Connection Admission Control - Closing the Loop

Guoqiang Mao
School of Electrical and Information Engineering
The University of Sydney
Daryoush Habibi
School of Engineering and Mathematics
Edith Cowan University

Abstract—**To date, many Connection Admission Control (CAC) schemes have been proposed. These schemes can be generally classified into two categories: measurement-based CAC and traffic descriptor-based CAC. Both categories adopt an open-loop architecture where traffic parameters are obtained from either on-line measurements or user-specifications and then input into a chosen traffic and network model to make connection admission decisions. No performance feedback is provided to the CAC scheme. However, no traffic model can be claimed to be accurate for all traffic sources (e.g. voice traffic, data traffic, video traffic) under all network scenarios (e.g. heavy traffic scenario or light traffic scenario). Modelling errors and traffic parameter errors are inevitable. An open-loop CAC scheme lacks the ability to account for modelling errors and adapt to changing network conditions to achieve an optimum performance. In this paper, we proposed a closed-loop architecture for connection admission control. The implementation of the closed-loop architecture in designing a closed-loop CAC scheme is illustrated. Simulation results indicate that the closed-loop architecture is able to overcome the inherent drawbacks of an open-loop CAC scheme to achieve better performance.**

I. Introduction

To date, many Connection Admission Control (CAC) schemes have been proposed. A comprehensive overview of the proposed algorithms can be found in [1], [2], [3], [4]. These schemes can be generally classified into two categories: measurement-based CAC and traffic descriptor-based CAC. Traffic descriptor-based CAC uses the a priori characterization provided by traffic sources at connection setup phase to compute the worst-case behavior of all existing connections in addition to the incoming one. Performance of traffic descriptor-based CAC relies on the accuracy of parameters provided by sources. However, it is difficult for the user to accurately characterize such parameters as mean traffic rate and maximum burst size. Moreover, in order to provide a tight Quality-of-Service (QoS) guarantee, traffic descriptor-based CAC has to allocate bandwidth according to the worst case behavior of connections. Network utilization under this scheme is acceptable when traffic flows are smooth; when traffic flows are bursty, however, it will inevitably result in low utilization [5]. Measurement-based CAC uses the *a priori* traffic characterizations only for the incoming connection and uses measurements to characterize existing connections. Therefore, network utilization does not suffer significantly if the traffic descriptions are not tight. However, because source behavior may be non-stationary, it is difficult for measurement-based CAC to obtain accurate online measurements. Measurement-based CAC can only deliver significant gain in utilization when there is a high degree of statistical multiplexing.

The proposed CAC schemes, whether measurement-based or traffic-descriptor based, have a common feature that they are all open-loop CAC schemes in the sense that there is no performance feedback in the system to allow the CAC scheme to adjust its parameters to achieve the optimum performance. They can be generally described by the two open-loop architectures shown in Fig. 1.

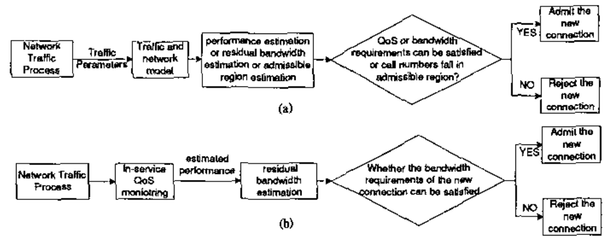

Fig. 1. Architecture of open-loop CAC schemes

Most CAC schemes adopt the architecture of Fig. 1.(a). They obtain traffic parameters from either online measurements or traffic sources. These parameters are then input into the chosen traffic and network model to carry out performance analysis. Based on the performance analysis of traffic sources in the network model, either an estimate of QoS parameters when the new connection is admitted, or an estimate of the residual bandwidth unused by existing connections, or an admissible region, is obtained. If the QoS parameters when the new connection is admitted satisfy QoS objectives, or the residual bandwidth is greater than the bandwidth requirement of the new connection, or the number of connections falls within the admissible region, the new connection is admitted. Otherwise, the new connection is rejected.

There are several other measurement-based CACs based on an in-service QoS monitoring algorithm. They adopt the architecture of Fig. 1.(b). These CAC schemes obtain an estimate of the QoS parameters of existing connections through an in-service QoS monitoring and estimation algorithm. These QoS parameters are then used to estimate the residual bandwidth. If the residual bandwidth is greater than the bandwidth requirement of the new connection, the new connection is admitted. Otherwise, the connection is rejected.

There are inherent drawbacks in these open-loop architectures. The performance of a CAC scheme based on an open-loop architecture relies on accurate traffic and network models, accurate model parameters, and accurate loss performance analysis. As a quick review of the existing literature indicates, no traffic model can be claimed to be accurate for all traffic sources

0-7803-7788-5/03/$17.00 ©2003 IEEE.

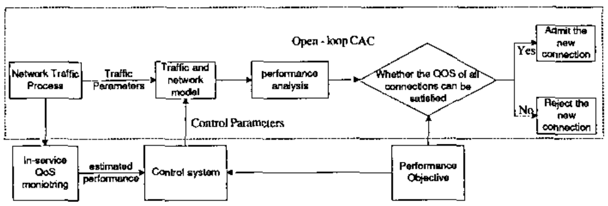

Fig. 2. Architecture of closed-loop CAC schemes

(e.g. voice traffic, data traffic, video traffic) under all network scenarios (e.g. heavy traffic scenario or light traffic scenario). Modelling errors, traffic parameter errors and errors in performance analysis are inevitable. An open-loop CAC scheme lacks the ability to account for these errors and adapt to changing network conditions to achieve the optimum performance.

In this paper, we propose a closed-loop architecture to overcome the inherent drawbacks of an open-loop CAC scheme. The rest of the paper is organized as follows. In section II, we present a closed-loop architecture for CAC scheme; the implementation of the closed-loop architecture in designing a closed-loop CAC scheme is introduced in section III; simulation study is performed in section IV; and finally some conclusions and further research are given in section V.

II. Closed-loop Architecture for CAC

The proposed architecture is shown in Fig. 2. The upper half of the figure is an open-loop CAC algorithm. An in-service QoS monitoring and estimation (ISME) unit is introduced in the system. The ISME unit online monitors and estimates the network performance achieved by the CAC scheme. This information is then used by a control system to control the performance of CAC algorithm. If CAC algorithm is too tight, the control parameters in the CAC algorithm are adjusted such that more resources are allocated to existing connections and vice versa. With a little bit abuse of terminology, we refer to it as a closed-loop architecture in the sense that performance feedback is provided in the system, through a control system, to allow the CAC algorithm to adjust its parameters to achieve the optimum performance.

Due to the real-time requirement of connection admission decisions, the open-loop CAC algorithm in the closed-loop architecture should be a CAC scheme that can function independently without waiting for samples of ISME information. Moreover, performance parameters monitored by the ISME unit are statistical parameters (e.g. cell loss ratio (CLR), cell transfer delay (CTD) and cell delay variation (CDV)) and measurement errors are also inevitable in ISME. Therefore, one or two samples of ISME information are insufficient to make decisions on whether the CAC scheme is tight or not. Thus the CAC scheme should not respond to one or two samples of ISME information too quickly. Instead, control system should change the control parameters gradually through small adjustments such that the CAC scheme converge to the optimum performance asymptotically.

This requirement is also based on the assumption that when the CAC scheme has reached a stable state, the corresponding values of the control parameters are indications of the network states (e.g. heavy traffic, light traffic, degree of self-similarity, etc.) which usually vary slowly with time. Therefore, a slowly-responding CAC scheme (i.e. with response time on the order of seconds) is enough to track network states and adapt to changing network conditions to achieve an optimum performance.

III. Implementation of the Closed-loop Architecture

In this section, we shall illustrate the implementation of the closed-loop architecture in connection admission control. The open-loop CAC scheme presented in [6] is chosen as an example to implement the closed-loop architecture. The CAC scheme in [6] is cited here for completeness.

A. Connection admission control scheme

Exponential on-off source model is used in [6]. A traffic source X_k is characterized by its peak rate pcr_k and mean rate mcr_k. Assuming there are n independent traffic sources $X_1, \dots, X_n$ on the link, the traffic rates of the traffic sources are normalized by $u = max\{pcr_1, pcr_2, ..., pcr_n\}$. A cell loss rate function (clrf) is computed from the n traffic sources.

$$F(k) = \begin{cases} m \times p & k = 0 \\ F(k-1) - 1 + \sum_{i=0}^{k-1} f(i) & k \geqslant 1 \end{cases} . \quad (1)$$

where $f(x)$ is a binomial function:

$$f(x) = \begin{cases} \begin{pmatrix} m \\ k \end{pmatrix} p^k (1-p)^{m-k} & x = k \\ 0 & else \end{cases} . \quad (2)$$

and $m = \lceil \sum_{k=1}^{n} PCR_k \rceil$ and PCR_k is the normalized peak rate of source X_k. p is determined by:

$$\widehat{p} = \frac{r}{m} + \alpha \sqrt{\frac{\frac{r}{m}(1 - \frac{r}{m})}{m}}, \quad (3)$$

where r is the mean rate of the aggregate traffic obtained from online measurements, and α is called the safety margin which is determined from the measurement period. If a new connection X_{n+1} with normalized peak rate PCR_{n+1} and mean rate MCR_{n+1} arrives, the CLR, if the new connection is admitted, is estimated as follows:

$$clr = \frac{(1 - MCR_{n+1})F(C) + MCR_{n+1}F(C-1)}{F(0) + MCR_{n+1}}. \quad (4)$$

where C is the normalized link capacity. If the estimated CLR is less than the CLR objective then the connection is admitted; otherwise the connection is rejected.

In the aforementioned CAC scheme, the safety margin α is a control parameter that controls the performance of the CAC scheme. For a large value of α, the CAC scheme will perform conservatively such that less connections are admitted and performance objective is tightly observed; for a small value of α the CAC scheme will perform aggressively such that more connections are admitted. Fig. 3 shows the variation of the performance parameter CLR with the safety margin, which is obtained from a group of simulations using the same heterogeneous on-off sources but different safety margin parameters. Fig. 3 demonstrates qualitatively that by controlling the safety margin α, the

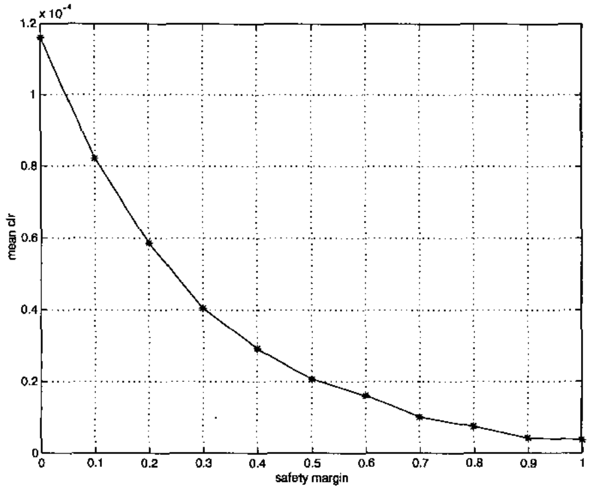

Fig. 3. Controlling effect of the safety margin

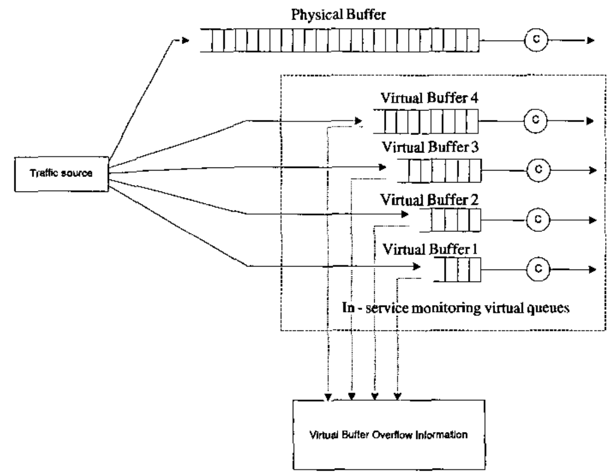

Fig. 4. System model of the CLR estimation algorithm

performance of the CAC scheme can be controlled. However, depending on the traffic sources being used, the exact relationship between CLR and the safety margin α may be different from that shown in the figure. The safety margin α is chosen as the control parameter in the closed-loop architecture.

It is worth noting that in some other measurement-based CAC schemes, there are similar parameters which can be used to control the performance of connection admission control schemes. Lee summarizes some features of those measurement-based CAC schemes [7]. Even for those schemes which do no naturally have such control parameters, they can be easily modified to embed the control parameter. For example, the mean rate m can be modified as $(1+\alpha)m$ or $m+\alpha\sigma$, where σ is the standard deviation of traffic rate. Therefore, the proposed closed-loop architecture is of general significance for CAC scheme design and can be implemented in a variety of CAC schemes.

B. In-service QoS monitoring and estimation scheme

An ISME scheme is designed to monitor the CLR performance online. One potential problem for ISME is that CLR is often specified in terms of the probability of occurrence of certain rare events. For example, in ATM networks CLR is often specified to be as small as 10^{-9}. Monitoring using direct statistical methods is impractical for such small CLR. As an example, in an OC3 link with a link utilization of 0.5 and a CLR of 10^{-9}, at least 10 billion ATM cells have to be monitored before any statistically meaningful information can be obtained (It is assumed that at least 10 cell loss samples have to be observed to obtain any statistically meaningful information). This will take 15 hours. The statistical information obtained after such long monitoring period may be obsolete and the network management system's reaction may be too late. Even in a network with a moderate CLR, an ISME scheme with a much reduced monitoring period than the direct monitoring method is still desirable to provide faster system response.

To address this problem, an ISME scheme using virtual buffer techniques is proposed [8]. The basic idea is to use the cell loss ratios observed from several virtual buffers with much smaller buffer sizes than that of the real buffer to estimate CLR of the real system. Since virtual buffers have much smaller buffer sizes, the cell loss ratios in the virtual buffers are much larger than that of the real buffer. In order to obtain statistical meaningful observations of the cell loss ratios of these virtual buffers, fewer cells need to be observed compared to those required for a large real buffer. Hence the observation of the cell loss ratios of the virtual buffers requires much less monitoring period. Then based on the asymptotic relationship between CLR and buffer size for both Markovian traffic [9], [10], [11] and self-similar traffic [12], [13], [14], [15], the cell loss ratios observed in these virtual buffers are used to obtain an estimate of CLR in the real buffer. Therefore much less monitoring period is required in the proposed ISME algorithm to obtain an estimate of CLR in the real buffer than that using direct monitoring method. It is worth noting that since the size of the virtual buffer is smaller than that of the real buffer, a virtual buffer can be simply implemented as a counter. The value of the counter varies between a threshold, which is the size of the virtual buffer, and zero. The counter is incremented by one with each cell arrival. When the threshold is reached, further cell arrivals are considered as lost and the counter remains constant at the threshold. At the same time, the counter is decremented by one with each cell departure in the real buffer. Therefore the implementation of virtual buffers will not add too much burden onto the real system.

Fig. 4 shows the system model using virtual buffers. Four virtual buffers are employed in the algorithm. Denote by B_1, B_2, B_3, B_4 the sizes of the four virtual buffers. B_1, B_2, B_3, B_4 are chosen such that:

$$B_1 < B_2 < B_3 < B_4 << B,$$

where B is the size of the real buffer. Denote by clr_t^1, clr_t^2, clr_t^3 and clr_t^4 the cell loss ratios observed in the four virtual buffers 1, 2, 3 and 4 respectively at discrete time t. These cell loss ratios of the virtual buffers are used to obtain an estimate of the CLR in the real buffer, which is denoted by $\widehat{clr}_t$.

Simulation results using both exponential on-off source

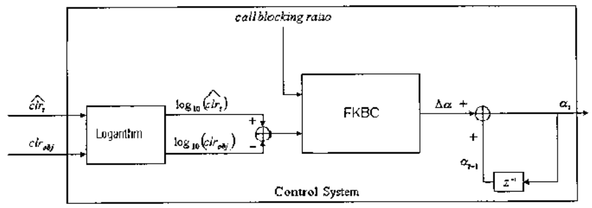

Fig. 5. Architecture of the control system

model and real self-similar variable-bit-rate (VBR) video sources showed that the proposed ISME scheme generally gives a CLR estimate that is accurate within one order of magnitude of the true value. The required monitoring period is reduced from several hours using direct monitoring method to several seconds.

C. Control system design

The control system is designed to control the safety margin α using the CLR estimation from ISME. A fuzzy-knowledge-based-controller (FKBC) with a center average defuzzifier, product-inference rule and and singleton fuzzifier [16] is used. Fig. 5 shows the architecture of the control system used in the closed-loop CAC scheme.

The CLR estimate made by the ISME unit at discrete time t, denoted by $\widehat{clr_t}$, and the CLR objective, denoted by clr_{obj}, are used as inputs to the control system. The output of the control system is the safety margin α, which can be used to control the connection admission and bandwidth allocation in the CAC scheme.

Based on the observation that CLR of the system is lower than the CLR objective does not necessary imply that the CAC scheme is conservative, it can also be caused by low call arrival rate, the call blocking ratio is also taken into account in the control system design. Here *call blocking ratio* is defined as the ratio of the number of calls rejected by the CAC scheme to the number of connection requests during the same CLR monitoring period when CLR estimation $\widehat{clr_t}$ is made. Only when the CLR of the system is lower than the CLR objective while a non-zero call blocking ratio is observed, it means that the CAC scheme is conservative. Thus the safety margin parameter of the CAC scheme needs to be changed such that less bandwidth is allocated to existing connections. Estimation errors of ISME unit are also taken into account in the design of the FKBC.

IV. Simulation Study

For comparison purpose, the same simulation scenarios as those in [6] are employed. Simulation parameters are chosen to be the same as those in [6] except that a large buffer size of $1,000$ ATM cells is chosen as opposed to 20 cells in [6]. This large buffer size is used to establish the ability of the closed-loop CAC scheme to overcome the restriction of model error. The open-loop CAC scheme reported in [6] is based on a bufferless fluid flow model. In a network with a large buffer size it fails to efficiently utilize the large buffer size due to the restriction of the bufferless fluid flow model. The CAC scheme reported in [6] achieves the same bandwidth utilization in a network with a large buffer size as that in a network with a small buffer size. However buffer size and bandwidth are exchangeable network resources under certain conditions. Therefore a higher bandwidth utilization should be achieved in a network with a large buffer size. We expect that the closed-loop CAC scheme is able to overcome the restrictions of model error and achieve higher bandwidth utilization in a network with a large buffer size. Therefore, a large buffer size of $1,000$ ATM cells is chosen in this section for comparison.

A. Simulation using exponential on-off sources

First, simulation using exponential on-off source model is performed. The duration of the ON and OFF periods of the on-off source are independently and exponentially distributed with means β and γ respectively. Three types of traffic sources are multiplexed on the link. The connection arrival process of each type of traffic source is a Poisson process with a mean of λ calls per second. The connection holding time for all traffic types is exponentially distributed with a mean of 100 seconds. The CLR objective is set to be 10^{-4}. The link capacity is set to be $10Mb/s$.

The connection arrival rates in this scenario, referred to as saturation scenario, are chosen to be very high. This scenario is used to establish the performance of the CAC scheme with regards to QoS guarantees, because if calls are offered at a very high rate, the rate at which calls are admitted in error becomes very large too [17].

The parameters of the three traffic types of the saturation scenario are listed in Table I.

TABLE I

Parameters of the three traffic types in the saturation scenario

	$\lambda\,(s^{-1})$	$pcr\,(kb/s)$	$\beta\,(s)$	$\gamma\,(s)$
type 1	10	100	0.424	3.816
type 2	50	50	0.424	1.696
type 3	100	10	0.424	0.424

The simulation was run for $20,000$ seconds. Fig. 6 shows the observed CLR, and Fig. 7 shows the variations of safety margin parameter in the simulation.

Fig. 7 shows that under the control of the FKBC, the safety margin α starts from a conservative guess of $\alpha = 0$ and arrives at a stable value of around -0.40. With the decrease of safety margin, the utilization achieved by the CAC scheme rises from an initial value of around 0.79 to 0.88. During this process, an increase in CLR is also observed. However the CLR is successfully controlled below the CLR objective.

This simulation builds the validity of the proposed closed-loop CAC scheme. Compared with the CAC scheme without closed-loop control in [6], network utilization increases from 0.76 to 0.88. Therefore the closed-loop architecture is able to significantly increase network utilization. This increase in utilization is achieved by utilizing the large buffer in the network efficiently. Therefore, this simulation demonstrates the ability of the closed-loop CAC scheme to overcome the restriction of model error, and achieves better network resources utilization.

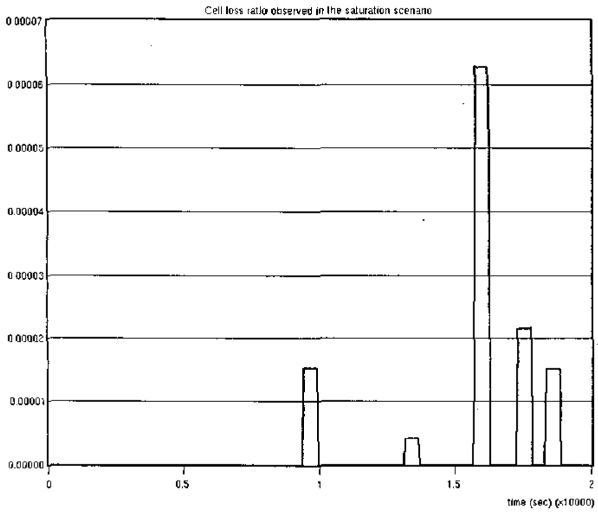

Fig. 6. CLR observed in the saturation scenario

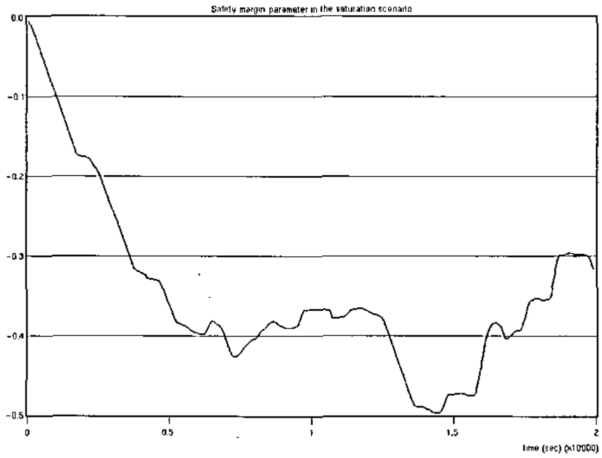

Fig. 7. Safety margin parameter in the saturation scenario

B. Simulation using real VBR video sources

In this section, we will further investigate the performance of closed-loop CAC scheme using variable bit rate video sources. Eight Motion-JPEG (M-JPEG) encoded movies are used in the simulation. The statistics of the M-JPEG encoded movies are shown in Table II. The frame rate of the M-JPEG encoded movies is $30 frames/s$. OC3 link is used in the simulation.

The simulation was run for $10,000s$. Fig. 8 shows the CLR observed in this scenario. Fig. 9 shows the variations of the safety margin parameter in the simulation.

When the open-loop CAC scheme in [6] based on bufferless fluid flow model and the on-off traffic source model is applied to real VBR video traffic sources, in addition to the error caused by the bufferless fluid flow model, the on-off traffic source model error also affects the performance of the CAC scheme. Therefore, the open-loop CAC scheme achieves much lower utilization when applied to real VBR video sources. When a closed-loop architecture is applied to the CAC scheme, a utilization of about 0.87 is achieved. This utilization is significantly higher than the utilization of 0.65 achieved by the open-loop CAC scheme under the same conditions, and is close to that achieved by the closed-loop CAC scheme in the saturation scenario where on-off sources are used. This simulation demonstrated that the closed-loop CAC scheme is able to overcome the restriction of traffic model errors to achieve higher network resources utilization. It is noticed that although the VBR video sources used in the simulation are self-similar [6], the closed-loop measurement-based CAC scheme successfully controls the CLR below the performance objective.

TABLE II

TRAFFIC RATE OF THE M-JPEG ENCODED MOVIES ($bytes/frame$)

Type	Name	peak rate	mean rate
1	Sleepless in Seattle	16617	9477.6
2	Crocodile Dundee	19439	10772.9
3	Home Alone, II	22009	11382.8
4	Jurassic Park	23883	11363.0
5	Rookie of the Year	27877	12434.9
6	Speed	29385	12374.4
7	Hot Shots, Part Duex	29933	12766.1
8	Beauty and the Beast	30367	12661.5

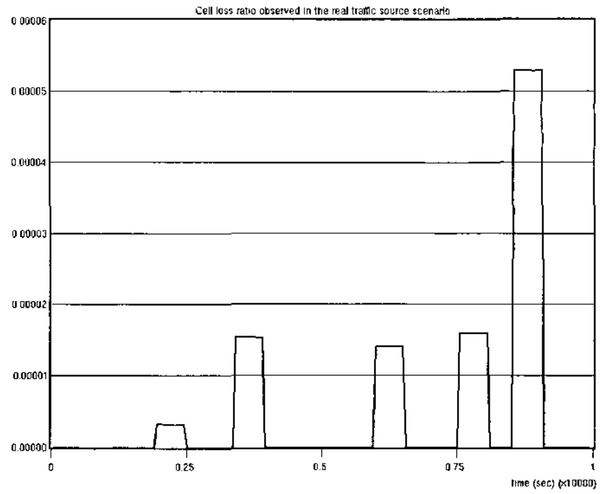

Fig. 8. CLR observed in the real traffic source scenario

V. CONCLUSION AND FURTHER RESEARCH

In this paper, we presented a closed-loop architecture for connection admission control. The implementation of the closed-loop architecture in designing a closed-loop CAC scheme was illustrated. Simulation was performed on the performance on the closed-loop CAC scheme. Simulation results showed that the closed-loop architecture is able to overcome the inherent drawbacks of an open-loop CAC scheme, in which performance is limited by the accuracy of traffic and network models as well as traffic parameters. Moreover, the closed-loop CAC scheme is able to adapt to real traffic conditions to achieve a better performance. In a real network, heterogeneity, complexity and dy-

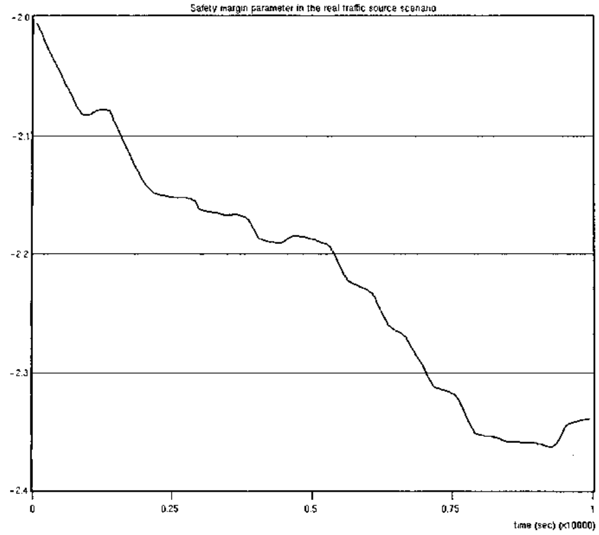

Fig. 9. Safety margin in the real traffic source scenario

namic nature of network traffic make accurate traffic modelling and obtaining accurate traffic parameters very difficult. The salient feature of a closed-loop CAC scheme that it is able to overcome these model errors, traffic parameter errors and adapt to changing network traffic conditions becomes very attractive.

Currently, the closed-loop architecture is only applied to CAC scheme where CLR is the performance objective. Further work is being done on online monitoring of other QoS parameters, i.e. traffic delay and delay variations, such that more performance objectives can be satisfied by the closed-loop CAC scheme. Moreover, simulation revealed interesting results that the closed-loop CAC scheme based on exponential on-off source models can meet the CLR QoS performance objective even under self-similar traffic conditions. Further research is being carried out on the performance of the closed-loop CAC scheme under self-similar traffic conditions.

REFERENCES

[1] G. Mao, *Statistical Multiplexing and Connection Admission Control*, Phd thesis, Edith Cowan University, 2001.

[2] Edward W. Knightly and Ness B. Shroff, "Admission control for statistical qos: Theory and practice," *IEEE Network*, vol. 13, no. 2, pp. 20–29, 1999.

[3] Harry G. Perros and Khaled M. Elsayed, "Call admission control schemes: A review," *IEEE Communications Magazine*, vol. 34, no. 11, pp. 82–91, 1996.

[4] Kohei Shiomoto, Naoaki Yamanaka, and Tatsuro Takahashi, "Overview of measurement-based connection admission control methods in atm networks," *IEEE Communications Surveys*, vol. First Quarter, pp. 2–13, 1999.

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

[6] G. Mao and D. Habibi, "Loss performance analysis for heterogeneous on-off sources with application to connection admission control," *IEEE/ACM Transactions on Networking*, vol. 10, no. 1, pp. 125–138, 2002.

[7] Breslau Lee, Jamin Sugih, and Shenker Scott, "Comments on the performance of measurement-based admission control," in *IEEE INFOCOM'2000*, 2000.

[8] G. Mao and D. Habibi, "Quality of service monitoring in multimedia network," in *SPIE Conference Multimedia Computing and Networking 2003*, Santa Clara, CA, USA, 2003, pp. 5019–7–1 – 5019–7–10.

[9] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of data-handling systems with multiple sources," *Bell System Technical Journal*, vol. 61, no. 8, pp. 1871–1893, 1982.

[10] Anwar I. Elwalid and Debasis Mitra, "Effective bandwidth of general markovian traffic sources and admission control of high speed networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 329–343, 1993.

[11] Anwar Elwalid, Daniel Heyman, T. V. Lakshman, and Debasis Mitra, "Fundamental bounds and approximations for atm multiplexers with applications to video teleconferencing," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1004–1016, 1995.

[12] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[13] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1566–1579, 1995.

[14] N. Likhanov, B. Tsybakov, and N.D. Georganas, "Analysis of an atm buffer with self-similar ("fractal") input traffic," in *IEEE INFOCOM 1995*, Boston, MA , USA, 1995, vol. 3, pp. 985–992.

[15] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, 1997.

[16] Li-Xin Wang, *Adaptive Fuzzy Systems and Control - design and stability analysis*, PTR Prentice Hall, Englewood Cliffs, New Jersey 07632, 1994.

[17] Richard J. Gibbens, Frank P. kelly, and Peter B. Key, "A decision-theoretic approach to call admission control in atm networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1101–1114, 1995.