A Simple Bandwidth Management Strategy Based on Measurements of Instantaneous Virtual Path Utilization in ATM Networks

Kohei Shiomoto, Member, IEEE, Shinichiro Chaki, Member, IEEE, and Naoaki Yamanaka, Senior Member, IEEE

Abstract—A new connection admission control method based on actual virtual path traffic measurements is proposed to achieve high bandwidth efficiency for various types of traffic. The proposed method is based on the measurement of instantaneous virtual path utilization, which is defined as the total cell rate of the active virtual channels normalized by the virtual path capacity. A low-pass filter is used to determine the instantaneous virtual path utilization from crude measurements. A smoothing coefficient formula is derived as a function of peak rate of the virtual channel. The residual bandwidth is derived from the maximum instantaneous utilization observed during a monitoring period. Simulation shows that the proposed method achieves statistical multiplexing gains of up to 80% of the limit possible with optimum control for similar traffic sources. It can be implemented with very simple hardware. The admission decision is simple: the requested bandwidth is compared with the residual bandwidth. This method is therefore well suited for practical asynchronous transfer mode switching systems.

Index Terms—Admission control, ATM, low-pass filter, measurement.

I. INTRODUCTION

ASYNCHRONOUS transfer mode (ATM) will serve as the platform technology in future multimedia networks. In ATM networks, communications are performed in a connection oriented manner. A virtual channel (VC) is established between a source and a destination. Bandwidth management is performed for each virtual path (VP) between adjacent nodes. Because multimedia traffic, such as voice, video, and computer data, are integrated into ATM networks, the cell stream flowing over each VC is considered to be very bursty [1]–[4]. We consider that the cell stream over each VC is modeled by an on–off source. The on–off source sends cells at the peak rate when it has data to send and it does not send cells when it has no data. With bursty traffic, the average rate of the cell stream over a VC is low compared to the peak rate. Therefore, if a peak rate assignment strategy is used, bandwidth efficiency is low. An efficient bandwidth management method is thus needed to obtain high network resource utilization while maintaining high quality of service (QoS).

The main objective of connection admission control (CAC) is to guarantee, for each service, an appropriate QoS while maximizing network utilization. In CAC, the decision as to whether a new VC is admitted or not is based on its impact on network performance, which depends on the traffic characteristics of the VC and on the QoS levels of the existing VC’s [5], [6]. The statistical multiplexing gain \( G \) is defined as the ratio of the maximum number of VC’s admitted into the VP by CAC versus that using the peak assignment strategy. The statistical multiplexing gain depends on the traffic characteristics of the VC’s such as their peak and average rates. We define the VC burstiness factor as the ratio of the peak rate to the average rate of a VC.

Fig. 1 shows the statistical multiplexing gain as a function of both the VC peak rate and the VC burstiness factor when maintaining a cell loss ratio (CLR) of \( 10^{-6} \). We assume that the VP capacity is 149.76 Mbit/s; the CLR is calculated by using the virtual cell loss probability as defined in the bufferless fluid-flow model [7]. The virtual cell loss probability is defined as the ratio of the overflow cell rate to the offered cell rate. No assumptions are made regarding the VC traffic characteristics; only the VC peak and average rates are used to obtain the virtual cell loss probability. According to Fig. 1, we can achieve a statistical multiplexing gain of up to 50 for traffic with a VC peak rate of 10 Mbit/s and VC burstiness factor of 100. This means that up to 50 times more VC’s
can be accommodated by using statistical multiplexing, which reduces the bandwidth cost by up to 98% (\(100 \times (1 - 1/30)\)).

Several CAC methods use traffic descriptors declared by the user at VC setup time [7]–[10]. They assume that the VC’s source traffic characteristics, such as the peak rate, average rate, and mean burst length, are given. Provided that these characteristics are accurate, the methods achieve optimum bandwidth efficiency. If the traffic over the VC could be characterized completely, we could admit the maximum number of VC’s while maintaining the QoS. The average rate and mean burst length are, however, hard to estimate at VC setup time, especially for data communication. (They are intrinsically a posteriori parameters.) To overcome this problem, a measurement-based dynamic CAC method was proposed [11].

Dynamic CAC uses the measured cell arrival distribution [11]. Its algorithm is summarized below.

1. **Measurement of cell arrival distribution:** Let \(s\) and \(T\) denote the measurement time-window size and measurement renewal cycle length, respectively. Empirical distribution \(\hat{a}(k; t)\) predicts that \(k\) cells will arrive in window \(s\) and is measured during period \([t - T, t]\). Actual cell arrival distribution \(\hat{a}(k; t)\) is evaluated using

\[
\begin{align*}
\hat{a}(0; t) &= \alpha a(0; t) + (1 - \alpha) \hat{a}(0; t - T) \\
\hat{a}(k; t) &= \alpha a(k; t) + (1 - \alpha) \hat{a}(k; t - T) \\
\hat{a}(M; t) &= \alpha a(M; t) + (1 - \alpha) \hat{a}(M; t - T)
\end{align*}
\]

where \(\alpha\) denotes a smoothing coefficient.

2. **Upperbound CLR:** If a new VC request with peak rate \(R\) [cells/window] arrives at time \(t\), the upperbound CLR \(CLR(t)\) is calculated as

\[
CLR(t) = \frac{\sum_{k=K}^{M+R}(K, R)(k - K) \hat{a}(k - R; t)}{\sum_{k=K}^{M+R} k \hat{a}(k - R; t)}
\]

where \(K\) denotes the buffer size.

Dynamic CAC was shown to achieve high efficiency for data communication [12]. Unfortunately, implementing the dynamic CAC requires very complicated circuits, such as digital signal processor (DSP’s) and extra hardware logic [13]. To overcome this hardware complexity problem, we propose a measurement-based CAC method that is considerably simpler than the Dynamic CAC, which requires estimating of the entire cell arrival distribution. It is based on an instantaneous VP utilization, which is defined as the total cell rate of the active VC’s normalized by the VP capacity. A low-pass filter (LPF) is used to obtain the instantaneous VP utilization from crude measurements.

In this paper, we describe our proposed bandwidth management method. In Section II, we describe the concept of instantaneous VP utilization and discuss the use of low-pass filter for measurement. We also discuss the relationship between the smoothing coefficient of the LPF and the traffic characteristics. In Section III, we examine how bandwidth efficiency is achieved using the proposed bandwidth management method. Implementation issues are addressed with reference to the dynamic CAC in Section IV. Finally, in Section V, we give a brief summary and comment on future work.

II. BANDWIDTH MANAGEMENT STRATEGY BASED ON INSTANTANEOUS VP UTILIZATION

A. ATM Traffic Modeling

ATM traffic is characterized by three levels of fluctuation: cell, burst, and connection [14]. It has been recognized that we should pay much attention to the burst level behavior in designing admission controls [15]. Fig. 2 shows the relationship between the CLR and buffer size when \(N\) VC’s are multiplexed onto a VP. Each VC is modeled by an on–off source whose on and off periods are exponentially distributed. Cells are sent at peak rate \(R = 10\) Mbit/s during the on period, the mean burst size \(B = 1, 10, 100\) Kbytes, and the burstiness factor \(\alpha = 10\).

In the region where the buffer size is less than around ten cells, the CLR decreases sharply as buffer size increases. This slope is determined by how often cells from different VC’s arrive simultaneously in a certain cell slot; it is approximated by the model with the same average offered load [16]. This region is called the cell level region. In this region, we can reduce the CLR dramatically by enlarging the buffer. In the region where the buffer size is larger than around ten cells, the slope becomes gentle. This region is called the burst level region. In this region, the buffer should be sufficiently large to absorb the excess bursts when overload conditions arise. As shown in Fig. 2, the reduction in CLR due to enlarging the buffer depends on the mean burst length. Thus, in the burst level region, we must consider the burst length distribution in order to keep the CLR below a target value. However, estimating the burst length distribution at VC setup time is difficult. Moreover, recent research shows that actual LAN traffic exhibits heavily tailed distributions, which cannot be absorbed using ATM buffers with a feasible size [17]. Accordingly, we should not rely on buffering excess bursts to attain the target CLR.

In Fig. 2, the CLR at the knee points corresponds to the virtual cell loss probability in the buffer-less fluid-flow model [7], [16]. Note that the virtual cell loss probability does not
Fig. 3. Conceptual view of definition of instantaneous VP utilization.

Fig. 4. Concept of proposed bandwidth management strategy.

B. Proposed Bandwidth Management Based on Instantaneous VP Utilization

We propose the concept of instantaneous VP utilization to capture burst level fluctuations and a method to measure it. Burst level fluctuations are characterized by the number of active VC’s, as illustrated in Fig. 3. Instantaneous VP utilization $\lambda(t)$ is defined as

$$\lambda(t) = \sum_{V(t) \text{ is active at time } t} \frac{R_i}{C}$$

where $R_i$ denotes the peak rate of VC $i$. The number of active VC’s changes with time and so does the instantaneous VP utilization.

Measurement of the instantaneous VP utilization plays an important role in our bandwidth management method. To measure the instantaneous VP utilization, we use a recursive LPF whose smoothing coefficient is $\alpha$

$$\lambda(t) = \alpha n(t) + (1 - \alpha)\lambda(t - \Delta), \quad 0 \leq \alpha \leq 1$$

where $n(t)$ denotes the number of cells arriving during the $t$th cell slot, and $\Delta$ denotes the single cell transmission time over the VP. Smoothing coefficient $\alpha$ controls the cutoff frequency of the LPF. As it approaches 1, the cutoff frequency increases. We regard time series data of the cells observed during unit time $\Delta$ as a mixture of many signals with different and diversified frequencies. As will be shown in Section II-C, if an appropriate cutoff frequency is set, we can obtain an accurate instantaneous VP utilization.

Our proposed bandwidth management strategy is illustrated in Fig. 4. The cells arriving over a VP during a cell transmission time slot are counted and this count is then converted into the instantaneous VP utilization by the LPF. Instantaneous VP utilization is calculated for every cell slot. Instantaneous VP utilizations are tracked for a monitoring period. The maximum instantaneous VP utilization observed during the monitoring period $\lambda_{\text{max}}(t)$ is used as the admission criteria. The residual bandwidth is defined as $1 - \lambda_{\text{max}}(t)$. If the requested bandwidth is lower than the residual bandwidth, the request is accepted; otherwise it is rejected. The maximum instantaneous VP utilization $\lambda_{\text{max}}(t)$ is defined as

$$\lambda_{\text{max}}(t) = \max_{t' \in [t - T_m, t]} \lambda(t')$$

where $T_m$ denotes the monitoring period. Thus, the admission criteria can be rewritten as

$$R/C < 1 - \lambda_{\text{max}}(t)$$

for a new VC setup request whose peak rate is $R$, arriving at a VP whose capacity is $C$.

If the VC is accepted, the observed instantaneous VP utilization is increased by an amount equal to the accepted VC’s peak rate

$$\lambda_{\text{max}}(t') = \lambda(t') + R/C, \quad \text{for } t' \in [t - T_m, t].$$

Equation (7) is used so as to improve robustness of our proposed method. Even if the newly admitted VC continuously sends cells at the peak rate, the instantaneous VP utilization cannot statistically exceed 1 for period $T_m$, provided that the admission criteria given in (6) are satisfied.

$T_m$ is related to the CLR: $T_m$ needs to be very long to achieve a very low CLR. To reduce $T_m$, we introduce target load $\lambda_{\text{target}} (< 1)$. The admission criteria is modified accordingly

$$R/C < \lambda_{\text{target}} - \lambda_{\text{max}}(t).$$

The effect of $\lambda_{\text{target}}$ and the relationship between $T_m$ and the target CLR will be discussed in more detail in Section II-D.

C. Smoothing Coefficient $\alpha$

The VC’s peak rate, which is equivalent to the minimum cell interval, is closely related to the instantaneous VP utilization in the frequency domain. Specifically, we should remove those frequency components corresponding to continuous bit rate
(CBR) signals whose cell rate is higher than the VC peak cell rate, so as to accurately obtain the instantaneous VP utilization. We assume that three VC’s (VC₁, VC₂, VC₃) with a peak rate $R = 30$ Mbit/s are multiplexed into a VP with a capacity $C = 150$ Mbit/s, as shown in Fig. 5. The pattern of cells arriving over the VP in the time domain varies depending on the differences in each VC’s cell transmission timing. (See the difference between pattern 1 and pattern 2 in Fig. 5.) In the frequency domain, the discrete Fourier transform (DFT) of the number of cells arriving over the VP is equal to the sum of the DFT’s of the number of cells arriving over each VC due to the linearity of the DFT’s, irrespective of the differences in each VC’s cell transmission timing. The $i$th frequency component of the number of cell arrivals over a VP, $Y_i$, is expressed as $\sum_j X_{i,j}$, where $X_{i,j}$ indicates the $i$th frequency component of the number of cell arrivals over VC$_j$; the basic frequency interval is the peak rate of VC$_j$. In the frequency domain, the direct current (dc) component $Y_0$ indicates the contribution of VC$_j$ to the instantaneous VP utilization. In other words, the magnitude of $Y_0 (= \sum_j X_{0,j})$ is equal to the instantaneous VP utilization. Thus, the instantaneous VP utilization is determined by the dc component of the multiplexed VC’s in the frequency domain. Consequently, we only have to remove all frequency components other than dc component $Y_0$ to obtain the instantaneous VP utilization. This is done by setting the cutoff frequency to the VC peak rate.

Smoothing coefficient $\alpha$ is set to remove all frequency components higher than the VC peak cell rate. Consider $\alpha$ in (4). The power spectral density function of the LPF given in (4) is given by

$$S(\omega) = |H(\omega)|^2 = \frac{\alpha^2}{1 + (1 - \alpha)^2 - 2(1 - \alpha) \cos(\omega \Delta)}$$  \hspace{1cm} (9)$$

where $\omega$ is frequency in radians ($\omega = 2\pi f$ and $f$ is the VC peak cell rate). As mentioned above, we can capture the burst level behavior by eliminating all frequency components higher than the peak cell rate. We do this by setting $\alpha$ such that $S(\omega_0)$ is lower than $\epsilon$, where $\omega_0$ denotes the frequency (in radians) corresponding to the peak cell rate. Solving (9) in terms of $\alpha$, we obtain

$$\alpha = \frac{-2(1 - K) + \sqrt{4(1 - K)^2 + 8(\epsilon^{-1} - 1)(1 - K)}}{2(\epsilon^{-1} - 1)}$$  \hspace{1cm} (10)$$

where $K = \cos(\omega_0 \Delta)$. Fig. 6 shows the relationship between the peak rate and smoothing coefficient $\alpha$ for three values of $\epsilon$. For example, an $\alpha$ of $4.156 \times 10^{-3}$ is adequate for VC’s with a peak rate of 10 Mbit/s on a VP whose capacity is 149.76 Mbit/s. Note that this holds as long as the VC/VP ratio remains the same.

To evaluate the effect of the LPF, we used computer simulations. We obtained time series data from crude measurement data (i.e., the number of cells counted per unit time), then calculated the instantaneous VP utilizations by using different smoothing coefficients. We assumed that the VP capacity
Fig. 7. Sample path of crude measurement and accurate instantaneous VP utilization from computer simulation. (a) Crude measurement (number of cells observed in a unit time). (b) Accurate instantaneous VP utilization.

Fig. 8. Sample path of the filtered instantaneous VP utilization from computer simulation. (a) Measurement $\alpha = 1.0 \times 1$. (b) Measurement $\alpha = 4.156 \times 3$. 

$C = 149.76$ Mbit/s, the number of the VC’s $(N) = 50$, the mean burst length of the VC’s $(B) = 100$ Kbytes, the peak rate of the VC’s $(P) = 10$ Mbit/s, and the burstiness factor $(\alpha) = 10$.

The crude measurement data and accurate instantaneous VP utilization rates are shown in Fig. 7(a) and (b), respectively. The crude measurements are far from the accurate instantaneous VP utilization. The crude measurement data fluctuates strongly, and the values are widely distributed because they reflect the cell level behavior. Fig. 8(a) and (b) show the sample paths of instantaneous VP utilizations observed using an LPF with different value of $\alpha$. A large $\alpha$ of $1.0 \times 1$ [Fig. 8(a)] does not eliminate the cell-level fluctuation, while an adequate $\alpha$ of $4.156 \times 3$ [Fig. 8(b)] yields an accurate instantaneous VP utilization. The key is to set the smoothing coefficient according to (10), which yields a quite realistic observed instantaneous VP utilization rate.

**D. Monitoring Period $T_m$**

Monitoring period $T_m$ needs to be dimensioned so that the actual CLR is lower than its target value. $T_m$ depends on the target CLR and the traffic characteristics of the multiplexed VC’s. We derived $T_m$ by approximating the 99% cumulative value of the duration of an underload state. We define an underload state as being that when the instantaneous VP utilization does not exceed target load $\lambda_{\text{target}}$.

1) Distribution of Underload Period: We will now derive the distribution of the underload period when $N$ on–off sources are multiplexed. The durations of the on and off periods are distributed exponentially with means of $\omega^{-1}$ and $\gamma^{-1}$, respectively. The number of active VC’s at time $t$ follows a birth–death process denoted by $N(t)$.

Let $T$ be a random variable for the duration of the underload period. It is the absorption time of a transient Markov process whose state space is $\{N(t) : N(t) = 0, \ldots, N_u\}$. The absorbing state is $N_u + 1$, where we denote the border between the underload and overload states by $N_u (= \lceil \lambda_{\text{target}}C/R \rceil)$. (If more than $N_u$ sources are active, the state is regarded as overload.)

The probability density function is given as

$$f(t) = \delta \exp(Qt)\gamma^\alpha$$

where

$$\delta = (0, \cdots, 0, 1)$$
and $Q$ is an infinitesimal generator whose $(i, j)$-elements $Q_{ij}$ are given by

$$Q^p = (0, \cdots, 0, (N - N_\mu + 1)\gamma)^T$$  \hspace{1cm} (13)

The first and second moments of $T$ are given by the following formulas [19, p. 45]:

$$E[T] = \delta(-Q^{-1})^2Q^p$$ \hspace{1cm} (15)
$$E[T^2] = 2\delta(-Q^{-1})^3Q^p.$$ \hspace{1cm} (16)

Standard deviation $\sigma$ is calculated from these first and second moments. By applying Chebyshev’s inequality, we can determine the 99% point of the distribution of $T$ [20, p. 388]:

$$\text{Prob}(\mid T - E[T] \mid \geq k\sigma) \leq \frac{1}{k^2}.$$  \hspace{1cm} (17)

2) Numerical Example: We will now evaluate the effect of the traffic characteristics on $T_m$ assuming that the CLR objective is $1.0 \times 10^{-6}$. Table I shows the maximum number of VC’s (with various VC traffic characteristics) that can be handled in a VP with a capacity of 149.76 Mbit/s. The number of VC’s is calculated by the virtual cell loss probability [7]. We approximated the 99% cumulative value of the underload period under the conditions listed in Table I. The effect of the peak rate, burstiness, and mean burst length on $T_m$ are plotted in Figs. 9–11, respectively. We can observe the following from these plots:

- Monitoring period $T_m$ can be reduced to a small value by introducing $\lambda_{\text{target}}$. A $\lambda_{\text{target}}$ of 0.8 reduces the $T_m$ to only 1/10th to 1/100th the values required for the normal condition, i.e., $\lambda_{\text{target}} = 1.0$.
- Monitoring period $T_m$ depends on mean burst length $B$, but is affected little by burstiness $a$ and peak rate $R$. Thus, the maximum burst length should be limited or $T_m$ should be modified adaptively by using a feedback mechanism. A $T_m = 10$ s is sufficiently large for the entire range of burstiness $a$ and peak rate $R$ (with $\lambda_{\text{target}} = 0.8$).

### III. Performance Evaluation

#### A. Bandwidth Efficiency

We evaluate the performance of our method through computer simulation. We modeled the cell, burst, and connection-level arrival and departure patterns. To investigate the robustness of our method, we assumed that the VC connection arrival
Fig. 12. Admitted VC’s as a function of time. For the effective bandwidth method proposed by Elwalid and Mitra [10], the buffer size is set to 1000 cells. (a) VC peak rate = 1.5 Mbit/s. (b) VC peak rate = 10 Mbit/s.

the VP capacity, though the proposed method is measurement-based while the virtual CLR method uses the user-declared VC peak rate and average rate. Note that these two methods yield comparable bandwidth efficiency.

The proposed method efficiently admits more VC’s than the effective bandwidth method, as shown in Fig. 12. The reason is as follows. The proposed method anticipates the statistical multiplexing gain between different VC’s. Generally speaking, this approach achieves high efficiency when the VC peak rate is small compared to the VP capacity. In contrast, the effective bandwidth method does not exploit a statistical multiplexing gain between VC’s, so is too conservative except when the VC peak rate is high and the buffer is sufficiently large (comparable to the VC average burst size). Here we assume that the VC peak rate and buffer size are small compared to the VC burst size. Therefore, the proposed method yields a higher bandwidth efficiency than does the effective bandwidth method.

The proposed method admits around 80% (around 65%) of the number of VC’s admitted by the virtual cell loss probability method for VC’s with $R = 1.5$ Mbit/s (10 Mbit/s). Note that the number of VC’s admitted by the virtual cell loss probability method may become lower than that shown in Fig. 12 if the declared VC average rate and average burst length contain the safety margin to offset their inaccurate estimations, as mentioned in Section I.

The efficiency at $R = 1.5$ and 10 Mbit/s differ due to the statistical multiplexing effect. That is, an $R = 1.5$ Mbit/s is more readily acceptable for the same residual bandwidth than an $R = 10$ Mbit/s.

Also as shown in Fig. 12, the proposed method more than triples the bandwidth obtained using the peak assignment method with an $R = 10$ Mbit/s, while the increase is tenfold for an $R = 1.5$ Mbit/s. This advantage is due to use of direct traffic measurements. In the proposed method, the measurements are obtained using a simple LPF described in Section II-B. If the average rate and average burst length cannot be accurately determined beforehand, the peak assignment method is the most suitable solution. Our proposed CAC method is more effective for multimedia traffic with unknown traffic characteristics.

IV. IMPLEMENTATION

Because admission control is invoked when the connection setup, the admission decision should be quick so as to minimize the connection setup delay. Namely, computational complexity in the admission decision procedure should be minimized. In addition, from the viewpoint of switching node cost, extra hardware and software control should be avoided. In this section, we address the computational complexity in the admission decision and describe a simple hardware implementation of our proposed method.

A. Hardware Implementation

As described in Section II-B, our proposed method uses an LPF to obtain the instantaneous VP utilization. The cells are counted and the instantaneous VP utilization is calculated
for every cell slot by using the LPF in (4). By expressing smoothing coefficient $\alpha$ as a power of two, i.e., $2^{-k}$, we can avoid the floating point operations and thus implement the LPF as a combination of an adder, a subtracter, and two shifters

$$\lambda(t) = 2^{-k}n(t) + (1 - 2^{-k})\lambda(t - \Delta).$$  \hspace{1cm} (18)

As shown in (18) and Fig. 13, the LPF is composed of one adder, one subtracter, and two $k$-bit-right shifters.

In our proposed method, the residual bandwidth is derived from the maximum of the observed instantaneous VP utilization. We therefore have to keep track of the instantaneous VP utilization over the monitoring period. Because the maximum instantaneous VP utilization is calculated every cell slot, we would need a tremendous amount of memory if we stored all the instantaneous VP utilization rates during the monitoring period. Fortunately, only a few of them are required to calculate the maximum instantaneous VP utilization rate. We reduce the storage capacity by dividing the monitoring period into $n$ bins, as shown in Fig. 14. Cyclic queue $\Lambda[i]$ ($i = 0, \cdots, n-1$) contains maximum instantaneous VP utilization rate associated with each bin. Two registers, $\lambda$ and $\Lambda$, contain the current instantaneous VP utilization and the maximum instantaneous VP utilization among all bins, i.e., $\Lambda = \max_i \Lambda[i]$. Now suppose that the current cell slot is the first cell slot of the $(n-1)$th bin. The instantaneous VP utilization is calculated for the current cell slot and loaded into $\Lambda[n-1]$. If it is larger than the value $\Lambda$, it is also loaded into $\Lambda$. In the next cell slot, i.e., the second cell slot of the $(n-1)$th bin, the instantaneous VP utilization rate is calculated again and checked to see if it is larger than $\Lambda[n-1]$ or not. If it is, it is loaded into $\Lambda[n-1]$. The current instantaneous VP utilization rate is also checked to see if it is larger than $\Lambda$ or not. If it is, $\Lambda$ is replaced by the current instantaneous VP utilization rate as well. The same procedure is repeated until the end of the bin is reached. At the end of the $(n-1)$th bin, the maximum instantaneous VP utilization is selected from among all $\Lambda[i]$, excluding $\Lambda[0]$, which is used to store the maximum instantaneous VP utilization observed in the next bin. The implementation complexity depends on the number of bins.

The relationship between the average number of connections and the number of bins is shown in Fig. 15. The impact of the number of bins is quite small. A two-bin configuration achieves almost the same average number of connections as a ten-bin one. The complexity of the measurement process is thus considerably reduced while the average number of connections is maintained.

B. Computational Complexity of the Admission Decision

The admission decision should also be quick to avoid long connection setup latency. We compare the computational complexity of our proposed method to that of dynamic CAC.

In dynamic CAC, the decision is made by comparing the target CLR to the estimated CLR by using the upper bound formula for the cell loss ratio

$$\text{CLR}(t) = \frac{\sum_{k=\max[K,R]}^{M+R} (k-K) \cdot \hat{a}(k-R;t)}{\sum_{k=R}^{M+R} k \cdot \hat{a}(k-R;t)}$$  \hspace{1cm} (19)

where $K$ denotes the buffer size and $M$ denotes the maximum number of cells arriving within a window whose size is $K$ cell transmission times on a VP. If the ATM switching fabric has $N$ input ports, $M = NK$. Even though the denominator in (19) is equal to the average number of arriving cells within the window, the numerator requires approximately $3K(N-1)$ floating point operations. Our proposed method admission decision procedure described in Section II-B is simpler. Admitting a new connection is simply judged by
comparing $\lambda_{\text{max}}(t) + R$ to $\lambda_{\text{target}}$, which requires only two operations.

The computational complexities of our proposed CAC and dynamic CAC are plotted in Fig. 16. We assume that the buffer size $K = 128$ cells. Dynamic CAC requires hundreds of operations to calculate the cell arrival distribution and a DSP to calculate the CLR. In contrast, our proposed CAC, which incurs much less computational complexity, can be implemented with only a couple of registers. Compared to the conventional admission control method, our proposed method is suited for real-time processing, a necessity in practical ATM switching systems. Because our proposed method can be implemented with reasonable hardware complexity and the admission decision is fast, it is better suited for practical ATM switching systems.

V. CONCLUDING REMARKS

To achieve high network efficiency, we proposed a connection admission control method that is based on measurement of the instantaneous VP utilization rate. A recursive LPF is used to measure the instantaneous VP utilization rate from the number of cells observed during one slot. The smoothing coefficient $\alpha$ of the LPF is dimensioned so as to eliminate frequency components higher than the peak cell rate of the VC. For example, $\alpha = 4.156e-3$ yields instantaneous VP utilization close to the theoretical values when 10-Mbit/s peak rate VC’s are multiplexed onto a 149.76-Mbit/s VP. We found that monitoring period $T_m$ depends on the mean burst length and not on the peak rate or burstiness. When the mean burst length is 10 Kbytes and the target load is 0.8, a monitoring period of 10 s is sufficiently large for all the peak rates and burstiness levels examined in this paper. To cope with the dependency of $T_m$ on the mean burst length, the maximum burst length should be limited or $T_m$ should be modified adaptively by using a feedback mechanism. Simulation showed that the proposed method achieves 65% or 80% of the optimum statistical multiplexing gain for VC’s with a peak rate of 10 or 1.5 Mbit/s. The proposed method can be implemented using very simple hardware, and the admission decision is simple: the requested bandwidth is compared with the residual bandwidth. Compared to conventional admission control methods, the computational complexity in the admission decision is much lower, so the proposed method is well suited for practical ATM switching systems.

Future tasks include developing a method for adaptively modifying $T_m$ based on the measured burst length by using a feedback mechanism. Also, in this paper, we assumed the multiplexing of homogeneous VC’s, which does not hold in practice. A method for determining a smoothing coefficient for heterogeneous VC’s is thus needed.

ACKNOWLEDGMENT

The authors are grateful to the reviewers for their helpful comments.

REFERENCES

Kohei Shiomoto (M’89) received the B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University, Osaka, Japan, in 1987, 1989, and 1998, respectively.

He is a Senior Research Engineer at the NTT Network Service Systems Laboratories, Tokyo, Japan, where he has been working in ATM traffic control and high-speed networking architecture since he joined NTT. He was engaged in development of the NTT’s first commercial ATM switching system from 1989 to 1994. From August 1996 to September 1997, he was engaged in research on high-speed networking as a Visiting Scholar at Washington University, St. Louis, MO.

Dr. Shiomoto is a member of the Institute of Electronics, Information and Communication Engineers (IEICE). He received the Young Engineer Award of the IEICE in 1995.

Shinichiro Chaki (M’93) received the B.E. degree from Sophia University and the M.E. degree from Waseda University.

He is a Senior Research Engineer at NTT Access Network Systems Laboratories, Chiba, Japan, where he does research on access network design.

Naoaki Yamanaka (M’85–SM’96) received the B.E., M.E. and Ph.D. degrees in engineering from Keio University, Japan, in 1981, 1983, and 1991, respectively.

In 1983 he joined Nippon Telegraph and Telephone Corporation’s (NTT’s) Communication Switching Laboratories, Tokyo, Japan, where he was engaged in the research and development of a high-speed switching system and high-speed switching technologies. Since 1989, he has been active in the development of Broadband ISDN based on ATM techniques. He is currently a Senior Research Scientist, Supervisor, Distinguished Technical Member in Broadband Network System Laboratory at NTT. He is now researching future ATM-based broadband ISDN architecture and traffic management and performance analysis of ATM networks.

Dr. Yamanaka is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan. He received Best of Conference Awards from the 40th and 44th IEEE Electronic Components and Technology Conference, the TELECOM System Technology Prize from the Telecommunications Advancement Foundation, twice received the IEICE Switching System Research Award, and received the IEEE CPMT Transactions Part B Best Transactions Paper Award in 1990, 1994, 1996, and 1998, respectively. He is Broadband Network Area Editor of IEEE Communication Surveys, Associate Editor of *IEICE Transactions*, IEICE Communication Society International Affairs Director, as well as Secretary of the Asia Pacific Board of the IEEE Communications Society.