SBrT 2000 Simpósio Brasileiro de Telecomunicações, 3 a 6 de Setembro, 2000, Gramado-RS

CONNECTION ADMISSION CONTROL MODELLING FOR ATM NETWORKS

TANIA R. T. FUDOLI$^{\diamond}$, IVANIL S. BONATTI*
$^{\diamond}$CPqD - R&D Telecom Foundation P.O. Box 1579
13088-061 - Campinas-SP-e-mail: tania@cpqd.com.br,
*State University of Campinas - P.O. Box 6101
13081-970 Campinas-SP - e-mail: ivanil@dt.fee.unicamp.br

ABSTRACT

The Connection Admission Control(CAC) is part of ATM traffic control and consists of a framework that optimizes network usage while ensuring the desired QoS for the services. The CAC approach we propose for ATM networks is based on a effective bandwidth allocation strategy. This paper addresses the effective bandwidth computation based on a probabilistic model for cell level QoS and call level QoS. Many papers concerning this subject usually consider only cell QoS parameter. We propose a call-admission criteria derived from load-sharing and Kaufman approximation that allows to simplify the blocking probability analysis of a multirate link.

1. INTRODUCTION

Asynchronous transfer mode networks are designed to support a variety of applications with different traffic requirements. When a network supports applications with delay and loss requirements, the network must decide whether there are enough resources in the network to admit a new call request without affecting the commitments made to network applications already admitted in the network.

A set of five service categories has been specified by the ATM Forum[1]: *Constant Bit Rate* (CBR); *non-real-time Variable Bit Rate* (nrt-VBR); *real-time Variable Bit Rate* (rt-VBR); *Available Bit Rate* (ABR) and *Unspecified Bit Rate* (UBR).

The CBR service category contains primarily the real-time applications with no compression of the source traffic and the rt-VBR category is intended for applications which require tightly constrained delay and delay variation, as would be appropriate for voice and video applications with variable rate.

The nrt-VBR service category is intended for applications which have bursty traffic characteristics and have no restrictions to cell delay occurrence, while the ABR category is used in applications with the ability to alter their information transfer rate as a function of the network flow state. In UBR service category the cells are transmitted whenever there is bandwidth available, and discarded in case of congestion. The UBR service does not specify traffic related service guarantee.

The ITU-T has been defined some similar ATM transfer capabilities [5].

The service categories are characterized by traffic descriptor parameters, as *Peak Cell Rate* (PCR) for the CBR service category, and by *Quality of Service* (QoS) parameters, like *Cell Loss Ratio* (CLR) in the nrt-VBR service category.

The Connection Admission Control (CAC) is part of the ATM traffic control and consists of a framework that optimizes network usage while ensuring the desired QoS for the services.

Resource allocation by CAC deals with determining the amount of bandwidth required by a connection for the network to provide the required QoS.

In a simple call admission control procedure if the peak bandwidth requirement of each connection is allocated (even for the VBR service categories), then the ATM operates like a STM (Synchronous Transfer Mode) system and no statistical advantages are obtained from the bursty nature of the broadband integrated traffic. On the other hand, if an average bandwidth requirement of each connection is allocated (the most efficient use of the network resources) no quality of service guarantee would be possible.

It can be possible to use resources more efficiently while still satisfying quality of service requirements by overbooking link bandwidth in statistical multiplexing schemes.

The main purpose of this paper is to investigate the best way to implement the statistical multiplexing in an ATM network in order to maximize the link bandwidth efficiency. The major constraint to be satisfied is to guarantee cell and call QoS requirements for different types of traffic even during periods of network congestion. The cell level QoS parameters considered are the cell loss, cell delay and cell delay variations. The call/connection blocking probability is considered as a call level QoS parameter taken into account in CAC procedures. The CAC approach proposed for ATM networks [4] is based on a *effective bandwidth* allocation strategy. A circuit-switched style call acceptance algorithm has been implemented, since each of the traffic sources sharing the network (the transmission links) can be considered as being served by its respective effective bandwidth.

In the proposed approach the effective bandwidth computation is based on a probabilistic model for cell level QoS (cell loss and a worst-case model for cell delay and cell delay variation) and call level QoS (call blocking probability). Many papers concerning this subject usually consider only one cell QoS parameter. A few papers consider all the cell level QoS parameters but using a single probabilistic model for their computation [2], [3], [6], [7]. Moreover, the formulation of effective bandwidth in the literature is mainly based on cell level QoS [8], [9]. We propose a call-admission criteria derived from load-sharing routing and Kaufman approximation that allows to simplify the blocking probability analysis of a multirate link.

2. CELL LEVEL CONTROL

The core of any CAC algorithm is the estimation of bandwidth required by a set of sources so as to respect the QoS requirements of each individual source. We consider the connection admission control problem for a statistical multiplexer system that consists of a link with transmission rate C and a buffer of size b. Our model considers effective bandwidth computation based on acceptable cell loss, cell delay and cell delay variation.

2.1 Cell Loss

The effective bandwidth based on acceptable loss probability of a source depends neither on the number of sources sharing the buffer nor on the model parameters of other types of sources sharing the buffer. The effective bandwidth computation is obtained on the assumption that the probability of finding more than b cells in the queue must be less than a certain given value. Mathematically, for an infinite size queue with service rate C *cps* (cells per second):

$$\Pr\{X > b\} \leq \exp(-b\delta) \tag{1}$$

X being the number of cells in the queue found by an arriving cell.

Supposing that there are k traffic sources multiplexed in a buffer, there exist functions c_i, depending only on source i and parameter δ, such that the grade of service required in (1) is achieved if, and only if,

$$\sum_{i=1}^{k} c_i \leq C \tag{2}$$

where c_i is the effective bandwidth computed as in [7]:

$$c_i = \frac{1}{\delta} \lim_{t \to +\infty} \frac{\ln E\{\exp[N_i(t)\delta]\}}{t} \tag{3}$$

$N_i(t)$ being the number of arrivals in the time interval $(0,t)$ for the source i, $i = 1,\ldots,k$ and $E\{X\}$ is the expected value of the X random variable.

All the possible cell loss ratio that are taken into account in the traffic contracts of standardization committees can be represented through two traffic models: constant rate and ON-OFF Markov fluid sources. The simplest case of effective bandwidth computation considers sources with constant arrival rate λ, where $N(t) = \lambda t$ is deterministic and $c = \lambda$.

A Markov fluid source is similar to a MMPP source [7], differing only on the arrival rate (in the case of Markov fluid sources, the arrival rate is constant in each phase). In the ON-OFF model, the source has two phases: it is either transmitting at rate λ or not transmitting at all. The source stays in each phase during an exponentially distributed random time interval.

Thus, the effective bandwidth for ON-OFF Markov fluid sources is given by:

$$c = \alpha + \sqrt{\alpha^2 + \beta} \tag{4}$$

with

$$\alpha = \frac{1}{2}\left(\lambda - \frac{1}{\delta T_{\text{on}}} - \frac{1}{\delta T_{\text{off}}}\right) \qquad \beta = \frac{\lambda}{\delta T_{\text{off}}} \tag{5}$$

where T_{on} and T_{off} are the mean time intervals spent in ON and OFF states, respectively.

2.2 Cell Delay

In this section the effective bandwidth is computed by means of a worst-case deterministic delay model derived from the work of Boudec [2].

A queue with infinite size buffer, served at rate c, can guarantee a limit Δ for the cell delay, if:

$$c = \max_t \frac{N(t)}{t + \Delta}; \quad t > 0 \tag{6}$$

$N(t)$ being the number of arriving cells in the interval $(0,t)$.

The deterministic effective bandwidth model does not take into account the cell loss; thus, the buffer must have a size such that:

$$b \geq \sum_i c_i \Delta_i \tag{7}$$

where c_i and Δ_i are the effective bandwidth and the maximum delay supported for each of the link connections, respectively.

The effective bandwidth for constant rate sources is given by the transmission rate, as:

$$\max_t \frac{\lambda t}{t+\Delta} \rightarrow \lambda \tag{8}$$

A variable rate source (defined by the parameters: mean rate λ, peak rate λ and maximum burst size L) has its worst-case delay when it is modeled as an ON-OFF source with all bursts having size L.

The burst duration can be derived from the relation:

$$T_{\text{on}} = \frac{L}{\lambda} \tag{9}$$

The interval T_{off} between bursts is derived from:

$$\bar{\lambda} = \frac{T_{\text{on}}}{T_{\text{on}} + T_{\text{off}}} \lambda \tag{10}$$

The effective bandwidth c is always greater than the mean rate λ, guaranteeing that the buffer is empty after each burst.

The buffer is filled obeying the function $(\lambda - c)t$ and is emptied with the rate c. The number of cells in the buffer when a burst finishes is $(\lambda - c)T_{\text{on}}$ and the buffer must be emptied within a maximum time interval Δ, at rate c. Thus, the effective bandwidth is given by:

$$c = \frac{\lambda T_{\text{on}}}{T_{\text{on}} + \Delta} \tag{11}$$

The parameter Δ may represent either the maximum delay admissible in the buffer or the maximum variation allowed for this delay.

The deterministic bandwidth model is conservative, provided that it is obtained from worst-case analysis which considers the collision of multiplexed sources transmitting at peak cell rate.

3. CALL LEVEL CONTROL

In this section, we analyze the general problem of admission control at call level in a network with arbitrary topology and different traffic classes. Consider a network with M nodes and L links.

Let $\mathcal{K}$ be the set of index of all possible node pairs that communicate each other in the network, i.e.,

$$\mathcal{K} = \{1, 2, \cdots, K\} \tag{12}$$

where $k \in \mathcal{K}$ denotes a communicating node pair (p,q) with node p being the origin node and q the destination node.

Associated with every possible communicating node pair is a set of admissible routes (a route is a sequence of links). We represent the route set for k and service class n ($n = 1, 2, \cdots, N$) by $\mathcal{R}_{k,n}$, i.e.,

$$\mathcal{R}_{k,n} = \{1, 2, \cdots, R_{k,n}\} \tag{13}$$

where $r \in \mathcal{R}_{k,n}$ represents the rth admissible route P_r for node pair k, service class n and consists of a sequence of node pair $[(j_0, j_1), (j_1, j_2), \cdots, (j_{d-1}, j_d]$, where j_0, j_d represents the node pair k and j_1, j_2, j_{n-1} the intermediate nodes of the route r. A load sharing routing is considered. The offered traffic $\lambda_{k,n}$ is shared between the possible routes $r \in \mathcal{R}_{k,n}$. Then

$$\lambda_{k,n} = \sum_{r \in \mathcal{R}_{k,n}} \lambda_{r,k,n} \tag{14}$$

where $\lambda_{r,k,n}$ is the Poisson arrival rate on the route r for the class n and node pair k.

For a specific link $j \in \mathcal{R}$ and class n, the Poisson arrival rate is $\lambda_{j,n}$ where

$$\lambda_{j,n} = \sum_{k \in \mathcal{K}} \sum_{r \in \mathcal{R}_{k,n}\ ;\ P_r \supset j} \lambda_{r,k,n} \prod_{l \in P_r\ ;\ l \neq j} (1 - B_{l,n}) \tag{15}$$

where $B_{l,n}$ is the blocking probability of link l for calls of class n.

For one traffic class, with Poisson arrivals and exponential holding-times (with average μ^{-1}), the blocking probabilities of the link j are approximated by

$$B_j = Er(\lambda_j / \mu, C) \tag{16}$$

where $Er(\bullet, \bullet)$ is the Erlang loss function, $1/\mu$ = holding time and link capacity C.

For N traffic classes we have $\lambda_{j,n}$ and $B_{j,n}$ for each class n. In order to simplify the notation, we will denote $B_{j,n}$ simply by B_n. A call-admission policy for a generic link is a rule that specifies, for each state of the link, the classes of call-arrivals that are accepted and those that are rejected in that state. The blocking probability B_n can be obtained by Markov decision process. This formulation for even the single-link has an n-dimensional state vector, defined by the numbers of calls of the n classes in progress on the link and becomes numerically intractable, even for small values of n. In order to reduce the complexity of computations with multi-dimensional Markov chains, we propose to reduce the Markov-chain by using the one-dimensional recursion for link occupancy probabilities derived by Kaufman [10]. With this approximation, the multirate admission-control problem on a link is solved by a single system of linear equations of size $(M+1)$, where M is the number of "trunks" (integer multiples of a basic bandwidth unit) in the link, regardless of number of traffic classes. Consider δ the basic bandwidth unit described by

$$\delta = gcf\{c_1, c_2, \cdots, c_N\} \tag{17}$$

gcf{} = greatest common factor. Let M be maximum basic bandwidth unities available given by

$$M = \lfloor C/\delta \rfloor \tag{18}$$

where $\lfloor x \rfloor$ is the greater integer less than x and

$$\bar{c_n} = c_n/\delta \tag{19}$$

the number of basic bandwidth unities required by the class n.

The state probabilities α (not normalized) is obtained by the following recursive scheme:

$$\alpha_0 = 1 \tag{20}$$

$$\alpha_m = \frac{1}{m}\sum_{n=1}^{N} a_n \alpha_{m-\bar{c_n}} \ ; \ m = 1, \cdots, M \tag{21}$$

$$\alpha_{m<0} = 0 \tag{22}$$

where

$$a_n = (\lambda_n/\mu_n)c_n/\delta \tag{23}$$

Then

$$B_n = \sum_{m=M-\bar{c_n}+1}^{M} q_m \tag{24}$$

where

$$q_m = \alpha_m \left(\sum_{m=0}^{M} \alpha_m\right)^{-1} \tag{25}$$

B_n denote the probability of blocking for calls of class n on the link of capacity C.

4. CONCLUSION

The connection admission control for ATM networks that unifies cell and call level QoS control has been addressed in this paper. The cell level effective bandwidth computation of the link traffic considering cell loss, cell delay and cell delay variation parameters as a single probabilistic model was presented. At the call level, we propose the load sharing method to routing the calls and the Kaufman model to calculate the blocking probability of a multirate link.

5. REFERENCES

[1] ATM Forum, af-tm-0056.000. *Traffic Management Specification, Version 4.0*, April, 1996.

[2] J.-Y. Le Boudec and R. Nagarajan. A CAC algorithm for VBR connections over a VBR trunk. In *Proceedings of 15th ITC*, vol. 1, pages 59–70, Washington, DC, USA, 1997.

[3] G. L. Choudhury, D. M. Lucantoni, and W. Whitt. Squeezing the most out of ATM. *IEEE Transactions on Communications*, 44(2):203–217, February 1996.

[4] ITU-T, Draft Rec. E.73y. *Dimensioning Methods for B-ISDN*, January 6-10, 1997.

[5] ITU-T, Rec. I.371. *Traffic control and congestion control in B-ISDN*, 1995.

[6] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. B. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141–168. Oxford University Press, 1996.

[7] G. Kesidis, J. Walrand, and C.-S. Chang. Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE Transactions on Networking*, 1(4):424–428, August 1993.

[8] D. Mitra, M. Reiman, and J. Wang. Robust Dynamic Admission Control for Unified Cell and Call QoS in Statistical Multiplexers. *IEEE Journal on Selected Areas in Communications*, 16(5):692–707, June 1998.

[9] K. Steenhaut, K. Degieter, W. Brissinck, E. Dirkx. Scheduling and admission control policies: A case study for ATM. *Computer Networks and ISDN Systems*, 29(1997) 539–554.

[10] J. S. Kaufman. Blocking in a shared resource environment *IEEE Transactions on Communications*, 29(10):1474-1481, 1981