論文 98-23-9-6

ATM 망에서의 통화품질 평가를 위한 근사화 기법과 이를 이용한 호 수락 제어

정회원 안 동 명*, 한 덕 찬*

A Call Admission Control in ATM Networks using Approximation Technique for QOS Estimation

Dongmyung Ahn*, Duckchan Han* *Regular Members*

요 약

호 수락 제어는 호 설정 단계에서 실행되는 밀집 제어 메카니즘 중 하나로 호 수락 여부를 실시간으로 결정할 수 있는 효율적인 통화 품질 예측 방법이 요구된다.

본 논문에서는 지연과 손실에 민감한 호의 수락 제어를 위한 셀 손실률과 평균 셀 지연 시간을 평가하는 간단한 근사 방법을 제안하였다. 통화 품질예측을 위한 큐잉 모델로서 새 호가 기 연결된 호들과 합성되는 합성입력(mixed input) 큐잉 시스템을 이용하였으며 이러한 호 들의 트래픽 특성을 나타내기 위하여 새로운 트래픽 파라메타를 제안하였다. 이 트래픽 파라메타에 의한 maximum entropy 해석과 renewal 근사기법을 사용하여 셀 손실율과 지연 시간을 closed form으로 구하였다. 또한 제안된 통화 품질 평가 방법을 이용하여 실시간 계산이 가능하고 새 호의 트래픽 형태에 무관하게 사용할 수 있는 호 수락 제어 기법을 소개하였다. 평균 지연과 셀 손실 확률은 적절한 근사화 식을 사용하여 결정하였다.

제안된 근사 기법의 정확성을 검증하기 위해 실제 ATM 상황하에서의 트래픽 모델을 사용하여 다양한 합성입력 큐잉 시스템을 해석 또는 시뮬레이션 하였으며 제안된 근사기법과의 비교를 통해 정확성을 입증하였다.

ABSTRACT

Admission control is one of the most important congestion control mechanism to be executed at the call set up phase by regulating traffic into a network in a preventive way. An efficient QOS evaluation or bandwidth estimation method is required for call admission to be decided in real time.

In this paper, we propose a computationally simple approximation method of estimating cell loss probability and mean cell delay for admission control of both delay sensitive and loss sensitive calls. Mixed input queueing system, where a new call combines with the existing traffic, is used as a queueing model for QOS estimation. Also

* 국방과학연구소
論文番號 : 98075-0219
接受日字 : 1998年 2月 19日

traffic parameters are suggested to characterize both a new call and existing traffic. Aggregate traffic is approximated by a renewal process with these traffic parameters and then mean delay and cell loss probability are determined using appropriate approximation formulas.

The accuracy of this approximation approach is examined by comparing their results with exact analysis or simulation results of various mixed input queueing systems. Based on this QOS estimation method, call admission control scheme which is traffic independent and computable in real time are proposed.

I. Introduction

Asynchronous transfer mode (ATM) has been widely accepted as a target technology to implement broadband integrated services digital network (B-ISDN) [1], [2]. ATM provides flexibility for supporting services each with a wide range of service requirements and various statistical traffic behaviors. Especially, it is expected that a large number of bursty traffic sources (voice, interactive data, compressed video, image, etc.) are supported in an ATM network. Statistical multiplexing of such traffics generated from several sources may lead to efficient bandwidth utilization and allow more calls to be handled.

However, severe network congestion and performance degradation may easily occur because of the dynamic behavior of traffic sources and the statistical multiplexing scheme. Traffic control is hence required to avoid the congestion and to allow for high utilization of network resources, while guaranteeing the quality of service (QOS) requirement for all connections. Since ATM is a connection-oriented high speed transport technique, various levels (path level, call level, cell level) of both preventive and reactive control methods are commonly considered suitable for an ATM network [3],[4],[5].

Admission control is one of the most important congestion control mechanisms to be executed at the call set up phase by regulating the calls into a network in a preventive way. When a new call arrives, the admission control function estimates the QOS (mean cell delay and cell loss probability, etc.) or required resources (link bandwidth and buffer) along the path between the source and the destination nodes. The estimation is based on the traffic characteristics of existing calls and a set of traffic descriptors, as well as the required QOS specified by a new call. Admission control accepts a new call if the available resources are able to meet the required QOS of a new call along with those of the existing calls. Hence, the method of evaluating the QOS performance or estimating the required resources are the most important issues of the admission control problems still under consideration. The main difficulty comes from the lack of accurate and computationally simple analytical models that can be used to evaluate the performance of multiplexed traffic, consisting of many heterogeneous calls with a wide range of service requirements and diverse traffic characteristics.

Various call admission control strategies have been proposed to evaluate the QOS performance or estimate the effective bandwidth using traffic descriptors such as peak bit rate, average bit rate, bit rate variance or average burst length, etc. Simple M/D/1/K model has been proposed by [6]. In this method, a new call is characterized by peak bit rate (wideband traffic) or average bit rate (narrowband traffic) and the overall aggregate traffic to the link is assumed to be Poisson with rate equal to the sum of the peak rates or sum of the average rate of the individual calls. This method is simple but quite inaccurate when the multiplexed traffic consists of bursty traffic sources.

Another approximation method, known as bufferless link overflow model, has been suggested. In the bufferless model, overflow occurs whenever aggregate traffic rate is larger than the link capacity. In [7], it is assumed that aggregate traffic has a gaussian distribution and the mean and variance are the sums of the average rate and bit rate variance of individual traffic sources respectively. Also in [8], overflow probability is determined by assuming an on-off traffic source and finite call classes. Since the overflow model neglects the buffering effect, overflow probability generally overestimates the actual cell loss probability in the queueing model.

There are schemes called a class related bandwidth assignment rule. In this scheme [9], [10], effective bandwidth estimation is obtained using simulation results or analysis, assuming an on-off arrival process and finite call classes. And such a precomputed bandwidth for homogeneous traffic is stored in a table and used as the guidelines to decide whether to accept a call or not. Also, bandwidth estimation for heterogeneous traffic is given by the linear sum of the required bandwidth of an individual call class or by the bandwidth obtained assuming that all calls are generated by traffic sources with the largest burstiness. The main disadvantage of this scheme is that the required table size and analysis or simulation to create the performance table are unrealistically large. Also, it is not easy to classify the wide range of ATM calls into finite call classes.

In [9], equivalent capacity (required bandwidth guaranteeing the required cell loss probability) for a single exponential on-off type traffic source is obtained using the fluid flow approximation technique. And the equivalent capacity for multiplexed traffic is determined by the minimum bandwidth between the sum of the individual equivalent capacity and the bandwidth obtained using gaussian approximation method. These two methods complement each other over different ranges of connection characteristics. For example, the linear sum of individual capacities, as in the STM network, substantially overestimates the required capacity unless the equivalent capacities of individual connections are close to their mean rates, i.e., low burstiness or number of connection is small. Meanwhile, the gaussian approximation method performs well as the number of connection is increased.

Most of the suggested admission control schemes have the disadvantage of the QOS or required bandwidth estimation being based on a particular arrival process (e.g. interrupted poisson process or on-off process). It may be very difficult or inaccurate to fit the new call with a wide range of traffic characteristics to a particular arrival model.

In this paper, a computationally simple approximation method of estimating cell delay and cell loss probability for admission control of both delay sensitive and loss sensitive calls is proposed. Also, since this method does not assume any particular arrival process, it can be applied for admission control of various types of calls.

The rest of the paper is organized as follows. Mixed input queueing models suitable for admission control and traffic characterization of new and existing calls are described in section II. In section III, we propose an approximation method to estimate the mean cell delay and cell loss probability. Section IV describes the implementation of admission control schemes using this approximation method. Section V investigates the accuracy of the approximation method by comparing with some exact analysis or simulation results and investigates the bursty call effect on the performance of existing calls. Section VI presents the summary and conclusion.

II. Queueing Model and Characterization of Traffic

We concentrate on the output queueing ATM switch and consider the case where a number of calls are statistically multiplexed onto the output ports (see Fig. 1). A mixed input queueing model, where a new call joins the calls which already exist on the output transmission link, seems appropriate for describing the queueing performance effect of adding a new call to the existing calls and deciding the call admission.

Existing traffic will consist of many calls with different bandwidth and statistical nature (CBR, VBR, on-off, etc). In general, it is very difficult to charac-

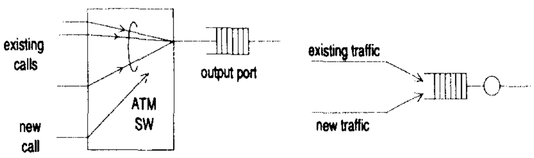


Fig. 1 Mixed input queueing model for admission control

terize the existing traffic which is the superposition of many heterogeneous traffics. Also, such a superposition (especially the superposition of a bursty source) induces the significant correlation among successive interarrival times. Therefore, to represent the various statistical natures of existing traffic, it seems reasonable to use a general point process instead of assuming a particular arrival process such as markov modulated poisson process (MMPP) [10], [11], poisson cluster process (PCP) [12].

In an ATM network, it is expected that many calls are bursty and has some correlation between successive interarrival times. Typically, a bursty source has been modeled by an on-off process or interrupted poisson process (IPP). This simple model captures the basic idea that a bursty source may be either active or inactive and can be fully characterized with a few parameters. However, IPP or on-off process are renewal processes, so there is no correlation between successive interarrival times. Also, the exponential period of on-off time may be severe restrictive to use as a model of bursty traffic source with a wide range of traffic characteristics. Due to such limitations in using the above bursty traffic models, we use a cluster point process as a bursty traffic model [15]. When a new call is bursty, our queueing model can be represented as point process + cluster point process/D/1 queue. This model, though general, seems impossible to be solved analytically. To use an approximation method it is required to characterize both the point process (existing calls) and cluster point process (new call) by a few well-defined parameters that significantly affect the QOS performance.

The rate and variability of offered traffic is commonly recognized as the most important parameters that determine the queueing performance. The precise relationship between variability and performance is very difficult to represent analytically. But generally, more variability corresponds to longer delay and larger packet loss.

The variability of an interarrival time can be partially characterized by its variance or coefficient of variation. The mean arrival rate and coefficient of variation of interarrival time Xn are defined by

$$\lambda \equiv E[Xn]^{-1} \quad (1)$$

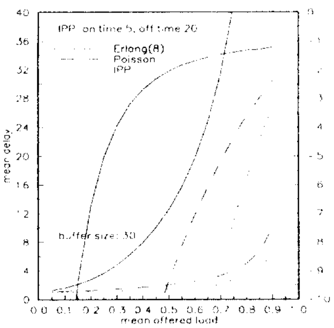
$$C_S^2 \equiv \frac{var(X_n)}{E[X_n]^2} \quad (2)$$

Another important characteristic that significantly affects the QOS performance is the correlation between consecutive interarrival times, which so far has received little attention. Also, such a correlation can be partially characterized by the coefficient of variation of the cumulative interarrival time defined by

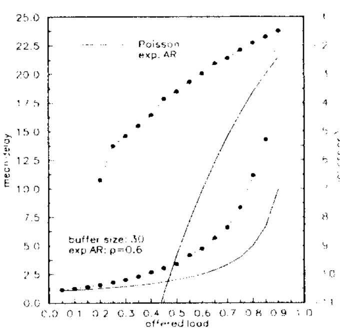
$$C_A^2 \equiv \lim_{k\to\infty} \frac{CV\left(\sum_{i=1}^{k} X_i\right)}{k} = CV(X_n)\left[1+2\lim_{k\to\infty}\sum_{i=1}^{k-1}\left(1-\frac{i}{k}\right)r_i\right]. \quad (3)$$

where CV(· x) represents the coefficient of variation and r_i is the correlation coefficient of interval with lag i. For a renewal point process C_S^2 is equal to C_A^2 but for a correlated point process C_A^2 is larger than C_S^2. We choose the above 3 parameters (λ, C_S^2, C_A^2) to represent both the existing traffic and a new call and QOS is estimated based on these parameters.

In order to investigate the effect of parameters C_S^2 and C_A^2 on the queueing performance, cell loss probability and mean cell delay for some arrival processes are compared in Fig. 2. Fig. 2-(a) represents the delay and loss characteristics of 3 different arrival processes - Poisson, IPP and Erlang arrival process respectively, when the mean service time of exponential server is normalized by 1. Since all these arrival processes are renewal processes, there is no correlation between interarrival times but the coefficient of variation of interarrival time is different. We can see that the mean delay and loss probability are very sensitive to the coefficient of variation of interarrival time C_S^2. Fig. 2-(b) shows the effect of the coefficient of variation of cumulative interval C_A^2. In Fig. 2-(b), the delay and loss probabilities of first order exponential autoregressive process are compared with the Poisson process using simulation. First order exponential AR process $\{Xn\}$ is defined by



(a)



(b)

Fig. 2 Delay and loss characteristics for various arrival process to investigate the effect of parameter C_S^2 and C_A^2. (a) Poisson, IPP and 8 stage Erlang arrival process (b) Poisson and exponential AR with correlation coefficient 0.6^k

$$X_n = \begin{cases} A_0 & n=0 \\ \alpha X_{n-1} + U_n A_n & n \geq 1 \end{cases}, \tag{4}$$

where An is *i.i.d.* exponential distribution with rate λ and U_n is also *i.i.d.* Bernoulli process with probability (1-α). Cell interarrival time distribution of exponential autoregressive process is exponential with rate λ same as the Poisson arrival process but it has correlation between intervals. Correlation coefficient of X_n and X_{n+k} is given by

$$c_k = \frac{\alpha^k}{\lambda^2}. \tag{5}$$

For the exponential AR process, though the variation of interarrival time itself is low, performance severely degrades due to the correlation between intervals. From Fig. 2, we can see that 3 parameters (λ, C_S^2, C_A^2) play critical roles in determining the queueing performance.

III. Computation Procedure

In this section, we present the approximation method of computing the mean cell delay and cell loss probability of point process + (cluster) point process/D/1 queue described in II. The approximation procedures are as follows. (1) characterize the aggregate traffic by its traffic descriptor from the traffic descriptors of existing traffic and a new call, (2) approximate the superposed traffic by a renewal process (3) determine the delay and loss probability of GI/D/1/K queue using appropriate approximation formula. The approximation method used in this paper is based on the renewal approximation method developed by [16], [17], [18]. and the maximum entropy analysis of the GI/G/1 queue [19], [20]. The renewal approximation approach was previously applied by [15], [21] to analyze the voice packet multiplexer and proved that this method is quite accurate.

1. Determining the traffic descriptors of superposed traffic

Let $N_E(t)$ and $N_B(t)$ be the point process indicating the existing traffic and a new call. And their traffic is characterized by traffic descriptors (λ_E, C_{SE}^2, C_{AE}^2) and (λ_B, C_{SB}^2, C_{AB}^2) respectively. We want to characterize the aggregate traffic $N_A(t)$ by its traffic descriptors (λ_A, C_{SA}^2, C_{AA}^2) using the traffic descriptors of $N_E(t)$ and $N_B(t)$. Obviously, arrival rate λ_A of aggregate traffic satisfy

$$\lambda_A = \lambda_E + \lambda_B . \tag{6}$$

C_{SA}^2 can be determined as follows. Interarrival time distribution of superposed traffic $F_A(t)$ is related with the distribution of component process $F_E(t)$ and $F_B(t)$ as [18]

$$\begin{aligned} F_A(t) &= 1 - \frac{\lambda_E \lambda_B}{\lambda_E + \lambda_B} \left\{ \overline{F_E(t)} \int_t^\infty \overline{F_B(u)}\, du \right. \\ &\quad \left. + \overline{F_B(t)} \int_t^\infty \overline{F_E(u)}\, du \right\} \end{aligned} \tag{7}$$

Using this relation, we can compute the C_{SA}^2. However, we know only the first and second moment of interarrival time of component processes (λ_E, C_{SE}^2) and (λ_B, C_{SB}^2). Whitt suggests a method called stationary interval method [16] which calculates the higher moment by fitting a hyperexponential or hypoexponential distribution to their first and second moments. When the coefficient of variation is greater than 1 (higher variability), we let the interval distribution be hyperexponential. The pdf, mean and coefficient of variation are given by

$$\begin{aligned} f(x) &= p_1 \alpha_1 e^{-\alpha_1 t} + p_2 \alpha_2 e^{-\alpha_2 t}, \\ \mu &= p_1 \alpha_1^{-1} + p_2 \alpha_2^{-1}, \\ C^2 &= 2\mu^{-1}(p_1 \alpha_1^{-2} + p_2 \alpha_2^{-2}) \geq 1. \end{aligned} \tag{8}$$

Under the assumption of balanced means ($p_1/\alpha_1 = p_2/\alpha_2$), p_1, p_2, α_1 and α_2 can be determined from given μ and C^2.

$$p_{1,2} = \frac{1}{2}\left(1 \pm \sqrt{\frac{C^2-1}{C^2+1}}\right), \qquad \alpha_{1,2} = 2p_{1,2}\,\mu^{-1}. \tag{9}$$

If the coefficient of variation is less than 1 (low variability), we can use hypoexponential distribution. It has density

$$f(x) = \beta e^{-\beta(x-d)} \quad x \geq d. \tag{10}$$

The two parameters β and d can be determined from given μ and C^2

$$\beta = \lambda C^{-1}, \qquad d = \lambda^{-1}(1-C). \tag{11}$$

Using these distribution fitting, we can compute the C_{SA}^2 from eq.(7). When both C_{SB}^2 and C_{SE}^2 are greater than 1, C_{SA}^2 is given by

$$C_{SA}^2 = 2\lambda_E \lambda_B (\lambda_E + \lambda_B) \left\{ \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{p_i q_j}{\alpha_i \beta_j (\alpha_i + \beta_j)} \right\} - 1. \tag{12}$$

By substituting the traffic parameters into eq (12), C_{SA}^2 can be represented as

$$C_{SA}^2 = \frac{1 - (V_B r_B^2 - V_E r_E^2)^2}{1 + (V_B r_B^2 - V_E r_E^2)^2 - 2(V_B r_B^2 + V_E r_E^2)}, \tag{13}$$

where V_B, V_E, r_B and r_E are defined by

$$V_B = \frac{C_{SB}^2 - 1}{C_{SB}^2 + 1}, \quad V_E = \frac{C_{SE}^2 - 1}{C_{SE}^2 + 1}, \quad r_B = \frac{\lambda_B}{\lambda_A}, \quad r_E = \frac{\lambda_E}{\lambda_A}. \tag{14}$$

When $C_{SE}^2 \leq 1$ and $C_{SB}^2 \geq 1$, we use hyperexponential distribution for a new call and hypoexponential distribution for the existing traffic. After computation, we obtain

$$C_{SA}^2 = 2\lambda_E\lambda_B(\lambda_E+\lambda_B)\sum_{i=1}^{2}\left\{\frac{p_i e^{-\alpha_i d}}{\alpha_i\beta(\alpha_i+\beta)} + \frac{p_i}{\alpha_i^2}(1-e^{-\alpha_i d})(d+\beta^{-1}) - \frac{p_i}{\alpha_i^3}(1-e^{-\alpha_i d}) - \alpha_i d e^{-\alpha_i d})\right\} - 1. \quad (15)$$

By substituting the traffic parameters into eq (15) and using the fact that $\alpha_i d << 1$ in the considering range, $C_{SA}{}^2$ can be represented as

$$C_{SA}^2 = \frac{2(C_{SB}^2+1)C_{SE}(r_B C_{SE}+r_E)}{2r_B^2 C_{SE}^2 + r_E(C_{SB}^2+1)(2r_B C_{SE}+r_E)} - 1. \quad (16)$$

Also, using the same procedure, we can determine the $C_{SA}{}^2$ when ($C_{SE}{}^2 \geq 1$, $C_{SB}{}^2 \leq 1$) and ($C_{SE}{}^2 \leq 1$, $C_{SB}{}^2 \geq 1$).

$C_{AA}{}^2$ can be determined as follows. For a renewal counting process, index of dispersion for counts and index of dispersion for interval are asymptotically equal [22], that is

$$\lim_{k\to\infty}\frac{var(\sum_{i=1}^{k}X_i)}{kE[X_n]^2} = \lim_{t\to\infty}\frac{var(N(t))}{tE[N(t)]}. \quad (17)$$

By applying this asymptotic property for the superposed point process $N_A(t)$ (see also asymptotic method in [16]), we have

$$C_{AA}^2 = \frac{1}{\lambda_A}\lim_{t\to\infty}\frac{var(N_A(t))}{t}. \quad (18)$$

Since $N_A(t)$ is the sum of two independent point processes $N_E(t)$, $N_B(t)$ and also $C_{AE}{}^2$ and $C_{AB}{}^2$ each satisfy the relation of eq(18), we have

$$C_{AA}^2 = r_E C_{AE}^2 + r_B C_{AB}^2. \quad (19)$$

Hence from eq(6), (13) or (16), and (19), we can compute the traffic descriptor of superposed process of existing traffic and new call.

The superposed point process $N_A(t)$ to the queue can be replaced to the GI/G/1 queue by the approximating renewal process. And hence we can use the approximation formula for GI/G/1 queue. We use hybrid renewal approximation method proposed by [17]. This method equates the coefficient of variation of renewal interval $C_H{}^2$ to the convex combination of $C_{SA}{}^2$ and $C_{AA}{}^2$. That is

$$C_H^2 = (1-w)\,C_{SA}^2 + wC_{AA}^2, \quad (20)$$

where weighting function w is given by

$$w = [1+6(1-\rho)^{2.2}n^*]^{-1}, \quad (21)$$

and n^* is given by

$$n^* = \frac{\lambda_E^2+\lambda_B^2}{\lambda_A^2}. \quad (22)$$

2. Mean delay and loss probability computation

The mean number of cells in the infinite buffer GI/G/1 system can be computed using the approximation formula [17]:

$$\overline{N} = \rho + \frac{\rho^2}{2(1-\rho)}(C_H^2+C_B^2)g(C_H^2, C_B^2, \rho). \quad (23)$$

In here, $C_H{}^2$ and $C_B{}^2$ are the coefficient of variation of arrival process and service process respectively. r is the traffic intensity to the queue and $g(C_H{}^2, C_B{}^2, r)$ is given by

$$g(C_H^2, C_B^2, \rho) = \begin{cases} \exp\left[-\frac{2(1-\rho)(1-C_H^2)^2}{3\rho\ \ C_H^2+C_B^2}\right] & C_H^2 \leq 1 \\ \exp\left[-(1-\rho)\frac{C_H^2-1}{C_H^2+4C_B^2}\right] & C_H^2 \geq 1 \end{cases}. \quad (24)$$

This approximation formula is basically an interpolation between known results for special systems and several results under heavy traffic conditions and has been validated by extensive simulations for a wide range of arrival and service process combinations.

The occupancy distribution, $pr\{N=n\}$, of infinite capacity GI/G/1 queue can be determined by choosing a distribution with maximum entropy, satisfying mean queue size constraint determined in eq. (23) and other given information. i.e., maximize the system's entropy function [21]

$$H(p) = -\sum_{n} p(n) \ln p(n) \tag{25}$$

subject to the constraints

$$\sum_{n} p(n) = 1, \quad p(0) = 1-\rho, \quad \sum_{n} np(n) = \overline{N}. \tag{26}$$

The maximum entropy solution can be computed using the Lagrange method of undetermined multipliers and this is given by

$$p(n) = \begin{cases} 1-\rho, & n=0 \\ (1-\rho)gx^n, & n \geq 1 \end{cases}, \tag{27}$$

where g and x can be determined from the constraints in eq. (26)

$$x = \frac{\overline{N}-\rho}{\overline{N}} \tag{28}$$

$$g = \frac{\rho^2}{(\overline{N}-\rho)(1-\rho)} \tag{29}$$

The maximum entropy solution is identical to the queue size distribution for an M/M/1 queue and it is known that for M/G/1 queue, maximum entropy solution becomes exact when the underlying pdf of the service time is characterized by the generalized exponential model. Also for G/M/1 queue, maximum entropy solution is identical to the exact solution based on an embedded MC at the arrival instant, with x equal to the probability of an arrival finding a busy server [13].

Given the distribution $pr\{N=i\}$ of GI/G/1 queue, the queue size distribution, $Pr\{N_K=n\}$, of the corresponding finite buffer system GI/G/1/K is approximated by [23]

$$P_r\{N_K = n\} = \frac{\Pr\{N=n\}}{\Pr\{N \leq K\}} \tag{30}$$

From eq (30), we can determine the mean cell delay and cell loss probability of GI/G/1/K queue. The cell loss probability is given by

$$\hat{P}_L = \Pr\{N_K = K\} = \frac{\rho^2}{\overline{N}-\rho} \frac{\left(\frac{\overline{N}-\rho}{\overline{N}}\right)^K}{1-\rho\left\{\frac{\overline{N}-\rho}{\overline{N}}\right\}^K}. \tag{31}$$

The mean cell delay can be readily obtained by the use of Little's formula; that is

$$\hat{D} = \frac{\sum_{n=0}^{K} n \Pr\{N_K = n\}}{\lambda_A(1-\hat{P}_L)} \tag{32}$$

where $\lambda_A(1-P_L)$ is the mean arrival rate of cells actually entering the system. After some computation, mean cell delay is given by

$$\hat{D} = \frac{1}{\lambda_A} \frac{\overline{N}\left[1-\left(\frac{\overline{N}-\rho}{\overline{N}}\right)^K\left(1+\frac{\rho K}{\overline{N}}\right)\right]}{1-\rho\left(\frac{\overline{N}-\rho}{\overline{N}}\right)^{K-1}}. \tag{33}$$

IV. Call Admission Control Scheme

This section describes a call admission control scheme based on the approximation method developed in III. The call admission control function maintains the link status vector $(\lambda_E, C_{SE}^2, C_{AE}^2)$ that characterizes the currently connected calls to this link. When a new call with traffic parameter $(\lambda_B, C_{SB}^2, C_{AB}^2)$ requests connection, the call admission control calculates the traffic parameters $(\lambda_A, C_{SA}^2, C_{AA}^2)$ that represent the link status after the requested call is added to the existing calls according to eq. (6), (13) or (16) and (19). And then call admission control evaluates the mean delay $\hat{D}$ and cell loss probability $\hat{P}_L$ using eq.(20), (31) and compares with the delay and loss requirement Q_D and Q_L. A new call is accepted if and only if

$$\hat{D} < Q_D \quad \text{and} \quad \hat{P}_L < Q_L \tag{34}$$

If admission control accepts a new call, the link status vector (λ_E, ${C_{SE}}^2$, ${C_{AE}}^2$) is updated by (λ_A, ${C_{SA}}^2$, ${C_{AA}}^2$). When a new call with parameter (λ_B, ${C_{SB}}^2$, ${C_{AB}}^2$) is released, the admission control updates the current link status vector (λ_E, ${C_{SE}}^2$, ${C_{AE}}^2$) according to the following equations.

$$\lambda_E \leftarrow \lambda_E - \lambda_B \tag{35}$$

When both ${C_{SB}}^2$ and ${C_{SE}}^2$ are greater than 1,

$$C_{SE}^2 \leftarrow \frac{(r_E^2 + V_B r_B^2) + \frac{C_{SE}^2}{C_{SE}^2+1} - \sqrt{r_E^2\{1 - r_B C_{SE}^2 + 1\}^2 - (V_B r_B^2 - 1)\left\{V_B r_B^2 - \frac{C_{SE}^2 - 1}{C_{SE}^2+1}\right\}}}{(r_E^2 - V_B r_B^2) - \frac{C_{SE}^2}{C_{SE}^2+1} + \sqrt{r_E^2\{1 - r_B C_{SE}^2 + 1\}^2 - (V_B r_B^2 - 1)\left\{V_B r_B^2 - \frac{C_{SE}^2 - 1}{C_{SE}^2+1}\right\}}} \tag{36}$$

When ${C_{SE}}^2 \leq 1$ and ${C_{SB}}^2 \geq 1$,

$$C_{SE}^2 \leftarrow \frac{r_E\{(1 - r_B(C_{SE}^2+1))\} - \sqrt{r_E^2\{1 - r_B(C_{SE}^2+1)\}^2 - 2 r_B r_E^2 (C_{SE}^2+1)\left\{\frac{C_{SE}^2+1}{C_{SB}^2+1} r_B - 1\right\}}}{2 r_B\left\{r_B \frac{C_{SE}^2+1}{C_{SB}^2+1} - 1\right\}} \tag{37}$$

${C_{AE}}^2$ is updated by

$$C_{AE}^2 \leftarrow \frac{(C_{AE}^2 - r_B C_{AB}^2)}{r_E} \tag{38}$$

Since the computationally simple approximation formula is used for QOS estimation, the admission control function can be carried out in real time. In addition, it can be applied to the admission decision for both delay sensitive and loss sensitive calls. Also, it is independent of the actual cell arrival process.

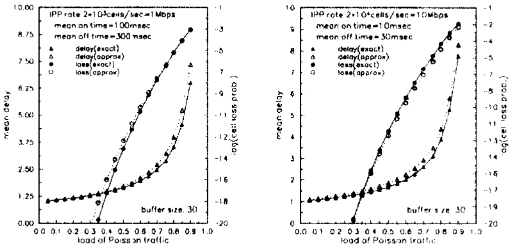
V. Numerical Results

In this section, the accuracy of the suggested approximation method is examined by comparing to the exact analysis or simulation results of various mixed input queueing models. We choose 4 mixed input queueing models - IPP+Poisson/D/1/K, IPP+MMPP/D/1/K, H_2+H_2/D/1/Kand ON−OFF+EXP. AR/D/1/K queue. IPP, H_2 (2 stage hyper exponential) and on-off processes can be interpreted as representations of new call with bursty traffic model and Poisson, MMPP, H_2 and EXP. AR as possible models of existing traffic. Mean delay and cell loss probability using approximation method described in III are compared to those obtained by exact analysis using Neut's matrix geometric solution method [24], [25] and simulation results.

The computation results are summarized in Table 1, 2, 3, 4 and Fig. 3, 4, 5 and 6. Here, the average queueing delay is shown by the normalized value i.e., one unit time corresponds to the cell transmission time.

Fig. 3 and Table 1 show the cell loss probability and average queueing delay of IPP+Poisson/D/1 queue obtained using exact analysis and approximation method as a function of mean offered load of existing traffic for two cases of IPP traffic, where the peak rate of IPP is 10Mbps and 30Mbps. The average on time and off time are 100 and 1000, respectively, for both cases. Fig. 4 and Table 2 show the cell loss probability and average queueing delay of IPP+MMPP/D/1 queue vs mean offered load of MMPP traffic. Two state MMPP model is used and its mean time on state 1 and 2 are 10^3 and 10^4 respectively. The ratio of arrival rate at state 1 and state 2 are

1.2. As we expected, since MMPP is more variable than the Poisson the mean delay and cell loss are larger than those of IPP + Poisson/D/1 case.



(a) (b)

Fig. 3 Delay and Loss Probabilities of IPP+Poisson/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of IPP = 1 Mbps, mean on time = 100 msec and mean off time = 300 msec (b) peak rate of IPP = 10 Mbps, mean on time = 10 msec and mean off time = 30 msec

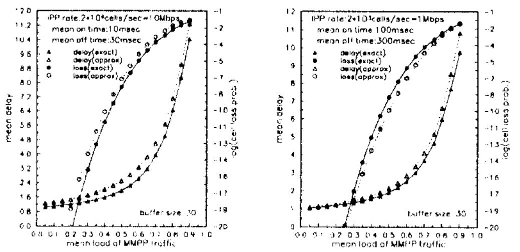
Table 1. Delay and Loss Probabilities of IPP + Poisson/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of IPP = 1 Mbps, mean on time = 100 msec and mean off time = 300 msec (b) peak rate of IPP = 10 Mbps, mean on time = 10 msec and mean off time = 30 msec

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.158	1.629	3.526	4.3e-18	6.5e-11	1.1e-5
approximation	1.174	1.703	3.988	6.6e-17	9.5e-10	1.1e-5

(a)

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.178	1.678	3.901	4.3e-16	1.2e-9	6.9e-5
approximation	1.250	1.844	4.371	2.4e-16	3.9e-10	2.3e-5

(b)



(a) (b)

Fig. 4 Delay and Loss Probabilities of IPP + MMPP/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of IPP = 10Mbps, mean on time = 10 msec and mean off time = 30 msec (b) peak rate of IPP = 1 Mbps mean on time = 100 msec and mean off time = 300 msec

Table 2. Delay and Loss Probabilities of IPP + MMPP/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of IPP = 10Mbps, mean on time = 10 msec and mean off time = 30 msec (b) peak rate of IPP = 1 Mbps mean on time = 100 msec and mean off time = 300 msec

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.206	1.920	6.469	1.4e-11	4.4e-6	2.3e-3
approximation	1.485	2.513	7.143	1.2e-10	1.9e-5	5.0e-2

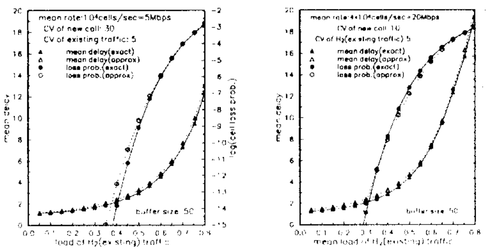
(a)

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.187	1.926	5.987	9.6e-13	1.1e-6	1.4e-3
approximation	1.255	2.130	6.625	5.5e-14	1.4e-7	9.1e-4

(b)

Fig. 5 and Table 3 represent the cell loss probability and average queueing delay of $H_2+H_2/D/1$ queue as a function of mean offered load H_2(existing traffic) for two cases of H2(new call) traffic where the mean rate and coefficient of variation are (10Mbps,

10) and (20Mbps, 20). Fig. 6 and Table 4 show the comparisons with simulation results of ON－OFF＋Exp. AR/D/1 queue when the correlation coefficient with lag k is 0.4^k. As we can see in the above computation results, the suggested approximation method for QOS estimation method has a reasonably good accuracy.



(a) (b)

Fig. 5 Delay and Loss Probabilities of H_2+H_2/D/1/K Queue obtained using approximation and exact analysis (a) mean rate of H_2(new call) = 5 Mbps (b) mean rate of H_2(new call) = 20 Mbps

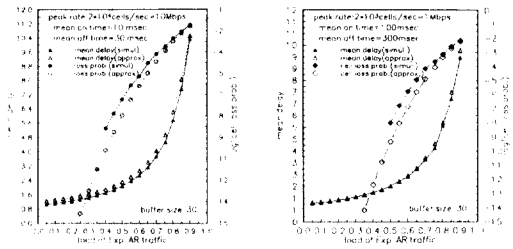
Table 3. Delay and Loss Probabilities of H_2+H_2/D/1/K Queue obtained using approximation and exact analysis. (a) when mean rate of H_2(new call) = 5Mbps (b) when mean rate of H_2(new call) = 10Mbps.

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.340	3.021	12.35	1.5e-14	1.1e-6	1.5e-3
approximation	1.458	3.271	13.01	3.3e-13	1.2e-6	1.7e-2

(a)

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.568	4.396	19.40	2.34e-9	1.34e-4	8.94e-3
approximation	1.815	4.674	20.08	1.42e-6	7.75e-5	9.04e-3

(b)



(a) (b)

Fig. 6 Delay and Loss Probabilities of ON－OFF＋Exp. AR/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of ON－OFF = 10Mbps mean on time = 10 msec and mean off time = 30 msec (b) peak rate of ON－OFF = 1 Mbps mean on time = 100 msec and mean off time = 300 msec.

Table 4. Delay and Loss Probabilities of ON－OFF＋Exp. AR/D/1/K Queue obtained using approximation and exact analysis (a) peak rate of ON－OFF = 10Mbps mean on time = 10 msec and mean off time = 30 msec (b) peak rate of ON－OFF = 1 Mbps mean on time = 100 msec and mean off time = 300 msec.

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.237	2.072	5.760	1.5e-9	4.0e-6	1.3e-3
approximation	1.437	2.304	6.236	5.9e-11	9.1e-7	8.2e-4

(a)

load of poisson traffic	mean delay			loss prob.		
	0.2	0.5	0.8	0.4	0.6	0.8
exact analysis	1.187	1.906	5.987	1.0e-8	7.1e-5	7.1e-4
approximation	1.255	2.130	6.625	4.8e-10	1.3e-5	4.6e-4

(b)

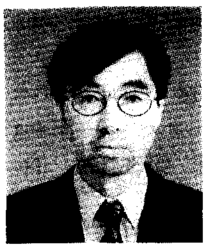
V. Conclusions

In this paper, we have investigated a call admission control for delay or loss sensitive calls. The traffic parameters specified by a new call are assumed to be average arrival rate, coefficient of variance of cell interarrival time and coefficient of variation of cumulative interarrival time. The existing calls are also characterized by these traffic parameters. Based on these parameters, mean cell delay and cell loss probability of superposed traffic are estimated without any assumption on the arrival process. The accuracy of approximation method for estimating cell loss probability and cell delay was examined by comparisons with exact analysis and simulation results. Implementation of admission control schemes was discussed with regard to their real time evaluation of cell loss and delay and traffic parameter update when a call is accepted or released.

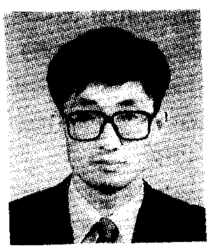
References

1. S.E. Minzer, "Broadband ISDN and asynchronous transfer mode(ATM)," IEEE Comm. Mag., pp. 17-23, Sep. 1989.
2. M. R. Asron and M. Decina, "Asynchronous transfer mode or synchronous mode," IEEE Comm. Mag., pp. 10-13, Jan. 1991.
3. D. Hong and T. Suda, "Congestion control and prevention in ATM networks," IEEE Network Mag., Sep. pp. 10-16, 1991.
4. A. Eckberg, B. T. Doshi and R. Zoccolillo, "Controlling congestion in BISDN/ATM : Issues and stratigies," IEEE Comm. Mag., Sep. 1991.
5. H. Saito, K. Kawashina and K. Sato, "Traffic control technologies in ATM networks," IEICE Trans., vol. E 74, pp. 761-771, no. 4, Apr. 1991.
6. H. Arai, M. Kawarasaki and S. Nogami, "An analysis of call admission control in the ATM network," IEICE Trans., vol. J 72-B-1, no. 11, pp. 1000-1007, 1989.
7. W. Verbiest, L. Pinnoo and B. Voeten, "Statistical multiplexing for variable bit rate video sources in asynchronous transfer mode networks," IEEE Globecom, pp. 7.2.1-7.2.6, 1988.
8. T. Murase, H. Suzuki and S. Sato and T. Takeuchi, "A call admission control scheme for ATM networks using a simple quality estimate," IEEE JSAC, vol. 9 no. 9, Dec. 1991.
9. G. Gallassi, G. Rigolio and L. Fratta, "ATM : Bandwidth assignment and bandwidth enforcement policies," IEEE Globecom, 49.6.1-49.6.6, 1989.
10. M. Decina and T. Toniatti, "On bandwidth allocation to bursty virtual connections in ATM networks," IEEE ICC, pp. 844-851, 1990.
11. R. Guerin, H. Ahamadi and M. Naghshieneh, "Equivalent capacity and its application to bandwidth allocation in high speed networks," IEEE JSAC, vol. 9, no. 7, Sep. 1991.
12. H. Heffs and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," IEEE JSAC, vol. 4, pp. 856-868, 1986.
13. H. Saito, M. Kawasaraki and H. Yamada, "An analysis of statistical multiplexing in an ATM transport networks," IEEE JSAC, vol. 9, no. 3, Apr. 1991.
14. K. Soharby, "Delay analysis of a single server queue with Poisson cluster arrival process arising in ATM networks," IEEE Globecom, pp. 17.5-17.5.6, 1989.
15. W. Whitt, "Approximating a point process by a renewal process, I : Two basic methods," OR, vol. 30, pp. 125-147, 1982.
16. S.L.Albin, "Approximating a point process by a renewal process, II : superposition arrival processes to queues," OR, vol. 32, no.5 Sep.-Oct. 1984.
17. P. J. Kuehn, "Approximation analysis of general queueing networks by decomposition," IEEE Trans. Comm., vol. 27, pp. 113-126, 1979.
18. Y. Jenq. "Approximations for packetized voice traffic in statistical multiplexer," IEEE INFOCOM, pp. 256-259, Apr. 1984.
19. K. Sriram adn W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," IEEE JSAC, vol. 4, no. 6, pp. 833-846, 1986.

20. P. A. W. Lewis, "A branching Poisson process model for the analysis of computer failure patterns," Jour. Royal Statis. Soc. B., vol. 26, 398-456, 1964.
21. D. Kouvatsos, "A maximum entropy analysis of the G/G/1 queue at equilibrium," J. Opl Res. Soc., vol. 39, no. 2, pp. 183-200, 1988.
22. M. A. El-Affendi and D. D. Kouvatsos, "A maximum entropy analysis of the M/G/1 and G/M/1 queueing system at equilibrium," Acta Informatica, vol. 19, pp. 339-355, 1983.
23. M. F. Neuts, "Matrix Geometric solutions in stochastic model," John Hopkins Univ. Press.
24. M. F. Neuts, "Structured Stochastic Matrices of M/G/1 type and their applications," Marcel Dekker, INC., 1989.



안 동 명(Dongmyung Ahn) 정회원
1977년 2월 : 서강대학교 전자공학과(학사)
1979년 2월 : 서울대학교 전자공학과(석사)
1993년 5월 : Brooklyn Polytech 전자공학과(박사)
1979년 3월 ~ 현재 : 국방과학연구소 근무
※주관심분야 : Wireless ATM, 위성통신



한 덕 찬(Duckchan Han) 정회원
1989년 2월 : 충남대학교 전자공학과(학사)
1991년 2월 : 충남대학교 전자공학과(석사)
1991년 2월 ~ 현재 : 국방과학연구소 근무
※주관심분야 : Wireless ATM, 위성통신