Approximate Cell Loss Performance in ATM Networks : In Comparison with Exact Results

Hoon Lee* ***Regular Member***

ABSTRACT

In this paper we propose an approximate method to estimate the cell loss probability(CLP) due to buffer overflow in ATM networks. The main idea is to relate the buffer capacity with the CLP target in explicit formula by using the approximate upper bound for the tail distribution of a queue. The significance of the proposition lies in the fact that we can obtain the expected CLP by using only the source traffic data represented by mean rate and its variance. To that purpose we consider the problem of estimating the cell loss measures from the statistical viewpoint such that the probability of cell loss due to buffer overflow does not exceed a target value. In obtaining the exact solution we use a typical matrix analytic method for GI/D/1/B queue where B is the queue size.

Finally, in order to investigate the accuracy of the result, we present both the approximate and exact results of the numerical computation and give some discussion.

I. Introduction

The main feature of ATM switch with asynchronous cell transfer capabilities lies in the statistical multiplexing of cells in a buffer. Since the buffer size is finite in a real system and there is a case when the number of cells bound for a particular output buffer from multiple inputs exceeds the available queue space in an arbitrary time slot, the buffer overflow occurs and a part of arriving cells are lost.

There are two ways to quantify the cell overflow for a buffer in ATM switching system : finite queue model and infinite queue model. The former can be obtained by G/D/c/B queueing model, where B is the buffer size[5]. However, in that case, we have to compute the steady-state probability for the queueing system(See the Appendix), which incurs a high cost for computation if B is very large. Also, one can not represent and explicit relationship between the source traffic parameter, the CLP value and the buffer size from the finite queue model.

The infinite queue model can be used as an approximation for the finite one. If the required CLP value for a source is assumed k for a queue capacity B, then the formula for the upper bound of the tail distribution of the queue contents x

$$\text{Prob}\{x \geq B\} \leq k, \tag{1}$$

where x is the queue length at observation point, can be used as an approximation of a CLP formula for a finite queue model[6].

The seminal works on the analytical model dealing with the approximations for the upper bound of the tail distribution of the queue contents have been proposed by literature [4,6]. Hui[3] applied the large deviation techniques to evaluate the probability of burst in the broadband networks.

Guérin et al.[2] made comprehensive discussions for the approximate analysis to the equivalent capacity for the networks, whose concept is similar to the statistical QoS measures.

* Korea Telecom, Telecommunications Network Laboratory
논문번호 : 99177-0506, 접수일자 : 1999년 5월 6일

The purpose of this paper is to use the large deviation principle and the concept of statistical QoS measures to obtain a simple formula for the upper bound of the buffer overflow probability (thus, the cell loss probability) such that the computational complexity does not scale with the capacity of the buffer and the arrival process. This can be realised by an approximation by exponential upper bound and the use mean and variance of arrival rate rather than the conventional parametric traffic description such as ON-OFF process.

This paper is organised as follows : In section Ⅱ, we describe the points of the discussion of this paper. Section Ⅲ presents the procedure for deriving the upper bound on the tail probability of GI/D/c queue. In section Ⅳ, we discuss about the cell loss measures. In section Ⅴ the discussion about the numerical results are given. Section Ⅵ summarises the paper.

Ⅱ. Problem description

As mentioned in the previous section, the swich should guarantee the cell rate and delay requirement to a connection under any situation during its connection. The maximum cell delay can be guaranteed by the size of the buffer. So, in designing and managing the ATM networks, one of the main performance issues should be taken into account is to guarantee the cell loss Quality-of-Service(QoS) negotiated during the contraction phase or to dimension the buffer(alternatively, queue) size in order not to violate the negotiated cell loss QoS.

This paper describes the issue, that is, we consider a method to estimate the cell loss performance in ATM networks for an arbitrary small buffer size with heavy offered load in doing that we assume a switch buffer with random traffic source.

In [8] it is stated that the random traffic source gives a worse CLP for the same load generated by CBR of VBR sources. So, we can a random traffic as a good upper bound for the buffer estimation

The main idea is to estimate the CLP performance for and arbitrary buffer from both the source traffic data represented by mean rate and its variance and the service capacity of the switch.

Following the ATM Forum's traffic description, the source traffic can be described in parameters such as the mean cell rate, peak cell rate, etc. However, the typical and simplest parameters which the system designer can obtain from a sample of measured sources are the mean rate and its variance.

We will not deviate our discussion by discussing the method to obtain the data for mean rate and the variance. It can be deduced from the previous experience on the traffic of similar kind or it can be computed from the original source traffic parameters

To our purpose, the statitical description as shown in the formula (1) is useful, and we will propose a method to derive a relationship between the source traffic profile, the channel capacity of the network, and the required QoS targets in the subsequent sections.

Ⅲ. Upper bound for the buffer tail probability

If we investigate the formula (1) we can find that there are three factors that are closely related to the CLP: the source characteristics, the service rate, and the queue capacity. The service rate is assumed to be a fixed value of c cells per time slot. The source characteristics are not shown explicitly in that formula, but it is embedded in the formula for the k.

Our concern in this paper is to represent the relationship between the service rate c and the queue capacity B that should be provided in order to guarantee the target cell loss QoS under the condition that the source traffic profiles are given.

Suppose that time is divided into slots of unit length which is the time required to transsmit a cell, and a batch of cells which arrive during each slot follows a general renewal process. First, let us define some variables concerning the system. Let the random variable x_k be the queue length at the boundary of time slot k, k=1, 2, ..., and we assume that it is zero just before the start of the slot 1.

Let the i. i. d. random variable y_k be the number of cells arrived during time slot k. Assume that c cells depart the queue just before the end of time slot k, if there is any. So, the order of cell input-output is arrival first. If we assume a with heavy load c cells can depart the queue every time slot with probability almost the same as 1.

We can obtain a formula for the number of cells remained in the queue at the beginning of time slot k, x_k, given by

$$x_k = [x_{k-1} + y_{k-1} - c, 0]^+, \tag{2}$$

where $[z, 0]^+$ denotes the larger number between z and zero.

Note that, under the abover assumptions, when the service discipline is FCFS (First Come First Served), x_k repesents an unfinished amount of work (unit : cell) in the queue seen by a first arriving cell of time slot k, which corresponds to the waiting time of the arriving cell in the queue.

Let us define the variation of the queue length in a time slot k to be

$$v_k = y_k - c. \tag{3}$$

From the stability condition v_k has to satisfy $\mathbf{E}[v_k] < 0$. Because $\mathbf{E}[v_k]$ is finite v_k has the moment generating function (mgf)[4]

$$v(\theta) = \mathbf{E}[e^{\theta v_k}]. \tag{4}$$

Kingman proposed the tail distribution of the queue occupancy for the GI/GI/1 queue[4]. If we rewrite it in discrete-time context and for the GI/D/c queue, where c is the unmber of the server, we have, for $k \to \infty$, the following formula for the tail distribution of the queue occupancy x in equilibrium :

$$Prob\{x \geq B\} \leq e^{-\theta B}, \tag{5}$$

where θ satisfies

$$\theta = \inf[\theta, \theta > 0 : \mathbf{E}[e^{\theta v_k}] < 1] \tag{6}$$

which is translated into the minimum of the positive solution of the inequality

$$\mathbf{E}[e^{\theta v_k}] < 1, \ 0 \leq \theta < \infty \tag{7}$$

In formula (5), θ mainly determines the behavior of buffer overflow (the derivative of the curve) and so it is called the decay rate. The formula (5) is considered to be a little conservative, but it is widely used in a approximate analysis because if its advantage of simplicity in computation. The decay equation (7) can be represented in a closed formula for a few specific arrival process such as a Poisson binomial process.

In order to represent the decay rate in an explicit formula for the arrival process without explicit distribution function, we have to develop a more tractable means, which is the main purpose of this section.

Let us rewrite (7) into

$$\mathbf{E}[e^{\theta v_k}] < e^{c\theta} \tag{8}$$

In order to obtain a formula for θ which satisfies the inequality (8), let us introduce an error ϵ such that the following equation is satisfied :

$$\mathbf{E}[e^{\theta v_k}] + \epsilon = e^{c\theta} \tag{9}$$

Now that ϵ should have a very small value of the difference between $\mathbf{E}[e^{\theta v_k}] - e^{c\theta}$. For that we essume that ϵ is a small fraction of $e^{c\theta}$, and it is given as follow :

$$\epsilon = \xi e^{c\theta}. \tag{10}$$

where ξ is a very small value ($\xi << 1$).

Let the mgf of the cell arrival $y(\theta) = \mathbf{E}[e^{\theta v_k}]$.

Using the power series expansion of the exponential function, $y(\theta)$ can be represented as follos :

$$y(\theta) = 1 + E[y]\theta + \ldots + E[y^n]\frac{\theta^n}{n!} + \ldots, \qquad (11)$$

where $E[y^n]$ is the n-th moment of random variable y.

We have an information for the mean and variance of the source traffic, so we truncate the series (11) into the order of two (term with the second moment), and let $\lambda = E[y]$ and let $\sigma^2 = E[y^2] - \lambda^2$, we obtain

$$y(\theta) \approx 1 + \lambda\theta + \frac{\sigma^2+\lambda^2}{2}\theta^2 \qquad (12)$$

On the other hand, the equation (9) is alternatively represneted by

$$y(\theta) + \varepsilon = e^{c\theta} \qquad (13)$$

Then, from (10), (12), (13), and also truncating the power series expansion of the formula $e^{c\theta}$ into second order of θ, we have a unique positive solution given by

$$\theta = \frac{(1-\xi)c - \lambda + \sqrt{\varphi}}{\sigma^2 + \lambda^2 - (1-\xi)c^2} \qquad (14)$$

where $\varphi = (\lambda - (1-\xi)c)^2 - 2\xi(\sigma^2 + \lambda^2 - (1-\xi)c^2)$

For the case of a more rough approximation of ε equal to zero, refer to [7]

If we put the formula (14) into (5) we can obtain an explicit formula for the upper bound of cell loss probability due to buffer overflow described by the service capacity and the source traffic profile.

Ⅳ. Discussion on Cell Loss Measures

As we assumed the CLP as the QoS measure, we will relate them with the result for tail distribution of the queue occupancy given is the previous section, This is represented in inequality (1), which means by the probability that the queue length x in equilibrium exceeds the threshold value B be not greater than k, From the inequalities (5) and (1) we obtain a formula for the upper bound on the CLP as a function of queue capacity B, which is given by

$$k = e^{-c\theta} \qquad (15)$$

Alternatively, we can rewrite a formula for the required buffer size from the above formula (Refer to [7] for more detailed discussion and some results for B as a function of CLP). The physical meaning of the last phrase is as follows : The queue capacity required by an ATM switch is easily calculated if one only knows the traffic parameters, the target CLP value for the source, and the channel capacity.

On the other hand, if we consider an accurate solution for the above mentioned problem, it can be obtained by using a finite queueing system GI/D/C/B, and the CLP is simpley given by a probability that the buffer is full. So, if the equilibrium probability that the buffer is full is given by $\pi(B)$, this corresponds to the CLP itself. For detailed procedure for obtaining $\pi(B)$, refer to Appendix.

Note, however, that for the finite queueing model we can not represent an explicit relationship between the CLP target and the buffer size as well as the source traffic profile, whereas we could represent it explicitly in infinite queueing model. This is the main motivation of considering the approximation for infinite queueing model.

In addition, we can also calculate the required channel capacity c if the traffic parameters and the target CLP value for the source as well as the buffer capacity B is known [7].

Ⅴ. Numerical Results and Discussion

In order to compare the performance obtained by the approximate and exact ananlyses, we have to assume an explicit and analytical cell arrival process to the multiplexer, So, let us assume that the aggregated input to the output buffer of ATM multiplxer from N sources is assumed to follow a binomial process, and we denote by the mean number of the binomial arrival process and its variance $\lambda = Nr$ and $\sigma^2 = Nrs$, respectively, where r is the probability that a cell is generated by a source in a slot, and $s = 1 - r$. The associated sources, with number equal to N, are assumed to have a very long holding time compared with a microscopic cell time slot, and it is assumed to be N=50. The service rate is assumed to be c=1.

Usually, the cell arrival in ATM network is bursty. For the result assuming more burstier process than the binomial one, refer to [7] by authors.

Fig. 1 illustrates the CLP versus the offered load for the buffer capacity B = 10. In Fig.1 three results are shown : Exact, Approx1, and Approx2, which is the result obtained by the exact solution, the approximate solution with $\xi = 0$, and the approximate solution with $\xi = 0.001$, respectively.

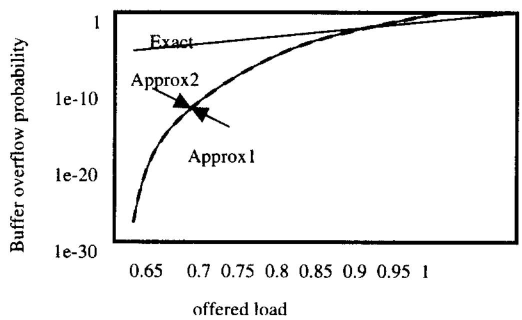

Fig. 1 CLP versus offered load for B=10.

As for the two approximate solutions, we could not find much difference between them. So, the assumption $\xi = 0$ can be applied for the simplicity of the computation.

From Fig. 1 we can find that the CLP obtained by approximate method sees the cell loss optimistically or pessimistically depending on the load in the buffer. When the buffer is lightly loaded the approximate method estimates the CLP more optimistically than that obtained by exact solution.

However, as the offered load increases to about 0.875 the two results almost coincide, after that the approximate results overstimates (sees pessimistically) the CLP compared with the exact result. This trend can be observed for the case of B greater than 10, which is presented in the following figures.

Fig. 2 illustrates the CLP versus the offered load for the buffer capacity $B = 20$. We can find that the CLP for $B = 20$ follows a little more steep curvature than that for $B = 10$, and with different order. We can find this trend for $B = 30$, too.

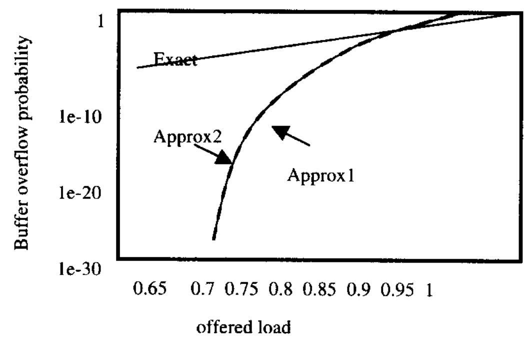

Fig. 2 CLP versus offered load for B = 20

Fig. 3 illustrates the CLP versus the offered load for the buffer capacity $B = 30$.

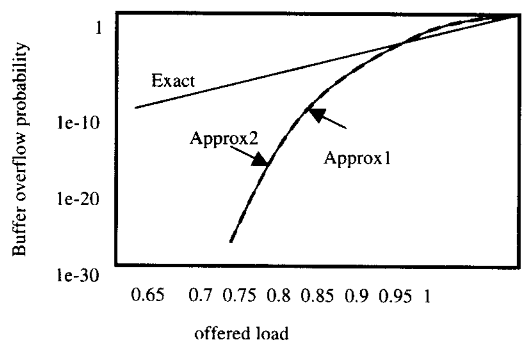

Fig. 3 CLP versus offered load for B = 30

From the three figures presented above, we can find one more fact : the crossover point between the approximate solution and the exact solution shifts toward right as the buffer capacity is becomes large.

We also found that the proposed method could be useful only the very heavily loaded network.

Ⅵ. Conclusion

In this paper, we presented a method for estimating the CLP in an ATM switch with output buffer and compared its results with those obtained by the exact solution.

In obtaining the approximate solutio for the tail distribution of the buffer occupancy, we used a large deviation approximation method for estimating the tail distribution of GI/D/c queue.

Via numerical experiment we investigated the accuracy of the proposed method, and found that the approximate solution can estimate relatively, well the CLP in high load regions. More over, even though we did not show the required buffer capacity for the target CLP value (refer to [7] for this mater), the proposed method can represent an explicit formula for the buffer capacity needed to guarantee the CLP as a function of traffic profile and CLP. This is a very strong point of the proposition compared with the exact solution.

The result of this paper can be used in the estimation of the buffer performance from the ststistical aspects for the high performance communication networks in which the error such as the cell loss due to queue overflow is considered to be a rare event and we have only the traffic profile of mean rate and its variance, which is actually the most probable case in the field.

Appendix

Let us begin from equation (2). It is easily found that the sequence (x_k), $k > 0$, constitutes a Markov chain[5], and its state transition probability is defined by

$$p(i,\ j) = Pr\{(x_{k+1} = j \mid x_k = i\} \qquad 0 \le i \le B,\ 0 \le j \le B \tag{16}$$

Then, we have, for $0 \le i \le c$,

$$p(i,\ j) = \begin{pmatrix} Pr\{y=j\} \equiv p_j, & 0 \le j \le B-1, \\ Pr\{y \ge B\} \equiv P_B, & j+B, \end{pmatrix} \tag{17}$$

and for $c+1 \le i \le B$,

$$p(i,\ j) = \begin{pmatrix} p_{j-i+i}, & i-1 \le j \le B-1, \\ P_{B-i+1}, & j = B, \\ 0, & otherwise, \end{pmatrix} \tag{18}$$

The state transition matrix $P=(p(i,\ j))$, $0 \le i \le B$ and $0 \le j \le B$, is given by

$$P = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ c \\ c+1 \\ \vdots \\ B-1 \\ B \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & \cdots & \cdots & \cdots & B\text{-}1 & B \end{array} \\ \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & \cdots & p_{B-1} & P_B \\ p_0 & p_1 & p_2 & \cdots & \cdots & p_{B-1} & P_B \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ p_0 & p_1 & p_2 & \cdots & \cdots & p_{B-1} & P_B \\ 0 & p_0 & p_1 & \cdots & \cdots & p_{B-2} & P_{B-1} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & p_0 & p_1 & P_2 \\ 0 & \cdots & \cdots & \cdots & 0 & p_0 & P_1 \end{pmatrix} \end{array} \tag{19}$$

If we let $\pi(n)$ be the probability that the queue length equals n in equilibrium and we denote the stationary probability vector of the Markov chain by π, $\pi=(\pi(0),\ \pi(1),\ \cdots,\ \pi(B))$, then π is the solution of the matrix equation given by $\pi \mathbf{P} = \pi$, $\pi e = 1$, where $\mathbf{P}$ is the state transition matrix of the queue occupancy and e is the $(B+1)\times 1$ column matrix with all elements equal to one.

References

[1] Y.J. Choi and H.M. Kim, "Optimal bandwidth and buffer allocation method for heterogeneous traffic with multiple QoS in ATM networks," *Electronics Letters,* 17th September 1998, Vol. 34, No. 19, pp. 1822~1823.

[2] R. Guérin, H. Ahmadi and M. Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks", *IEEE J. Select. Areas Commun.,* Vol. 9, No.7, pp.968-981, Sep. 1991.

[3] J. Y. Hui, "Resource Allcation for Broadband Networks," *IEEE J. Select. Areas Commun.,* Vol. 6, No 9, pp. 1598-1608, Dec. 1988.

[4] J.F.C. Kingman, "Inequalities in the Theory of Queues," *J. Roy. Stat. Soc.,* Series B, 32, pp. 102~110, 1970.

[5] L. Kleinrock, *Queueing Systems Volume 1 : Theory,* John Wiley & Sone Inc., New York,

1975
[6] H. Kobayashi and A.G.Konheim, "Queueing Models for Computer Communications System Analysis," IEEE Trans. Commun. Vol. COM-25, No.1, pp.2-29, Jan 1977.
[7] Hoon Lee, "Dimensioning the output buffer in ATM networks," To be presented at ICT '99, June 15-18, 1999, Cheju, Korea.
[8] J.M. Pitts and J.A.Schormans, *Introduction to ATM design and performance*, John Wiley & Sone Inc., New York, 1996.

이 훈(Hoon Lee)

Feb. 1984 : B.E. in electronics Kyungpook National University, Daegu, Korea

Feb. 1986 : M.E. in Communications from Kyungpook National University, Daegu, Korea

March 1996 : Ph.D. in Electrical and communication engineering from Tohoku University, Sendai, Japan

In Feb. 1986, he joined Korea Telecom, where he has been engaged in the research on the teletraffic theory, network planning, traffic modeling and management of broadband computer and telecommunication networks.

He is now a senior research staff in Telecommunications network laboratory of Korea Telecom, Daejeon, Korea.

His current research interests include the performance analysis, QoS guarantee, pricing strategies and design / control of multi-service broadband networks.

Dr. Lee is a member of IEEE, KICS and IEEK