Design of a Real-Time Call Admission Controller for ATM Networks

Tsern-Huei Lee, *Member, IEEE*, Kuen-Chu Lai, and Shii-Tyng Duann

Abstract— In this paper, we present a real-time computation algorithm based on the bufferless fluid flow model [6] for call admission control (CAC) on one link of an asynchronous transfer mode (ATM) network with heterogeneous bursty traffic. Cell loss probability is adopted as the measure of quality-of-service (QoS). Our computation algorithm requires a constant memory size and needs only two multiplications and one division to determine whether a connection request can be accepted or not. It is known [7], that due to the interference between different types of traffic the individual cell loss probability may not meet the requirement even though the global one does. In this paper, we provide a close upper-bound for individual cell loss probability which can easily be obtained with our computation algorithm. Numerical examples using typical traffic parameters are studied to corroborate the upper-bound. We also compare the performance of the investigated CAC scheme with that of the effective bandwidth technique [9].

I. INTRODUCTION

THE asynchronous transfer mode (ATM) has been chosen as a flexible and efficient transmission standard for broadband integrated service digital network (B-ISDN). Since ATM adopts the statistical multiplexing technique, traffic control is necessary to avoid possible congestion at each network node and to achieve the quality-of-service (QoS) requested by each connection. Traffic control methods are often divided into two categories, reactive control and preventive control. Because of real time constraint, preventive control is considered to be more suitable than reactive control for high speed networks. Call admission control (CAC) and bandwidth policing are two effective forms of preventive traffic control.

Several CAC schemes for individual links of ATM networks have been proposed and analyzed recently [1]–[9]. Long-term time-averaged cell loss probability and cell delay are often adopted as measures of QoS. Cell delay can usually be controlled within a desired bound by engineering the buffer size, hence in this paper we choose cell loss probability as the QoS. Control schemes based on exact analysis of cell loss probability can achieve good utilization of link capacity. However, the analysis is often too complicated to implement and, moreover, the result is quite sensitive to burst length, a traffic parameter which is difficult to be declared and policed.

T.-H. Lee and S.-T. Duann are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, ROC (e-mail: thlee@atm.cm.nctu.edu.tw).
K.-C. Lai is with the Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan, ROC.

Control schemes based on simplified models have also been investigated. An example is the effective bandwidth technique [9]. The effective bandwidth (called equivalent capacity in [8]) of a set of connections multiplexed on a transmission link is defined as the minimum amount of bandwidth required to achieve a desired QoS [8]. A computationally simple approximate expression for the effective bandwidth of individual and multiplexed connections was derived in [8]. Based on the approximate expression, CAC can be performed in real time. Unfortunately, the approximate expression loses the benefit of statistical multiplexing gain and thus, becomes inappropriate when statistical multiplexing is significant.

Another example is the bufferless fluid flow model studied in [6] and [7]. However, the control schemes suggested in [6] and [7] require computation of convolution which is difficult to be performed in real time. A virtual bandwidth technique was suggested in [7] to replace the convolution. The virtual bandwidth of each type of traffic is derived based on the homogeneous traffic model and thus, cannot guarantee the QoS under heterogeneous traffic conditions. In fact, the interference between different types of traffic is very difficult to determine [7].

In this paper, we present a fast implementation for the CAC scheme proposed in [6] and [7]. We show that if information about the current traffic on the link is stored in an appropriate manner, then the system can decide whether or not to accept a connection request with only two multiplications and one division.

The rest of this paper is organized as follows. In Section II, we review the bufferless fluid flow model. Our computation algorithm is presented in Section III. In Section IV, we extend the computation algorithm to obtain a close upper-bound of individual cell loss probability. Some numerical examples are studied in Section V. Comparisons of the performance of the investigated CAC scheme with that of the effective bandwidth technique is also presented in this section. Conclusion is finally drawn in Section VI.

II. THE BUFFERLESS FLUID FLOW MODEL

For simplicity, we study on–off sources, i.e., all traffic sources alternate between active and idle periods. On–off models had been adopted to approximate real traffic sources [10] and [11]. The computation algorithm to be presented in Section III can be generalized for sources described by Markov-modulated fluid flow models [9].

Consider an on–off source. In active periods, cells are generated at a peak bit rate denoted by (MAX). In idle periods,
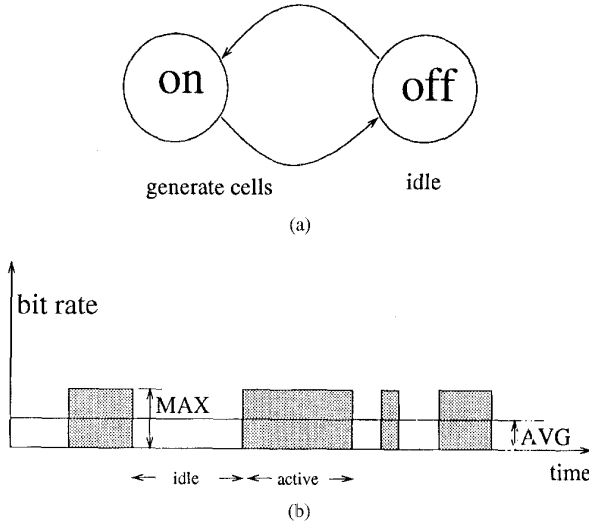
Fig. 1. (a) Source model and (b) traffic model.

no cells are generated. Let (AVG) denote the average bit rate of a traffic source. The probability that a traffic source is active or idle is then given by AVG/MAX or $1 - $ AVG/MAX, respectively. Fig. 1 illustrates the source and traffic models.

Under the bufferless fluid flow model, cell losses due to overflow occur if and only if (iff) the aggregate MAX R exceeds the link capacity C, where R is the sum of the MAX's of all active calls. Assume there are N existing calls. Let MAX_i and AVG_i denote, respectively, the MAX and AVG of call i, $1 \leq i \leq N$. Also, let $f_i(x)$ represent the probability density function (pdf) of the traffic generated by call i, i.e.,

$$f_i(x) = \begin{cases} \dfrac{\text{AVG}_i}{\text{MAX}_i}, & \text{if } x = \text{MAX}_i \\ 1 - \dfrac{\text{AVG}_i}{\text{MAX}_i}, & \text{if } x = 0. \end{cases} \quad (1)$$

Assuming the calls are independent, the density function of the aggregate traffic generated by the N existing calls, denoted by $q(x)$, is equal to the convolution of $f_1, f_2, \cdots$, and f_N, i.e.,

$$q(x) = (f_1 * f_2 * \cdots * f_N)(x). \quad (2)$$

Let ET denote the excess traffic and ρ represent the traffic load. The virtual cell loss probability Pv is defined as [7]

$$Pv = \frac{ET}{\rho} \quad (3)$$

where

$$\begin{aligned} ET &= E[(R - C)^+] \\ &= \sum_x (x - C)^+ q(x) \end{aligned} \quad (4)$$

and

$$\rho = \sum_{i=1}^{N} \text{AVG}_i. \quad (5)$$

Notice that the virtual cell loss probability depends only on MAX_i and AVG_i. Moreover, it is an upper-bound of the actual

cell loss probability [7] and thus, can be used as a criterion for CAC.

Unfortunately, to use (3)–(5) to calculate the virtual cell loss probability, one needs to perform convolution which is not real-time computable if the number of existing calls is large. Suppose the N existing calls can be categorized into L classes so that two calls are in a same class iff they have the same peak and AVG's. Furthermore, assume that class j has N_j calls such that $\sum_{j=1}^{L} N_j = N$. To determine whether or not a connection request can be accepted, one has to determine the pdf of the traffic generated by the existing calls and the call requesting connection. Let $f_{N+1}(x)$ denote the pdf of the traffic generated by the call requesting connection. Since the number of nonzero terms in $q(x)$ could be as large as $\prod_{j=1}^{L} (N_j + 1)$, $2 \prod_{j=1}^{L} (N_j + 1)$ multiplications may be needed to obtain $(q * f_{N+1})(x)$. Moreover, after obtaining $(q * f_{N+1})(x)$, one has to perform $2 \prod_{j=1}^{L} (N_j + 1)$ (if the connection request does not belong to any existing category) or $(N_l + 2)/(N_l + 1) \prod_{j=1}^{L} (N_j + 1)$ (if the connection request belongs to category l, $1 \leq l \leq L$) multiplications to obtain the value of ET in (4). Therefore, the computational complexity involved in determining whether a connection request can be accepted is of order $O(N_1 \times N_2 \times \cdots \times N_L)$. The number of multiplications that need to be performed when a connection is finished is also of order $O(N_1 \times N_2 \times \cdots \times N_L)$. Notice that the complexity is enumerated assuming the density function $q(x)$ is available. Without the density function, the complexity is of order $O(N_1^2 \times N_2^2 \times \cdots \times N_L^2)$ [6]. As an example, for $L = 7$ and $N_1 = N_2 = \cdots = N_7 = 10$, one needs to perform more than ten million multiplications to determine whether a connection request should be accepted or denied. Obviously, this cannot be done in real time. Storing the density function is another difficulty for the convolution method, because the number of nonzero terms in $q(x)$ varies and the locations where $q(x)$ is nonzero are irregular (they depend on the traffic type of exiting calls).

In [7], the virtual bandwidth technique was proposed to replace convolution. The virtual bandwidth of each traffic type is derived based on the homogeneous traffic assumption. It was found [7], that in a heterogeneous environment the virtual bandwidth may increase because of the interference between different types of traffic. Consequently, the individual QoS cannot be guaranteed using the virtual bandwidth technique. The effect of interference is very difficult to determine, especially when there are many different types of traffic sources. In the following section, we present a computation algorithm which requires a constant memory size and can be performed in real time.

III. A REAL-TIME COMPUTATION ALGORITHM

Let C denote the link capacity or transmission speed. Select a data rate u as a unit so that C is an integral multiple of u. Assume the MAX of each type of traffic is also an integral multiple of u. We shall consider later the case where the MAX's of some types of traffic are not integral multiples of u. Without loss of generality, we assume $u = 1$ and thus, the density function $f(x)$ of a traffic source whose MAX equals

j times the unit rate u can be expressed as

$$f(k) = \begin{cases} \dfrac{\text{AVG}}{\text{MAX}}, & \text{if } k = j \\ 1 - \dfrac{\text{AVG}}{\text{MAX}}, & \text{if } k = 0. \end{cases} \qquad (6)$$

Define $I(m)$ as follows:

$$I(m) = E[R - m]^+$$
$$= \sum_{k=m+1}^{\infty} (k - m)q(k). \qquad (7)$$

It is clear that $I(m)$ represents the expected excess traffic when the link capacity is m under the bufferless fluid flow model and $I(0)$ is equal to the traffic load ρ. Moreover, the virtual cell loss probability is given by $I(C)/I(0)$. We show below that real-time call admission control is achievable as long as the values of $I(m)$, $0 \le m \le C$, are stored.

Suppose there are N existing calls and a connection request with peak rate MAX_{N+1} and average rate AVG_{N+1} arrives. Let $\hat{I}(m)$ denote the updated value of $I(m)$ assuming the connection request is accepted. The value of $\hat{I}(m)$ can be obtained as follows:

$$\hat{I}(m) = \sum_k (k - m)^+ q * f_{N+1}(k)$$

$$= \left(1 - \frac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}}\right) \sum_k (k - m)^+ q(k)$$

$$+ \frac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}} \sum_k (k - m)^+ q(k - \text{MAX}_{N+1})$$

$$= \begin{cases} \left(1 - \dfrac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}}\right) I(m) \\ \quad + \left(\dfrac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}}\right)[(\text{MAX}_{N+1} - m) + I(0)], \\ \quad \text{if } m < \text{MAX}_{N+1} \\ \left(1 - \dfrac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}}\right) I(m) \\ \quad + \left(\dfrac{\text{AVG}_{N+1}}{\text{MAX}_{N+1}}\right) I(m - \text{MAX}_{N+1}) \\ \quad \text{if } m \ge \text{MAX}_{N+1}. \end{cases} \qquad (8)$$

Since $I(m) = I(0) - m$ for $m < 0$, one can rewrite $\hat{I}(m)$ as

$$\hat{I}(m) = f_{N+1}(0)I(m)$$
$$+ f_{N+1}(\text{MAX}_{N+1})I(m - \text{MAX}_{N+1}) \qquad (9)$$

where $f_{N+1}(0) = 1 - (\text{AVG}_{N+1}/\text{MAX}_{N+1})$ and $f_{N+1}(\text{MAX}_{N+1}) = \text{AVG}_{N+1}/\text{MAX}_{N+1}$. Notice that the initial value of $I(m)$ when $N = 0$ is zero for all m. For CAC, the connection request is denied if $\hat{I}(C)/\hat{I}(0) > \eta$, where η is a predetermined cell loss probability. In this case, the values of $I(m)$, $0 \le m \le C$, need not be updated. Therefore, to make a decision, one needs to perform only two multiplications to get $\hat{I}(C)$ and one division to get $\hat{I}(C)/\hat{I}(0)$. The value of $\hat{I}(0)$ is equal to $I(0) + \text{AVG}_{N+1}$ and thus, can easily be obtained. Suppose $\hat{I}(C)/\hat{I}(0) \le \eta$ and the connection request is accepted; then the system needs

to perform a total of $2(C + 1)$ multiplications to update $I(m)$, $0 \le m \le C$. When a call finishes, $I(m)$ always needs to be updated. Suppose a call with parameters MAX_i and AVG_i finishes. Then $\hat{I}(m)$, the updated value of $I(m)$, is given by

$$\hat{I}(m) = f_i^{-1}(0)\{I(m) - f_i(\text{MAX}_i)\hat{I}(m - \text{MAX}_i)\} \qquad (10)$$

where $f_i(0) = 1 - \text{AVG}_i/\text{MAX}_i$ and $f_i(\text{MAX}_i) = \text{AVG}_i/\text{MAX}_i$. Notice that (10) is derived from (9) by interchanging the roles of $I(m)$ and $\hat{I}(m)$. Since $\hat{I}(0) = I(0) - \text{AVG}_i$, one can obtain $\hat{I}(m)$, $m = 1, 2, \cdots, C$, recursively.

It is clear, that the computation algorithm requires a constant memory size (i.e., $C+1$ words). Moreover, it can be performed in real time. For example, for $C = 150$ Mb/s and $u = 16$ Kb/s, one needs to store only 9376 values and it takes 18 752 multiplications to update these values when a connection request is accepted or a call is finished. If a multiplication takes 2 μs, then the update can be completed in 0.04 s.

The quantization technique can be adopted when the MAX of some traffic type is not an integral multiple of u. That is, a traffic source with parameters MAX and AVG is regarded as one with parameters $\lceil \text{MAX} \rceil$ and AVG, where $\lceil x \rceil$ is the smallest integer greater than or equal to x. Note, that the quantized source model is burstier than the unquantized one (the burstiness of a traffic source with MAX and AVG is defined as MAX/AVG) and the virtual cell loss probability of the quantized system is an upper-bound of that of the unquantized one. This is proved in the Appendix. Consequently, one can use the quantized model for CAC. The loss in utilization due to quantization will be studied in Section V.

IV. INDIVIDUAL CELL LOSS PROBABILITY

The cell loss probability defined in Section III is for global traffic. In this section, we study the cell loss probability for each individual type of traffic. Tag an existing traffic source of type i. Let $q_0(k)$ denote the pdf of random variable (r.v.) X, the aggregate traffic generated by all the traffic sources except for the tagged one. For notational convenience, let M_i and A_i denote, respectively, the MAX and AVG of the tagged source. If Y denotes the r.v. of the traffic generated by the tagged source, then the virtual cell loss probability Pv_i of the tagged source (or type i sources) is given by

$$Pv_i = \frac{1}{A_i} E\left[(X + Y - C)^+ \frac{Y}{X + Y}\right]$$

$$= \frac{1}{A_i} E\left[E[(X + Y - C)^+ \frac{Y}{X + Y}|Y]\right]$$

$$= \frac{1}{A_i} E\left[(X + M_i - C)^+ \frac{M_i}{X + M_i}\right] \frac{A_i}{M_i}$$

$$= E\left[(X + M_i - C)^+ \frac{1}{X + M_i}\right]$$

$$= \sum_{k \ge C - M_i} (k + M_i - C)q_0(k) \frac{1}{k + M_i}. \qquad (11)$$

Define r.v. Z_i to be $Z_i = (X + M_i - C)^+$. The virtual cell loss probability of the tagged source is then given by

$Pv_i = E[Z_i/(Z_i + C)]$. Since $g(z) = z/(z + C)$ is a concave function in the interval $[0, \infty)$, by Jensen's inequality [13], we get

$$
\begin{aligned}
Pv_i &= E\left[\frac{Z_i}{Z_i + C}\right] \\
&\leq \frac{E[Z_i]}{E[Z_i] + C} \\
&= \tilde{P}v_i.
\end{aligned} \tag{12}
$$

Therefore, one can use $\tilde{P}v_i$ as the virtual cell loss probability for type i traffic sources. Notice that $\tilde{P}v_i$ is a close upper-bound of Pv_i. This can be seen from

$$
\begin{aligned}
|g''(z)| &= \left|-\frac{C}{(z + C)^3}\right| \\
&< \left|\frac{1}{C^2}\right| \\
&\ll 1
\end{aligned} \tag{13}
$$

which means $g(z)$ is very close to a straight line and thus, Jensen's inequality provides a close upper-bound. We shall corroborate this upper-bound with numerical examples in the next section.

It should be pointed out that $Pv_i = E[Z_i/(Z_i + C)]$ may underestimate the cell loss probability of type i traffic if some traffic types are quantized and type i is not. However, Jensen's inequality always gives an upper-bound. To see this, suppose there is no quantization and Pv_i, the cell loss probability of type i traffic, is computed using the exact density functions. By Jensen's inequality, we have

$$
Pv_i \leq \frac{E[Z_i]}{E[Z_i] + C} \tag{14}
$$

where $E[Z_i] = E[X + M_i - C]^+$ is also evaluated assuming there is no quantization. Let $\tilde{Z}_i$ be the r.v. of Z_i with some quantized traffic sources. From the results proved in the Appendix, we know that $E[\tilde{Z}_i] \geq E[Z_i]$. Consequently, we have

$$
\frac{E[Z_i]}{E[Z_i] + C} \leq \frac{E[\tilde{Z}_i]}{E[\tilde{Z}_i] + C} \tag{15}
$$

because $E[Z_i]/(E[Z_i] + C)$ is a proper fraction. Therefore, Jensen's inequality always gives an upper-bound on individual cell loss probability regardless of whether there are quantized sources or not.

To compute $\tilde{P}v_i$, one needs $E(Z_i)$, which can be obtained as follows:

$$
\begin{aligned}
E[Z_i] &= E[(X + M_i - C)^+] \\
&= \sum_{k \geq C - M_i} (k + M_i - C) q_0(k) \\
&= I_i(C - M_i)
\end{aligned} \tag{16}
$$

where $I_i(m)$ is the expected excess traffic generated by the N existing calls except for the tagged one when the link capacity is m. Notice that $I_i(m) = I_i(0) - m$ for $m < 0$. Therefore, as long as the values of $I_i(m)$ $(0 \leq m \leq C)$ are available, one can obtain $\tilde{P}v_i$ for type i traffic sources. In

general, if there are L types of traffic sources, then $(L + 1)$ arrays $I_0(m)$, $I_1(m)$, $\cdots$, and $I_L(m)$ are needed, where $I_0(m)$ is used for the global cell loss probability and $I_i(m)$ is used for the individual cell loss probability of type i traffic sources.

CAC based on individual cell loss probabilities can be implemented as follows. When a connection request with MAX and AVG arrives, the system performs the following three steps. In Step 1), W denotes the r.v. of the traffic generated by the call requesting connection.

Step 1) Compute $\tilde{P}v_i = E[Z_i]/E[Z_i] + C$ for $1 \leq i \leq L$, where

$$
\begin{aligned}
E[Z_i] &= E[(X + M_i + W - C)^+] \\
&= \left(1 - \frac{\text{AVG}}{\text{MAX}}\right) E[(X + M_i - C)^+] \\
&\quad + \frac{\text{AVG}}{\text{MAX}} E[(X + M_i + \text{MAX} - C)^+] \\
&= \left(1 - \frac{\text{AVG}}{\text{MAX}}\right) I_i(C - M_i) \\
&\quad + \frac{\text{AVG}}{\text{MAX}} I_i(C - M_i - \text{MAX}).
\end{aligned} \tag{17}
$$

Step 2) If the connection request creates a new traffic type (type $L + 1$), compute

$$
\tilde{P}v_{L+1} = \frac{E[Z_{L+1}]}{E[Z_{L+1}] + C}
$$

where

$$
\begin{aligned}
E[Z_{L+1}] &= E[(X + \text{MAX} - C)^+] \\
&= I_0(C - \text{MAX}).
\end{aligned} \tag{18}
$$

Step 3) Accept the connection request and update $I_i(m)$ for all i and m according to

$$
\begin{aligned}
\hat{I}_i(m) &= \left(1 - \frac{\text{AVG}}{\text{MAX}}\right) I_i(m) \\
&\quad + \frac{\text{AVG}}{\text{MAX}} I_i(m - \text{MAX})
\end{aligned} \tag{19}
$$

iff $\tilde{P}v_i \leq \eta$ for all i.

When a call of type j finishes, $I_i(m)$ is updated according to $\hat{I}_i(m) = f_j^{-1}(0)\{I_i(m) - f_j(\text{MAX}_j)\hat{I}_i(m - \text{MAX}_j)\}$ for all i and m. If there is no other call of type j in the system, then $I_j(m)$ is cleared (reset to zero) for all m.

Notice that the calculations in updating $I_i(m)$ involve floating point arithmetic, which unavoidably will cause round-off errors. The round-off errors may accumulate to make $I_i(m)$ no longer sufficiently accurate for CAC decisions. To avoid accumulation of round-off errors, one can use a separate processing unit to compute $I_i(m)$ based on current traffic sources and replace the old values by the newly computed ones. To be more precise, suppose at time instant t_1, the separate processing unit starts to compute $I_0(m)$ [using (9)] and completes its job at time instant t_2. For convenience, we denote the values computed by the separate processing unit by $I_0'(m)$, $0 \leq m \leq C$. There may be some connection requests accepted and some calls terminate during time interval $(t_1, t_2]$. The separate processing unit further updates $I_0'(m)$

in accordance with these changes and completes the update process at time instant t_3. More updates are needed if there are changes during time interval $(t_2, t_3]$. At some instant t_i where there is no change during interval $(t_{i-1}, t_i]$, the vector $I_0(m)$ is replaced with $I'_0(m)$. Once the new $I_0(m)$ vector is obtained, one can quickly compute the new $I_i(m)$ $(1 \leq i \leq L)$ vector [using (10)] to replace the old one. As a result, the round-off errors before time t_1 are removed. The separate processing unit can start the procedure all over again as soon as all the stored vectors are renewed. For a link with $C = 150$ Mb/s and $u = 32$ Kb/s, the interval $(t_1, t_2]$ is about 20 s, assuming the number of existing calls is 1000. The number of changes is expected to be not large in a time interval of 20 s. Therefore, the scheme is feasible. If a higher speed for computing $I_0(m)$ is required, then one can use dedicated hardware to implement the update process [12].

V. NUMERICAL EXAMPLES

In the examples studied in this section, we assume link capacity $C = 150$ Mb/s. Two types of traffic sources are considered. Type I sources have $MAX_1 = 64$ Kb/s and $AVG_1 = 32$ Kb/s. The two parameters for type II sources are $MAX_2 = 5$ Mb/s and $AVG_2 = 0.5$ Mb/s. The desired cell loss probability is restricted to be at most 10^{-9} for both types of traffic sources.

In Figs. 2–4, we use global cell loss probability for CAC. Fig. 2 shows the admissible regions for combinations (n_1, n_2), where n_i denotes the number of type i sources, for $u = 8$ Kb/s, 64 Kb/s, and 100 Kb/s. For convenience, the combinations lying on the boundary are called maximum combinations. Notice that there is no quantization loss for $u = 8$ Kb/s because both MAX_1 and MAX_2 are integral multiples of 8 Kb/s. It can be seen from Fig. 2 that the quantization loss is quite small for $u = 64$ Kb/s and 100 Kb/s. In other words, it is possible to reduce the required memory size and updating time by selecting a larger value of u with only a slight sacrifice in utilization. Fig. 3 illustrates the utilized link capacity versus the number of type II sources for maximum combinations. The utilized link capacity for combination (n_1, n_2) is defined as $n_1 \times AVG_1 + n_2 \times AVG_2$. Fig. 4 shows the quantization loss in percentage of utilized link capacity for maximum combinations. The percentage of loss is defined as $(UC_8 - UC_x)/UC_8$, where UC_x is the utilized link capacity for $u = x$ Kb/s. It can be seen that the quantization loss is small except when there is almost no type II traffic.

The oscillation in Fig. 4 is due to the constraint of integer number of sources, i.e., both n_1 and n_2 need to be integers. Let (k_1, k_2), and (k'_1, k_2) be maximum combinations for the unquantized system and the quantized system, respectively. When n_2 is increased from k_2 to $k_2 + 1$, both k_1 and k'_1 are decreased. In general, the decrement of k'_1 is greater than that of k_1, because the more the type II sources, the more the quantization loss. Moreover, the constraint of integer number of sources results in different losses to the unquantized and quantized systems. When more type II sources are added, the losses can accumulate to accommodate a type I source.
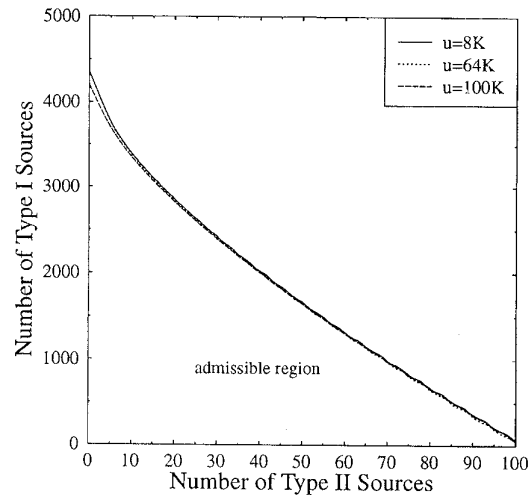

Fig. 2. Admissible region ($MAX_1 = 64$ Kb/s, $AVG_1 = 32$ Kb/s, $MAX_2 = 5$ Mb/s, $AVG_2 = 0.5$ Mb/s).
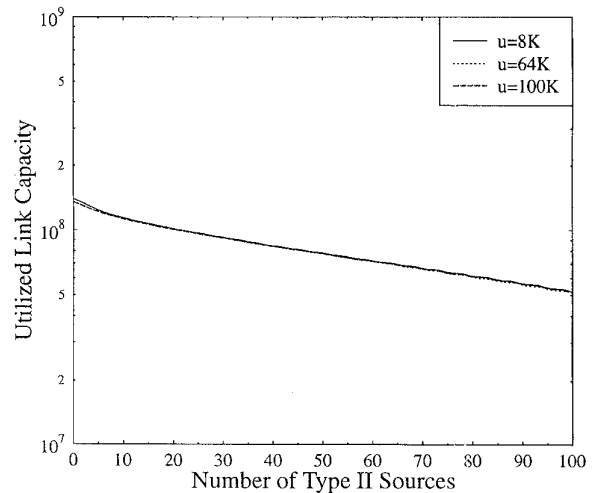

Fig. 3. Utilized link capacity versus number of type II sources for maximum combinations.

In Fig. 5, we examine the upper-bound derived from Jensen's inequality in Section IV. In this figure, the number of type II sources is fixed at 50. One can see from the figure that the upper-bound is close to the actual value. In Figs. 6–8, we consider individual cell loss probabilities for CAC. Fig. 6 shows the admissible regions, which are slightly smaller than those shown in Fig. 2 because of the interference between type I and type II traffic. Fig. 7 shows the utilized link capacity for maximum combinations. The decrease in utilization caused by quantization and the adoption of Jensen's inequality is small for our investigated example. For example, the utilized link capacity is about 76 Mb/s for $u = 8$ Kb/s and about 75.5 Mb/s for $u = 100$ Kb/s (with quantization and Jensen's inequality) when the number of type II sources is equal to 50. The percentage of loss in utilization caused by quantization and adoption of Jensen's inequality is shown in Fig. 8. One can see that the percentage of loss is always smaller than 4%.

Figs. 9–11 show the admissible regions of the investigated CAC scheme and the effective bandwidth technique [9] with
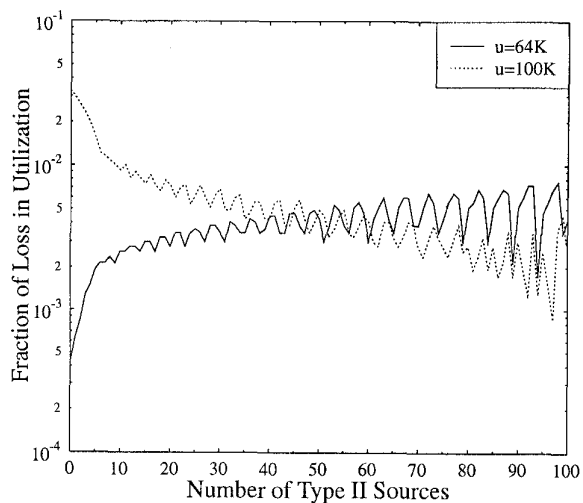
Fig. 4. Fraction of loss in utilized link capacity versus number of type II sources.
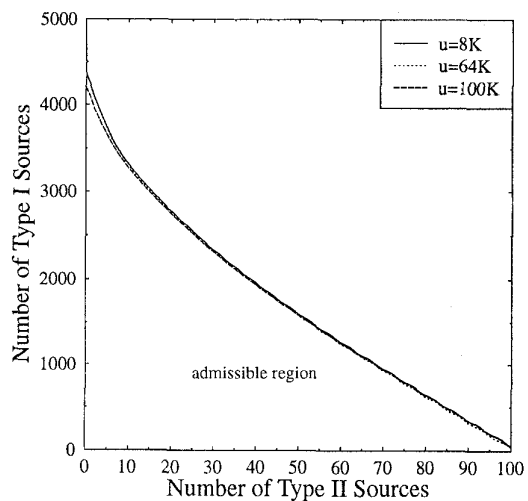

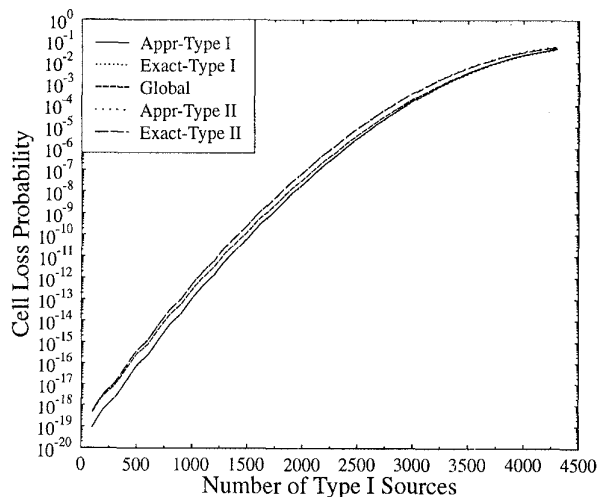Fig. 6. Admissible region based on individual cell loss probability.

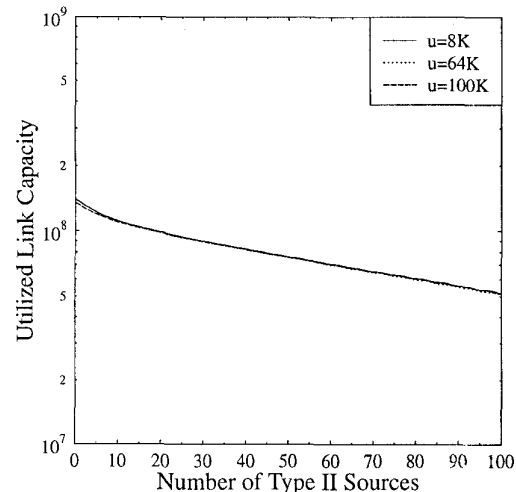Fig. 5. Jensen's inequality with 50 type II sources.

Fig. 7. Utilized link capacity versus number of type II sources for maximum combinations based on individual cell loss probability.

various buffer sizes. For the investigated scheme, buffer size is not a parameter. For the effective bandwidth technique, state transition rates have to be considered. Let r_i^{01} (r_i^{10}) denote the transition rate of type i sources from "off" ("on") state to "on" ("off") state. Notice that the stationary probability of "on" state is given by $r_i^{01}/(r_i^{01} + r_i^{10}) = \text{AVG}_i/\text{MAX}_i$, which implies $r_i^{10} = \text{MAX}_i/\text{AVG}_i \times r_i^{01}(1 - \text{AVG}_i/\text{MAX}_i)$. In Figs. 9–11, we choose $r_1^{01} = r_2^{01} = 0.1$; $r_1^{01} = r_2^{01} = 1$; and $r_1^{01} = r_2^{01} = 10$, respectively. It can be seen in these figures that, for the effective bandwidth technique to outperform the investigated scheme, the required buffer size would become unreasonably large. Since the effective bandwidth technique requires large buffer size, we would expect that it is used for delay in-sensitive traffic and our investigated scheme for delay sensitive traffic.

VI. CONCLUSION

Call admission control for ATM networks is an important issue that has attracted the attention of many researchers

in recent years. Exact analysis based on various queuing models can achieve good utilization of network resources; such analysis often is too complicated to be performed in real time, however, and the results are sensitive to traffic parameters such as burst rate and burst length, which cannot be easily declared and policed. In this paper, we have presented a fast implementation of a CAC scheme which uses only two traffic parameters, the MAX and AVG, of each traffic source. The computation algorithm can be generalized for a system with sources described by Markov-modulated fluid flow models. An interesting further research topic is to extend the results to a system with prioritized traffic sources.

APPENDIX

In this Appendix, we prove that the virtual cell loss probability for the quantized system is greater than or equal to that of the unquantized system. Assume there are N existing calls and let Y_i, $1 \leq i \leq N$, denote the r.v. of the traffic generated
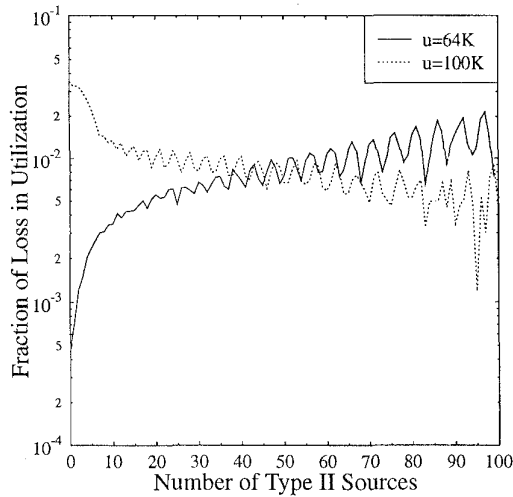
Fig. 8. Fraction of loss in utilized link capacity versus number of type II sources based on individual cell loss probability.
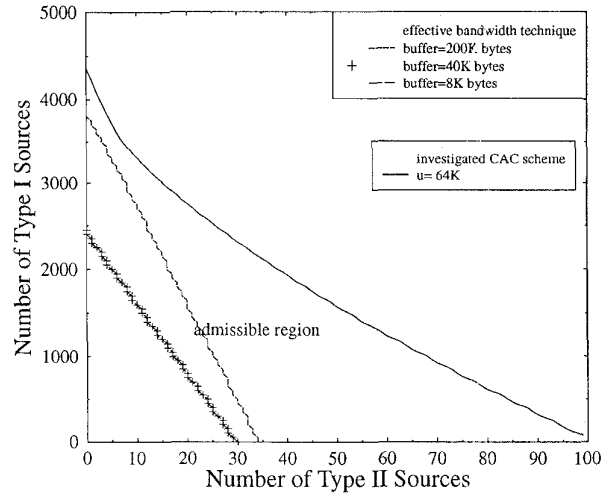

Fig. 10. Admissible regions of the investigated CAC scheme and the effective bandwidth technique with medium transition rates ($r_1^{01} = r_2^{01} = 1$).
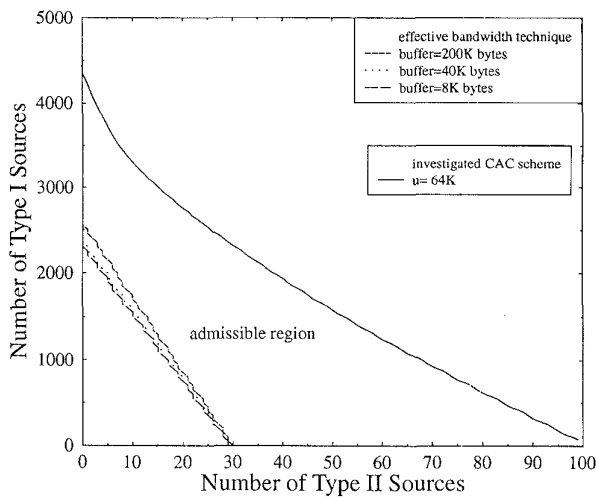

Fig. 9. Admissible regions of the investigated CAC scheme and the effective bandwidth technique with small transition rates ($r_1^{01} = r_2^{01} = 0.1$).
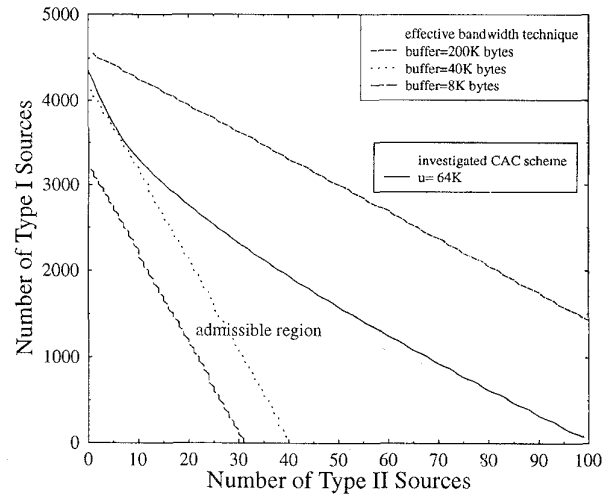

Fig. 11. Admissible regions of the investigated CAC scheme and the effective bandwidth technique with large transition rates ($r_1^{01} = r_2^{01} = 10$).

by the ith source. That is, the density function $f_i(x)$ of Y_i is given by

$$f_i(x) = \begin{cases} \dfrac{\text{AVG}_i}{\text{MAX}_i}, & \text{if } x = \text{MAX}_i \\ 1 - \dfrac{\text{AVG}_i}{\text{MAX}_i}, & \text{if } x = 0. \end{cases} \tag{20}$$

Let $\tilde{Y}_i$ denote the traffic generated by the ith source after quantization. In other words, the density function $g_i(x)$ of $\tilde{Y}_i$ is given by

$$g_i(x) = \begin{cases} \dfrac{\text{AVG}_i}{\lceil \text{MAX}_i \rceil}, & \text{if } x = \lceil \text{MAX}_i \rceil \\ 1 - \dfrac{\text{AVG}_i}{\lceil \text{MAX}_i \rceil}, & \text{if } x = 0 \end{cases} \tag{21}$$

where $\lceil t \rceil$ is the smallest integer greater than or equal to t. We need to prove $E[\tilde{Y} - C]^+ \geq E[Y - C]^+$, where $\tilde{Y} = \sum_{i=1}^N \tilde{Y}_i$

and $Y = \sum_{i=1}^N Y_i$. The proof of $E[\tilde{Y} - C]^+ \geq E[Y - C]^+$ is provided in the corollary of the following theorem. In the following theorem, $u(t)$ is the unit step function, i.e.,

$$u(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0. \end{cases} \tag{22}$$

Theorem 1: Let X be an arbitrary r.v. with distribution function $F(t)$. Assume U and V are two r.v.'s which are independent of X and have, respectively, distribution functions $F_U(t) = (1 - A/M)u(t) + (A/M)u(t - M)$ and $F_V(t) = (1 - A/m)u(t) + (A/m)u(t - m)$, where $A > 0$ and $M \geq m > 0$. It holds that $E[X + U]^+ \geq E[X + V]^+$.

Proof: The theorem can be proved as follows:

$$E[X + U]^+ - E[X + V]^+$$
$$= E[X + M]^+ P(U = M) + E[X]^+ P(U = 0)$$
$$\quad - E[X + m]^+ P(V = m) - E[X]^+ P(V = 0)$$

$$= \frac{A}{M} E[X + M]^+ + \left(1 - \frac{A}{M}\right) E[X]^+$$

$$- \frac{A}{m} E[X + m]^+ - \left(1 - \frac{A}{m}\right) E[X]^+$$

$$= A \left\{ E\left[\frac{X}{M} + 1\right]^+ - E\left[\frac{X}{M}\right]^+ \right.$$

$$\left. - E\left[\frac{X}{m} + 1\right]^+ + E\left[\frac{X}{m}\right]^+ \right\}$$

$$= A \left\{ \int_{x \geq -M} \left(\frac{x}{M} + 1\right) dF(x) - \int_{x \geq 0} \frac{x}{M} dF(x) \right.$$

$$\left. - \int_{x \geq -m} \left(\frac{x}{m} + 1\right) dF(x) + \int_{x \geq 0} \frac{x}{m} dF(x) \right\}$$

$$= A \left\{ \int_{0 > x \geq -M} \left(\frac{x}{M}\right) dF(x) - \int_{0 > x \geq -m} \frac{x}{m} dF(x) \right.$$

$$\left. + \int_{-m > x \geq -M} dF(x) \right\}$$

$$= A \left\{ \int_{0 > x \geq -m} \left(\frac{x}{M} - \frac{x}{m}\right) dF(x) \right.$$

$$\left. + \int_{-m > x \geq M} \left(\frac{x}{M} + 1\right) dF(x) \right\}$$

$$\geq 0. \tag{23}$$

Notice that the last inequality holds because the integrands in both integrations are nonnegative.

Corollary: $E[\tilde{Y} - C]^+ \geq E[Y - C]^+$.

Proof: From Theorem 1, we have

$$E[\tilde{Y} - C]^+ = E\left[\sum_{i=2}^{N} \tilde{Y}_i + \tilde{Y}_1 - C\right]^+$$

$$\geq E\left[\sum_{i=2}^{N} \tilde{Y}_i + Y_1 - C\right]^+$$

$$\geq E\left[\sum_{i=3}^{N} \tilde{Y}_i + Y_1 + Y_2 - C\right]^+$$

$$\cdots$$

$$\geq E[Y - C]^+.$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments on an earlier version of this paper.

REFERENCES

[1] M. Decina, T. Toniatti, P. Vaccari, and L. Verri, "Bandwidth assignment and virtual call blocking in ATM networks," in *IEEE INFOCOM*, 1990, pp. 881–888.

[2] H. Esaki, K. Iwamura, and T. Kodama, "A simple and effective call admission method for an ATM network," in *IEEE GLOBECOM*, 1990, pp. 300.5.1–300.5.6.

[3] H. Esaki, "Call admission control method in ATM networks," in *IEEE ICC*, 1992, pp. 354.4.1–354.4.6.

[4] T. Kamitake and T. Suda, "Evaluation of an admission control scheme for an ATM network considering fluctuations in cell loss rate," in *IEEE GLOBECOM*, 1989, pp. 49.4.1–49.4.7.

[5] B. Jabbari and F. Yegenolu, "An upper bound of cell loss probability of bursty sources in broadband packet networks," in *IEEE ICC*, 1991, pp. 23.3.1–23.3.6.

[6] ———, "An efficient method for computing cell loss probability for heterogeneous bursty traffic in ATM networks," *J. Dig. Analog Comm.*, pp. 39–48, 1992.

[7] T. Murase, H. Suzuki, S. Sato, and T. Takeuchi, "A call admission control scheme for ATM networks using a simple quality estimation," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 1461–1470, Dec. 1991.

[8] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 968–981, 1991.

[9] A. I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," in *IEEE INFOCOM*, 1993, pp. 256–265.

[10] B. Maglaris, D. Anaastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, no. 7, pp. 834–844, July 1988.

[11] J. A. S. Monteiro, M. Gerla, and L. Fratta, "Statistical multiplexing in ATM networks," in *Proc. 4th Int. Conf. Data Commun. Perform.*, Barcelona, Spain, 1990, pp. 148–162.

[12] T. H. Lee, K. C. Lai, and S. T. Duann, "Call admission control for ATM networks with nonrenewal inputs," National Chiao Tung University, Tech. Rep. NSC-83-0408-E009-026, 1994.

[13] S. M. Ross, *Stochastic Processes*. New York: Wiley, 1983.

Tsern-Huei Lee (S'86–M'87) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, ROC, in 1981, the M.S. degree from the University of California, Santa Barbara, in 1984, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1987, all in electrical engineering.

Since 1987, he has been a member of the faculty of the National Chiao Tung University, Hsinchu, Taiwan, ROC, where he is a Professor of the Department of Communication Engineering and a Member of the Center for Telecommunications Research. His current research interests are in communication protocols, broadband switch architectures, and traffic management.

Dr. Lee received the Outstanding Paper Award from the Institute of Chinese Engineers in 1991.
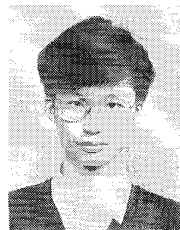
Kuen-Chu Lai received the B.S. degree in communication engineering from the National Chiao Tung University, Hsinchu, Taiwan, ROC, in 1989. He is currently pursuing the Ph.D. degree in electronic engineering at the National Chiao Tung University, Hsinchu, Taiwan, ROC.

His principal areas of interest include ATM switch architecture, connection admission control, and stochastic ordering.
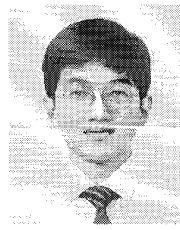
Shii-Tyng Duann received the B.S. degree in electrical engineering from the National Sun Yat-Sen University, Kaohsiung, Taiwan, ROC, in 1991, and the M.S. degree in communication engineering from the National Chiao Tung University, Hsinchu, Taiwan, ROC, in 1993.

He then worked as a Group Leader in the Department of Hardware and Telecommunication Network, ROC Army, until he finished his mandatory military service in 1995. He is currently a graduate student of the Department of Electrical Engineering, Stanford University, CA. His research interests include congestion control, multimedia communications systems, and wireless networks.