A Hybrid ATM Connection Admission Control Scheme based on On-Line Measurements and User Traffic Descriptors

Daryoush Habibi , Guoqiang Mao
Edith Cowan University
Perth, Australia
d.habibi@cowan.edu.au , g.mao@cowan.edu.au

Abstract

Connection Admission Control (CAC) is a preventive congestion control mechanism and it is essential for providing Quality of Service (QoS) requirements for calls established over ATM. The strategies for CAC in an ATM network can be classified into two categories: measurement-based CAC and traffic descriptor-based CAC. Both categories have their advantages and disadvantages. This paper proposes a simple CAC algorithm which utilizes both traffic descriptors and parameters that can be reliably and accurately derived from on-line measurements. This hybrid CAC algorithm avoids the disadvantage of traffic descriptor-based CAC in which some traffic parameters can not be tightly characterized by sources and the disadvantage of measurement-based approach in which some measurements are not reliable and/or time-consuming. CAC decision can be made easily in our algorithm. Simulation studies are presented which are indicative of the good performance of our algorithm.

1 Introduction

To date, many CAC schemes have been proposed [3]-[6]. These schemes can be classified into two categories: measurement-based CAC and traffic descriptor-based CAC. Traffic descriptor-based CAC uses the a priori characterization provided by sources at call setup phase to calculate the worst-case behaviour of all existing calls in addition to the incoming one. Performance of traffic descriptor-based CAC relies on the precision of parameters provided by sources at call setup. However, it is difficult for the user to accurately characterize such parameters as mean cell rate and maximum burst length. Moreover, in order to provide a tight QoS bound, traffic descriptor-based CAC has to allocate bandwidth according to the worst case behaviour of calls. Network utilization under this algorithm is acceptable when traffic flows are smooth; when traffic flows are bursty, however, it will inevitably results in low utilization [2]. Measurement-based CAC uses the a priori source characterizations only for incoming call and uses measurements to characterize existing calls. Therefore, network utilization does not suffer significantly if the traffic descriptions are not tight. Because of its reliance on on-line measurements and the fact that source behavior is not static in general, the measurement-based CAC can never provide a completely reliable QoS bound for calls [3]. Furthermore, when there are only a few calls present, the unpredictability of individual call's behavior dictates that these measurement-based approaches must be very conservative. Thus, a measurement-based CAC can only deliver significant gain in utilization when there is a high degree of statistical multiplexing.

In this paper we propose a CAC which takes a different approach. We identify that some parameters such as peak cell rate can be characterized by sources with reasonable accuracy while other parameters as mean cell rate of VP can be reliably derived from on-line measurements. We base our CAC model on those reliable parameters from both traffic descriptors of sources and on-line measurements. Our algorithm aims at achieving the maximum benefit from statistic multiplexing while at the same time providing reasonably reliable QoS guarantees for admitted calls. The calculations involved with CAC decision of our algorithm are very simple which makes it suitable for the real time requirements of CAC in ATM networks.

Cell loss 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 cell loss is used in many papers as the QoS index[4], [5], [7], [8]. Therefore, we use cell loss as the QoS index in our CAC scheme.

The rest of the paper is organized as follows: in section two we introduce our traffic model and perform some theoretical analysis. Our CAC algorithm is given in section three. Section four describes our simulation procedures and outlines some simulation results. Finally, we make some

0-7695-0777-8/00 $10.00 © 2000 IEEE

concluding remarks in section five.

2 Bufferless Traffic Model

Many measurement-based CAC attempt to capture the instantaneous traffic rate of a VP, or capture other parameters such as variance and autocorvariance from on-line measurements. However, sources are not static and can sometimes be very bursty. This makes the estimation of instantaneous traffic rate or other parameters either unreliable or computationally ineffective for accurate measurements.

In our CAC algorithm, we try to avoid measurements and estimation of such parameters as instantaneous traffic rate, variance and autocovariance of traffic rate which either can not be estimated with reasonable accuracy or will consume much time for accurate estimation. We build our model on those parameters which can be reliably derived from traffic descriptors and measurements, specifically, peak cell rate from traffic descriptors and mean VP traffic rate from on-line measurements. We identify the accuracy of the two parameters are much more reliable than other parameters. In our model, we use the measured mean cell rate of the VP and the declared peak cell rates of calls to predict the cell loss ratio of the VP and make CAC decisions accordingly. For the purpose of modelling the aggregate traffic of the VP we use binomial distribution.

When the number of multiplexed connections is large the traffic distribution can be best modeled by Gaussian distribution. However, using central limit theorem, the binomial distribution $f(k) = \binom{N}{k} p^k(1-p)^{N-k}$ approximates to Gaussian distribution when N is large. The condition of N being large is equivalent to the number of multiplexed connections being large, under which Gaussian distribution applies. Fig. 1 shows the comparison between Gaussian distribution and binomial distribution with the same mean and variance. As shown in Fig. 1, within the precision range required for the calculation of less stringent Cell Loss Ratio (CLR), binomial distribution approximates Gaussian distribution well when N is large. For measurement-based CAC, limited by the accuracy of measurements and estimation, it makes little sense to choose a very stringent CLR objective, say for example, more stringent than 10^{-8}. Therefore, the use of binomial distribution to approximate the aggregate traffic distribution is a reasonable choice.

Moreover, binomial distribution has wide applications in traffic modeling. For example, multiplexing of homogeneous on-off sources has binomial traffic distribution. By choosing binomial distribution in mathematical modelling of our CAC algorithm, as shown later, we avoid the estimation of high order moments of traffic rate. Our algorithm also shows that binomial distribution suits real-time calculation.

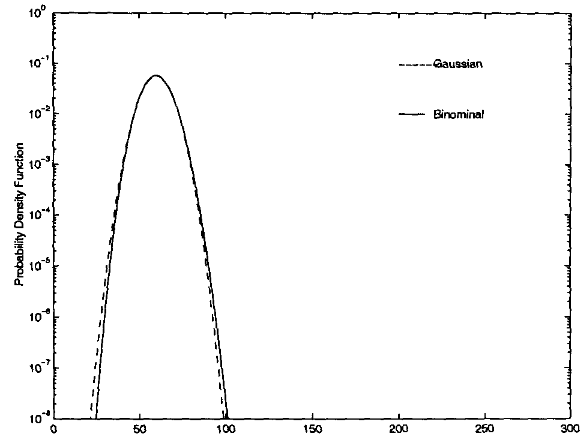


Figure 1. Comparison between binomial distribution and Gaussian distribution with $mean = 60$ **and** $\sigma^2 = 48$

For simplicity let us consider the case of an on-off source. On-off source models have been widely adopted to approximate real traffic sources [1], [7]. In active periods cell are generated at a peak cell rate denoted by PCR. In idle periods no cells are generated. Let AVG denote the average cell rate of a traffic source. Hence, the traffic source generates traffic at peak cell rate with probability AVG/PCR, and stays in idle state with probability $1 - AVG/PCR$. Let $f_{new}(x)$ represent the probability density function (pdf) of the new call:

$$f_{new}(x) = \begin{cases} \frac{AVG_{new}}{PCR_{new}} & if \quad x \quad = \quad PCR_{new} \\ 1 - \frac{AVG_{new}}{PCR_{new}} & if \quad x \quad = \quad 0 \end{cases} \tag{1}$$

where the subscript new denotes parameters of the new call. The traffic distribution of existing calls, denoted by $f(x)$, is assumed to be a binomial distribution. Therefore if the new call is admitted, the aggregate traffic distribution will change to:

$$q(x) = f(x) * f_{new}(x) \tag{2}$$

where $*$ denotes convolution operation. Let ET denote the excess traffic after the new call is admitted and ρ represent the traffic load. The cell loss ratio when the new call is admitted can be estimated as follows:

$$CLR = \frac{ET}{\rho} \tag{3}$$

where

$$ET = E\left[(R-C)^+\right] = \sum_x (x-C)^+ q(x) \tag{4}$$

$$\rho = M + AVG_{new} \quad (5)$$

In the above equations R denotes the instantaneous VP traffic rate, C denotes the VP capacity, and M is the mean traffic rate of the VP before the new call is admitted. If the estimated CLR is less than the cell loss ratio objective, the new call is admitted, otherwise, it will be rejected.

3 CAC Algorithm

Let us select a traffic rate u, less than or equal to the smallest value of the peak cell rates of all calls in the VP, as a traffic rate unit so that the VP capacity C is an integer multiple of u. Furthermore, u should be much smaller than the VP capacity: $u < \frac{C}{100}$.

The sum of the declared peak cell rates of all calls in the VP is recorded and denoted by S. The mean traffic rate of the VP is measured and denoted by M. Since we assume the aggregate traffic has binomial distribution we need to calculate N and p for binomial distribution $f(k) = \binom{N}{k} p^k(1-p)^{N-k}$.

N is calculated as follows:

$$N = \left\lceil \frac{S}{u} \right\rceil \quad (6)$$

where $\lceil \bullet \rceil$ denotes the smallest integer greater than or equal to $\bullet$.

p can be estimated directly as follows:

$$p = \frac{M}{N \times u} \quad (7)$$

This estimation may be improved by measuring over a longer period. However, since network traffic is bursty in nature errors in measured mean traffic rates are inevitable. Such estimation errors may significantly affect the robustness of the CAC scheme with regards to QoS guarantees [8]. Since $\frac{M}{N \times u}$ has an approximate normal distribution $N(p, p(1-p)/N)$, p can also be estimated as: [5]

$$p = \frac{M}{N \times u} + \alpha \sqrt{\frac{\frac{M}{N \times u}(1 - \frac{M}{N \times u})}{N}} \quad (8)$$

The choice of safety margin α is related to many factors. As a general rule, the larger the measurement interval, and the larger the link capacity, the smaller the α. The introduction of α enables us to greatly shorten the required measurement interval while at the same time maintain the robustness of the CAC scheme. Moreover, the safety margin can introduce an additional benefit. For some ATM switches with large buffers, α can be set to be very small or even zero to reflect the fact that large buffer will effectively absorb bursty traffic therefore reduce cell loss. It introduces some flexibility into the CAC scheme which enables us to efficiently utilize network resources. Hence, Eq. 8 is used to estimate p in our CAC scheme.

Therefore the traffic distribution can be calculated from the declared peak cell rate of each call and the measured mean cell rate of the VP. We define

$$I(m) \triangleq E\left[(R-m)^+\right] \quad (9)$$

$$= \sum_{k=m+1}^{\infty} (k-m)f(k) \quad (10)$$

Note that $f(k) = 0$ for $k > N$. It is easy to derive that $I(0) = Np = M$ and $I(m) = I(0) - m$ for $m < 0$. For $m > 0$, $I(m)$ is calculated as follows:

$$I(m) = I(m-1) - 1 + \sum_{k=0}^{m-1} f(k) \quad (11)$$

As shown later, in our CAC algorithm only the calculation of $I(m)$ from $m = 0$ to $m = C$ is involved. The same conclusion applies to the calculation of $f(k)$. Therefore we only need to calculate $I(0)$ to $I(C)$ here, and in the value update process shown later. The computation complexity of our CAC algorithm is determined by the choice of the traffic rate unit u and the VP capacity.

When a new call arrives, parameters of the new call's traffic descriptor will be quantized for the purpose of calculation. After quantization the traffic parameters of the new call may change. Let $\widetilde{Y}_{new}$ denote the quantized call whose traffic parameters are the quantized traffic parameters of the new call. The density function $g(x)$ of $\widetilde{Y}_{new}$ is then given by

$$g(x) = \begin{cases} \frac{AVG_{new}}{\lceil PCR_{new} \rceil} & if \quad x = \lceil PCR_{new} \rceil \\ 1 - \frac{AVG_{new}}{\lceil PCR_{new} \rceil} & if \quad x = 0 \end{cases} \quad (12)$$

It has been shown [7] that if the VP has enough bandwidth for the quantized call $\widetilde{Y}_{new}$ then the new call can be admitted. Hence from Eq. 3 $\sim$ 5, the estimated cell loss ratio when the new call is admitted will be:

$$CLR = \frac{ET}{\rho} \quad (13)$$

where

$$ET = \sum_{k=C+1}^{\infty} (k-m)f(k) * g(k) \quad (14)$$

$$= \left(1 - \frac{AVG_{new}}{\lceil PCR_{new} \rceil}\right) I(C) \quad (15)$$

$$+ \left(\frac{AVG_{new}}{\lceil PCR_{new} \rceil}\right) I(C - \lceil PCR_{new} \rceil) \quad (16)$$

$$\rho = I(0) + AVG_{new} \tag{17}$$

If the estimated cell loss ratio is less than our CLR objective then the new call is admitted, otherwise the new call is rejected.

If the new call is admitted, $I(m)$ will be updated. It can be shown that $\widehat{I}(0) = I(0) + AVG_{new}$ and $\widehat{I}(m) = \widehat{I}(0) - m$ for $m < 0$. For $m > 0$ we will have:

$$\widehat{I}(m) = \left(1 - \frac{AVG_{new}}{\lceil PCR_{new} \rceil}\right) I(m) \tag{18}$$

$$+ \left(\frac{AVG_{new}}{\lceil PCR_{new} \rceil}\right) I(m - \lceil PCR_{new} \rceil) \tag{19}$$

where the superscript $\widehat{}$ denotes the updated value.

Accordingly, S and M will be updated in the following manner:

$$\widehat{S} = S + PCR_{new} \tag{20}$$

$$\widehat{M} = M + AVG_{new} \tag{21}$$

If an existing call leaves, S will also be updated to keep track of the sum of the peak cell rates of all calls currently in the VP:

$$\widehat{S} = S - PCR_{release} \tag{22}$$

However $I(m)$ and M are not updated for a leaving call. The changes in $I(m)$ and M due to a leaving call are caught up automatically by measurements. Their value will be updated after one measurement interval.

In Eq. 22 the subscript $release$ denotes the parameter of the call to be released. In order to enhance the performance of our CAC algorithm, when the sum of the peak cell rates of existing calls and the new call is less than the VP capacity the new call is admitted directly. Only when the sum of the peak cell rate of all existing calls and the new call is more than the VP capacity, our CAC algorithm is invoked to make the CAC decisions. The flowchat of our CAC algorithm is shown in Fig. 2.

Once our CAC algorithm is invoked, $I(m)$ is calculated and updated using Eq. 6 $\sim$ 11, if during the measurement interval no new calls are accepted.

If there are existing calls that are released during the measurement interval, the released calls contribute to the measured mean traffic rate of the VP. However, S is the sum of the peak cell rates of calls in the VP at the instant of updating $I(m)$, not including the peak cell rates of calls released during the measurement interval. Hence in Eq. 7, the calculated activity parameter p will be greater than the actual activity parameter. Therefore our CAC algorithm becomes more robust and efficient with regard to the CLR objective by not considering the effect of call departure in periodical update of $I(m)$.

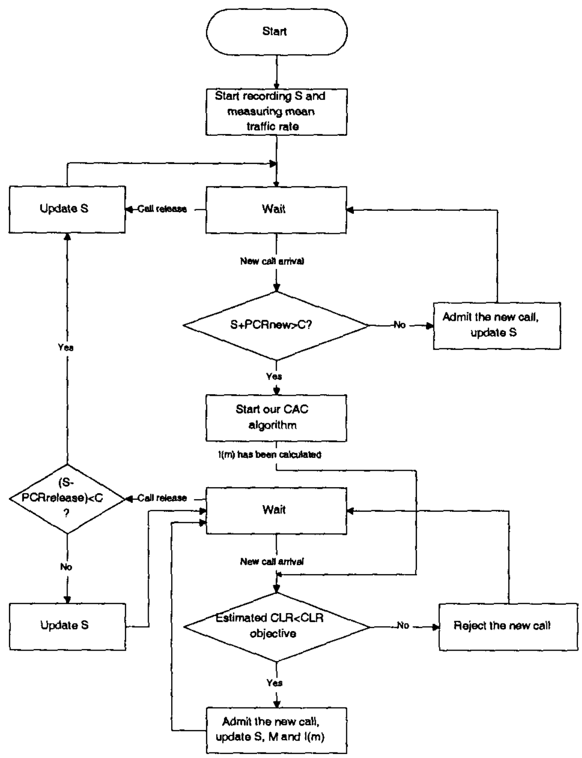


Figure 2. Flowchart of CAC algorithm

4 Simulation Study

Connection admission control algorithms using on-line measurements can only be verified through experiments on either real networks or simulations. In this section we present the simulation results of the proposed CAC algorithm. The aim of our simulation experiments is to evaluate the performance of our approach with respect to resource utilization and the effectiveness of our CAC algorithm in terms of its ability to guarantee the QoS constraints required by the connections.

4.1 Simulation Model

In our simulation we assume that traffic sources follow the exponential ON-OFF source model. The duration of the ON and OFF periods are independent and exponentially distributed with means β and γ respectively. During each ON period an exponentially distributed random number of cells, with mean L, are generated at peak cell rate. During off periods no cells are generated. We define the burstiness

Table 1. Parameter settings for simulation

		$\lambda(s^{-1})$	PCR(Mb/s)	B	L(cells)
1	Type 1	10	1	2	10
	Type 2	2	5	10	20
2	Type 1	10	1	2	10
	Type 2	5	2	10	20
	Type 3	2.5	5	10	30
3	Type 1	10	1	2	10
	Type 2	2	5	10	30
	Type 3	1	10	20	50

(B) of a traffic source as:

$$B = \frac{PCR}{AVG} = \frac{\beta + \gamma}{\beta} \tag{23}$$

Furthermore, the following parameters are used for our simulation: cell loss ratio objective is set to 10^{-4}, the link capacity is set to $150Mb/s$, the measurement interval is 0.05 second, and the safety margin is chosen to be 0.5, the switching speed of the ATM switch is set to be infinity, hence every incoming cell is buffered in the output buffer; and the output buffer size is set to 100 cells. The simulation measures VP utilization and cell loss ratio. These results are presented and discussed later.

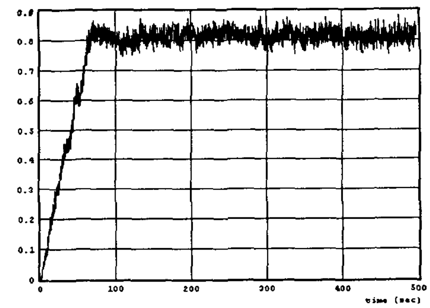
4.2 Call Model

We study the scenarios in which calls of a particular traffic type arrive according to an exponential distribution with the mean λ calls per second. Since CAC schemes are expected to perform worse with regards to QoS guarantees under high call arrival rates, high call arrival rates are chosen for the simulation to establish the robustness of our CAC algorithm. Burstiness and mean burst length L of each traffic type increase with the increase of the peak cell rate of calls of that traffic type. Each call has a call duration which is exponentially distributed with a mean of 100 seconds.

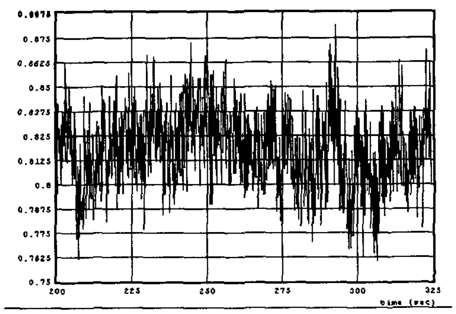
4.3 Simulation Result

Extensive simulations have been carried out and and the results of three scenarios are presented in this section. Two traffic types are multiplexed in the VP in scenario 1 and three traffic types are multiplexed in the VP in scenarios 2 and 3. Traffic rate unit for the three scenarios is set at $1Mb/s$. Parameters of the three scenarios are listed in Table 1.

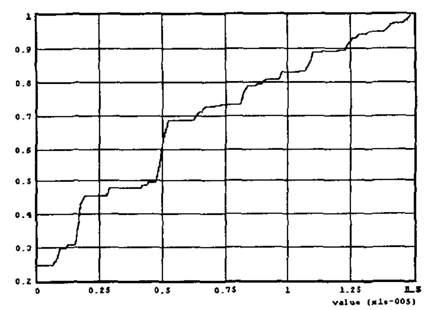
Simulation results for the three scenarios are shown in Fig. 3 ~ Fig. 5. For each scenario, results are presented for the achieved utilization, the cumulative distribution function (cdf) of the observed cell loss ratio, and the number of



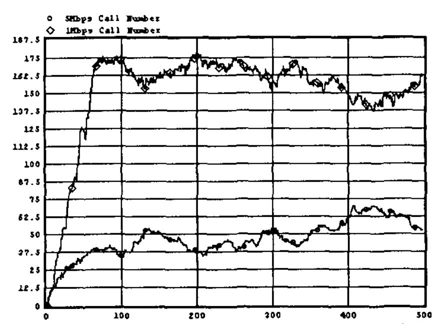
(a) Utilization



(b) Utilization over 200s~325s (for a closer view)

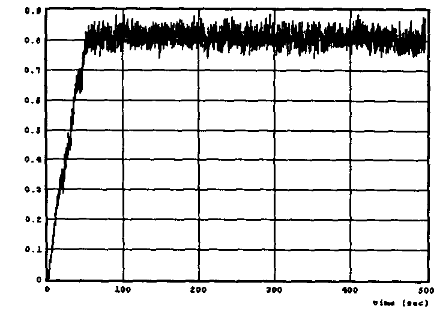


(c) CDF of observed CLR

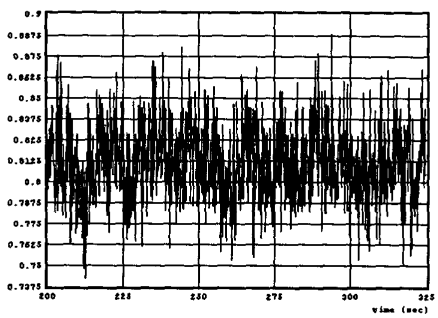


(d) Multiplexed call number

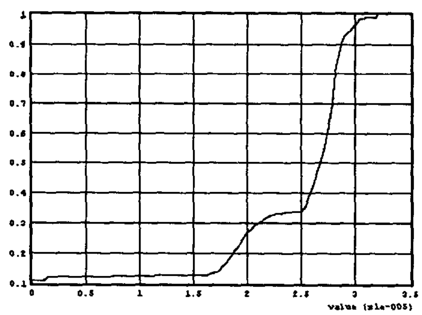
Figure 3. Simulation results for scenario 1



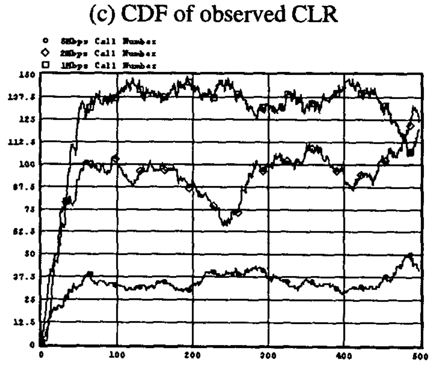
(a) Utilization



(b) Utilization over 200s~325s (for a closer view)

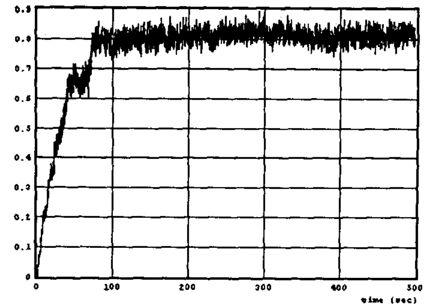


(c) CDF of observed CLR

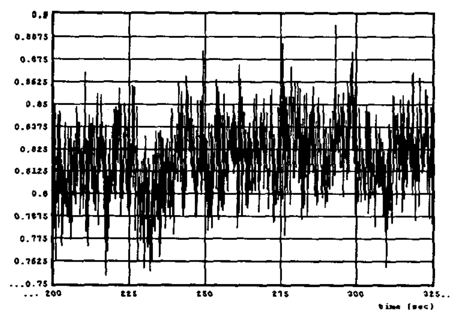


(d) Multiplexed call number

Figure 4. Simulation results for scenario 2

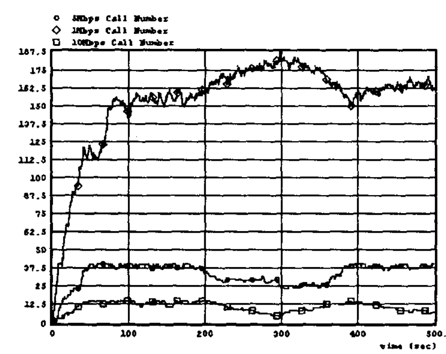


(a) Utilization



(b) Utilization over 200s~325s (for a closer view)

(c) CDF of observed CLR



(d) Multiplexed call number

Figure 5. Simulation results for scenario 3

calls of each traffic type multiplexed in the VP. Simulation results show that the proposed algorithm achieves high VP utilization whilst meeting the stringent QoS requirements of calls multiplexed in the VP. Define the multiplexing gain as the ratio of VP utilization achieved by our CAC algorithm to what would be achieved by peak cell rate CAC algorithm in the same environments. An average multiplexing gain of 2.8 is achieved in scenario 1, an average multiplexing gain of 3.4 is achieved in scenario 2 and an average multiplexing gain of 3.0 is achieved in scenario 3. Fig. 3 shows the simulation results for the scenario when two traffic types are multiplexed together and Fig. 4 shows the simulation results for the scenario when three traffic types are multiplexed together. Binomial distribution is usually used to model homogeneous traffic and our simulations indicate that it is also a good approximation for the traffic distribution of heterogeneous traffic when the maximum peak cell rate of the heterogeneous calls in the VP is not large, say, not larger than 5 percent of the VP capacity.

In scenario 3, a traffic type with peak cell rate of $10Mb/s$ is multiplexed in the VP. It shows the impact of large bandwidth calls on the performance of our CAC algorithm. As a rule of thumb, when the peak cell rate of a traffic type approaches 0.1 of the VP capacity, no significant multiplexing gain can be achieved for that traffic type. Moreover, many large bandwidth calls will also cause the aggregate traffic distribution to significantly deviate from Gaussian distribution. However, Fig. 5 shows that the proposed algorithm still achieves high VP utilization and is robust when the number of large bandwidth calls is small so that these calls do not dominate the aggregate traffic distribution.

5 Concluding Remarks

This paper proposes a CAC algorithm based on on-line measurements and traffic descriptors supplied by sources. Extensive simulations show that the proposed CAC scheme is both efficient and robust. Computation complexity of the proposed algorithm is solely determined by the capacity quantization unit we choose and the capacity of the VP, and it suits real time requirements of ATM connection admission control. Considering the difficulty for sources to tightly characterize their traffic parameters such as mean cell rate and maximum burst size in traffic descriptor-based CAC schemes, and the difficulty in precisely measuring and estimating instantaneous VP traffic rate and high order moments of traffic rate in measurement-based approaches, we only use peak cell rate from traffic descriptor and mean cell rate from traffic measurements which can be easily and reliably derived.

The aggregate traffic distribution of heterogeneous traffic types is approximated using binomial distribution. It is a novel approach and proves to be an effective and reliable method. Aggregate traffic distribution of heterogeneous traffic types can be better modeled by other more complex distribution functions. However, the complexity in CAC design not only lies in how well we can model the traffic, but also in whether the sources or the network can provide enough accurate parameters for the model. Some simplification and approximations in traffic modeling will actually make a CAC scheme performs better and more robust than those complex, well modeled ones based on many unreliable parameters.

A major feature of our CAC scheme is its simplicity. Such simplicity is, we believe, essential for practical implementation of CAC schemes.

References

[1] Chia-Lin Hwang, San-qi Li, "On Input State Space Reduction and Buffer Noneffective Region" *INFOCOM*, pp. 1018-1028, 1994.

[2] H.Zhang, D. Ferrari, "Improving utilization for deterministic service in multimedia communication" *Proceedings of the International Conference on Multimedia Computing and Systems*, 1994.

[3] Sugih Jamin, Peter B. Danzig, Scott J. Shenker and Lixia Zhang, "A Measurement-based admission control algorithm for integrated service packet networks", *IEEE/ACM Transactions on Networking*, vol. 5, pp. 56-70, 1997.

[4] R. R. John, T. Lewis, Fergal Toomey, Brian McGurk, Simon Crosby, Ian Lesie, "Practical Connection Admission Control for ATM Networks based on on-line measurements", *Computer Communications*, vol. 12, pp. 1585-1596, 1998.

[5] Richard J. Gibbens, Frank P. Kelly, Peter B. Key, "A Decision-Theorectic Approach to Call Admission Control in ATM Networks", *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1101-1114, 1995.

[6] J. S. Sukyong Lee, "A Measurement-based Admission Control algorithm using variable-sized window in ATM Networks", *Computer Communications*, vol. 21, pp. 171-178, 1997.

[7] Tsern-Huei Lee, Kuen-Chu Lai and Shii-Tyng Duann, "Design of a Real-Time Call Admission Controller for ATM Networks" *IEEE/ACM Transactions on Networking*, vol. 4, pp.758-765, 1996.

[8] Matthias Grossglauser and David N. C. Tse "A Framework for Robust Measurement-based Admission Control" *IEEE/ACM Transactions on Networking* vol. 7, pp. 293-309, 1999.