal state when the arriving LP cell finds the queue in overflow, in between [T+1,B-1] or in between [0,T], respectively.

Note that we denoted x-dangerous and x-cautious (x=h or l) for HP and LP cells, respectively, since the meanings for h-dangerous and h-cautious for HP cells are different from those of LP cells.

If we summarise the above definition, we obtain the following table.

Table 1 Classification of penalties

	$x \geq B$	$x = [T+1, B-1]$	$x = [0, T]$
HP cell	Fatal(black)	h-dangerous(red)	h-cautious(yellow)
LP cell	l-dangerous(red)	l-cautious(yellow)	normal(green)

Note that in Table 1 we described colors for each item for convenience of easy understanding and notation.

3.2 Penalty functions

In order to describe a penalty function to each class, let us denote as follows: ϕ_z^x be the penalty function of the system when the cell arrives to the system given that the cell is in class x and it finds the queue in region z. x has an index H and L for HP and LP cell, respectively. On the other hand, z has an index b, r, y and g for black, red, yellow, and green, respectively.

Let us denote penalty function to each case. For HP cells, we have penalty functions given as follows:

$$\phi_b^H = \alpha(x), x \geq B, \tag{10}$$

$$\phi_r^H = \beta(x), x \in [T+1, B-1], \tag{11}$$

$$\phi_y^H = \gamma(x), x \in [0, T]. \tag{12}$$

(13)

For LP cells, we have penalty functions given as follows:

$$\phi_r^L = \delta(x), x \geq B, \tag{14}$$

$$\phi_y^L = \epsilon(x), x \in [T+1, B-1], \tag{15}$$

$$\phi_g^L = \zeta(x), x \in [0, T]. \tag{16}$$

(17)

3.3 Measure of goodness

Let us define that the performance of the system is good if the following function is minimum.

$$\psi = aP_{loss} + bP_{delay} \tag{18}$$

where a and b are weighting factors between cell loss and delay performance, and P_{loss} and P_{delay} are the penalties due to cell loss and delay, respectively. Note that a and b can be used as a design parameter which can be determined considering the priority between loss and delay depending on the application under consider. P_{loss} is given as follows:

$$P_{loss} = \begin{cases} \alpha(B)\pi(B) & \text{for HP cell,} \\ \delta(B)\pi(B) + \sum_{x=T+1}^{B-1} \epsilon(x)\pi(x) & \text{for LP cell.} \end{cases} \tag{19}$$

P_{delay} is given as follows:

$$P_{delay} = \begin{cases} \sum_{x=0}^{T} \gamma(x)\pi(x) + \sum_{x=T+1}^{B-1} \beta(x)\pi(x) & \text{for HP cell,} \\ \sum_{x=0}^{T} \zeta(x)\pi(x) & \text{for LP cell.} \end{cases} \tag{20}$$

4 NUMERICAL RESULTS

In order to evaluate the goodness of the performance, we have to assume the source model as well as the penalty functions which are defined in the previous section.

4.1 Assumptions

Assume that N homogeneous and mutually independent Bernoulli like sources are superposed, and they form a bursty source which follows a binomial distribution. In each time slot a batch which is composed of HP and LP cells arrives with probability σ. The probability density function for the aggregated HP and LP cell arrivals from N sources is given by $p_n = \binom{N}{n}\sigma^n(1-\sigma)^{N-n}$. Assume that the proportion of HP cells and LP cells in a batch is the same. Then, we can obtain the probability density functions for the HP and LP cell arrivals as $q_n = p_{2n}$ and $\tilde{p}_n = p_{2n}$, respectively.

The number of source is assumed $N=40$, which corresponds to the offered load, $\rho = N\sigma$, ranging from 0.16 to 0.96 for $\sigma = 0.004$ to 0.024. The queue size is assumed to be $B = 30$ and the threshold is assumed to be $T = 25$. These assumptions on the parameters are effective unless they are specified explicitly.

As to the penalty functions, we can have a tremendous different kind of functions. The assessment of the system performance depends directly on the type of the penalty function. So, we have to be very cautious in assuming them. However, we do not know which type of function is best suited in assessing the performance in ATM switching system. In reality, we have no alternatives except determining them intuitively. The only intuition we can have is as follows: A customer (cell in this case) will feel worriness as the resource (available buffer space in this case) becomes scarce. We also think that the degree of worriness will increase exponentially even though the resource dries up linearly. A more detailed discussion about this intuition is described in [Lee(1998)]. Thus, let us assume as described in Table 2:

Table 2 Penalty functions

function	values
$\alpha(x)$	$\alpha = 10$
$\beta(x)$	sigmoid function (defined in (21))
$\gamma(x)$	$\gamma = 0$
$\delta(x)$	$\delta = 1$
$\epsilon(x)$	$\epsilon = 1$
$\zeta(x)$	sigmoid function (defined in (22))

Note that we assumed a ten-fold weight on α with respect to δ since the overflow of HP cells is more serious than that of LP cells. The value of γ is given zero since there is no trouble for HP cells to find the queue to be light. As to the functions $\beta(x)$ and $\zeta(x)$ we will use sigmoid functions, which are defined as follows:

$$\beta(x) = \frac{3.8}{1 + e^{-(x-T-6)}}, \tag{21}$$

$$\zeta(x) = \frac{1}{1 + e^{-0.5(x-T/2)}}. \tag{22}$$

The curves $\beta(x)$ and $\zeta(x)$ are obtained by empirical manipulation. Figure 1 illustrates the curves for the $\beta(x)$ and $\zeta(x)$ for $T = 25$ and $B = 30$.

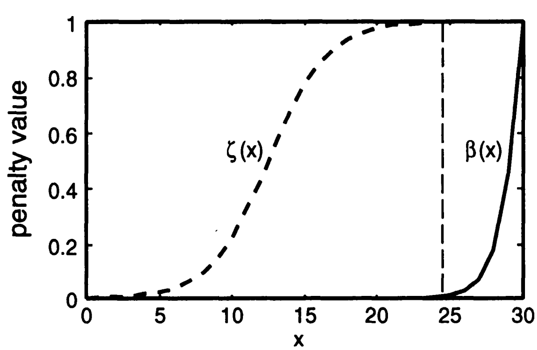

Figure 1 Curves for penalty functions.

4.2 Results and discussion

Figure 2 illustrates the loss penalty functions for HP and LP cells. Note that

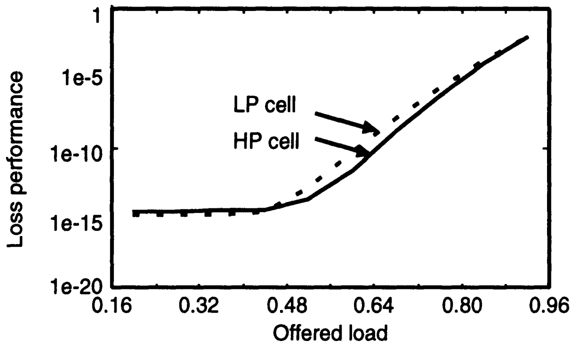

Figure 2 Loss performance

the two curves cross at the point of offered load of 0.44. When the offered load is less than 0.44, the two curves almost coincides. .However, for the offered load greater than 0.44 the cost reverses. This trend illustrates that the offered load should be limited to a certain value if one wants to obtain a certain level of performance from the system.

Figure 3 illustrates the delay performance for HP and LP cells. There

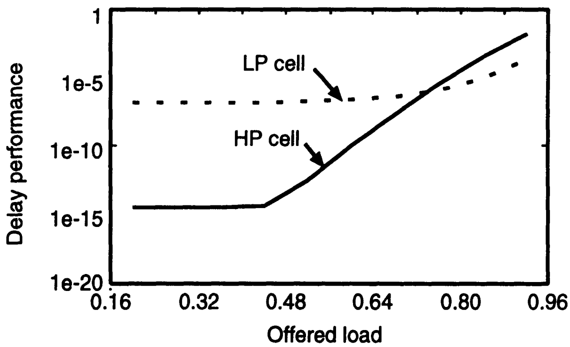

Figure 3 Delay performance.

exists a great difference in the cost of two curves for the lightly loaded case. As the offered load increases, they approach each other, and when the offered load is 0.73 they reverses. So, for highily loaded system, the number of HP cells that imposed penalty to the system may be greater than that of LP cells. For lightly loaded system, the reverse holds.

Note that the loss performance for HP cell and the delay performance of HP cell has the same order. So, we can assume that the simplest selection is assignning the same value for the coefficients a and b; that is $a = b = 1$.

Figure 4 illustrates the overall penalty functions for HP and LP cells. From Figure 4 we can deduce that the cost imposed by the LP cells for

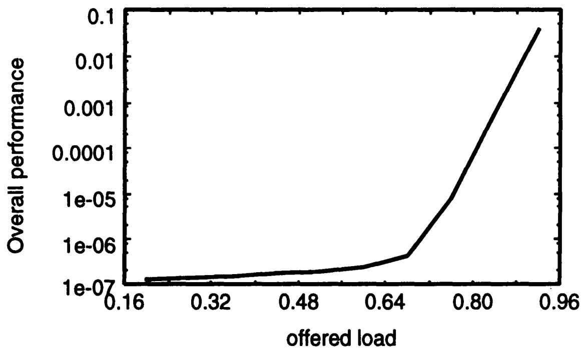

Figure 4 Overall performance

delaying in a buffer is dominant among other ones.

5 CONCLUSIONS

We presented a new approach to assessing the penalties and evaluating the performance of the ATM switch from the view point of the cell loss and delay of HP and LP cells, respectively. First, we presented a system model under PBS scheme using the finite capacity queue with a threshold. Next, we defined a new measure of system penalty function, and investigated the goodness of the system using the proposed method.

From the numerical experiments, we obtained the following results : The proposed method can clarify the weight of the system performance with respect to the position in a buffer as well as the class of cell loss priority. The proposed measure can measure the loss and delay performance of the system simultaneously and with different weight.

Therefore, we expect that the results can be applied to the design and control of the output buffer for the ATM switch.

6 REFERENCES

Awater, G.A. and Schoute, F.C. (1991) Optimal queueing policies for fast packet switching of mixed traffic. *IEEE Journal on Selected Areas in Communications*, vol.9, no.3, pp.458-467.

Ferrari, D. (1990) Client requirements for real-time communication services. *IEEE Communication Magazine*, pp.65-72.

Kemeny, J.G. and Snell, J.L. (1976) Finite Markov Chains. Springer-Verlag, New York.

Kröner H., Hébuterne G., Boyer P. and Gravey A. (1993) Priority Management in ATM switching nodes. *IEEE Journal on Selected Areas in Communications,* Vol. 9, No. 3, pp.418-427.

Lee Hoon (1998) A new performance assessment method in ATM networks. *Proceedings of the international conference on probability and its applications,* February 24-26, Korea.

Li, S.Q. (1989) Study of information loss in packet voice systems. *IEEE Transactions on Communications,* vol.37, no.11, pp.1192-1202.

Marafih, N.M. and Zhang, Y-Q. (1994) Modeling and queueing analysis of variable-bit-rate coded video sources in ATM networks. *IEEE Transactions on Circuits and Systems for Video Technology,* vol.4, no.2, pp.121-128.

Neuts, M.F. (1981) Matrix-geometric solutions in stochastic models. The Johns Hopkins University Press, Baltimore and London.

Towsley, D. (1993) Providing quality of service in packet networks. *Performance evaluation of computer and communication systems (ed. L. Donatiello, R. Nelson),* pp.560-586, Springer Verlag.

Yegenoglu, F. (1994) Characterization and modeling of aggregate traffic for finite buffer statistical multiplexers. *Computer Networks and ISDN Systems,* 26, pp.1169-1185.

Yuan, C. and Silvester, J.A. (1989) Queueing analysis of delay constrained voice traffic in a packet switching system. *IEEE Journal on Selected Areas in Communications,* vol.7, no.5, pp.729-738.

7 BIOGRAPHY

Hoon Lee obtained B.E. and M.E. in electronics in 1984 and 1986, respectively, both at Kyoungpook National University, Daegu, Korea. He obtained Ph.D. in electrical and communication engineering in 1996 at Tohoku University, Sendai, Japan. Since 1986, he has worked on network planning, dimensioning, and teletraffic engineering, with a particular interest in the field of traffic modeling, traffic control, performance analysis, and QoS guarantee of ATM networks. Dr. Lee is a member of IEEE and IEEK of Korea.