A Novel Effective Bandwidth Approach to CAC in ATM networks

M. USHARANII1, K. SAILAJA2

1Associate Professor, 2Research scholar, SPMVV, Tirupati

musha_rohan@yahoo.com, sailaja.k20@gmail.com

ABSTRACT

In ATM networks, connection admission control can be regarded as the most important of the preventive control functions which aim at restraining congestion in the network nodes. In this paper, I presented the CAC scheme which is based on the concept of effective bandwidth. Since this approach takes the statistical multiplexing effect into account, it can lead to higher resource utilisation. The use of effective bandwidth simplifies the CAC procedure by estimating the total bandwidth of the aggregate traffic (including the new connection). In this CAC scheme, the effective bandwidth of the aggregate traffic is calculated directly rather than summing up individual bandwidths, hence overcoming the drawback of conventional methods. Simulation tests show that our approach is simple and results in higher utilisation compared with conventional methods. In this section, I presented a novel effective bandwidth approach to perform real-time CAC.

Keywords : ATM Networks, CAC scheme, Statistical Multiplexing, Novel Effective bandwidth approach, aggregate traffic.

1.INTRODUCTION

One of the important issues in ATM networks is bandwidth management and allocation. The theory of effective bandwidth is an active area of research and is being applied to the design, simulation and analysis of high-speed ATM networks. CAC is defined as a set of actions taken by the network during the call setup or a call renegotiation phase to establish a connection without degrading network performance. At connection setup, a route through the network is selected. Then, the QOS of each affected link is estimated, taking the effect of the new connection into consideration. The connection request is accepted if each link can offer sufficient QOS to all connections. According to [1], admission control methods must address three major issues:

• What parameters are needed to accurately describe the traffic generated by a connection? The parameters that are used to describe a connection must be complete enough to allow the admission control method to accurately predict the effect of the newly admitted connection on network performance. However, the set of parameters should be as small as possible in order to limit the computational complexity, latency and other resources needed for CAC. It is also crucial to provide an accurate characterization of a connection's burstiness.

• How does CAC decide whether or not admit a new connection? The network must guarantee some QOS level to each connection that it admits. The two most important QOS measures are cell delay and cell loss probability. Both are closely related to the level of network congestion. In this chapter we focus only on the cell loss probability assuming that delay requirements can be met by restricting the buffer size.

• How does the network performance depend on the traffic parameters? The network performance should be a function of the traffic parameters that are chosen to characterize each connection. Many CAC schemes have been proposed, as reviewed in previous chapters. They vary widely in terms of computational complexity, computational accuracy, traffic models and parameters, memory usage, and in other ways.

Typical CAC methods employ the concept of effective bandwidth. The use of effective bandwidth simplifies the CAC procedure by estimating the total bandwidth of the aggregate traffic (including the new connection). The new call is accepted only if the updated total effective bandwidth is less than the link capacity. In the last chapter, we describe a CAC approach which uses NNs for effective bandwidth estimation. However, in this chapter, we attack the problem from a different angle. Considering statistical multiplexing between traffic sources, we directly calculate the effective bandwidth of the aggregate traffic rather than summing up individual bandwidths, hence overcoming the drawback of conventional methods. The aggregate arrival traffic is
characterized by four appropriately selected parameters and then accurately modelled by a two-state Markov Modulated Poisson Process (MMPP) via matching four important statistics. If the buffer size is large, admission control can be achieved by computing the effective bandwidth of the two-state MMPP. Figure 1 shows a block diagram of the proposed method. Simulation tests show that our approach is simple and results in higher utilisation compared with conventional methods.

2 THE MODEL
For traffic control mechanisms, we must define a unique set of parameters which can describe a wide range of service characteristics. Here we adopt the four parameters given in \([2]\) because they can be easily evaluated and monitored by traffic control.

\[
\begin{align*}
\varphi_1 & = \frac{\lambda}{\lambda + \mu} \\
\varphi_2 & = \frac{\mu}{\lambda + \mu} \\
\tau & = \frac{\lambda + \mu}{\lambda} \\
\alpha & = \frac{\lambda + \mu}{\lambda}
\end{align*}
\]

The aggregate traffic from \(A\) independent sources can also be represented by the four parameters \(m, \alpha, \tau, \nu\). The relationship between the aggregate traffic parameters and the individual ones are:

\[
\begin{align*}
\varphi_1 & = \sum_{i=1}^{A} \varphi_{i1} \\
\varphi_2 & = \sum_{i=1}^{A} \varphi_{i2} \\
\tau & = \sum_{i=1}^{A} \tau_i \\
\alpha & = \sum_{i=1}^{A} \alpha_i
\end{align*}
\]

From the above equation, we observe that all these statistics are of direct incremental or additive nature, which is important since connection requests arrive sequentially.

Next we approximate the aggregate traffic by a two-state MMPP because of its versatility and simplicity. The MMPP is a doubly stochastic Poisson process where the two states of a continuous-time Markov chain correspond to two Poisson processes. An MMPP model can be completely characterized by four parameters, \(R_1\), \(R_2\), \(\varphi_1\), and \(\varphi_2\), as shown in Figure 2. Parameter \(R_j\), \(j = 1, 2\), is the mean rate of a Poisson process in state \(j\), and the state duration time at each state has an exponential distribution with mean \(\varphi_j^{-1}\). The MMPP models can accurately characterize the aggregate arrival process because a large number of statistics can be matched and the correlation among the arrival rates can be captured over large time intervals. In \([3]\), Heffes and Lucantoni use an MMPP to successfully model average delay of voice packets through an infinite buffer multiplexer. In \([4]\), different sets of MMPP parameters are used to model the superposition of ON/OFF sources and packet loss in finite-buffered multiplexers. It has been shown that the approaches based on MMPPs are many orders of magnitude better than modeling the superimposed sources simply as a Poisson process. In \([3]\), Heffes and Lucantoni introduce a scheme to determine the MMPP parameters by matching them to the statistical moments of real traffic. However, the resulting MMPP model turns out to underestimate the queueing performance when the arrival traffic consists of bursty sources with high peak rate such as compressed video. So in \([2]\), a new scheme is proposed to evaluate the MMPP parameters based on the four traffic parameters of the aggregate traffic, i.e., \(m, \alpha, \tau, \nu\). The MMPP parameters \(R_j\), \(\varphi_j (j = 1, 2)\) are given by:

\[ q = \frac{\alpha}{1 + \alpha}. \]
Fig. 3: Cell loss probability vs. buffer size with traffic intensity = 0.6, τ = 0.1, T = 10.0, α = 5.0

The simulation results in [2] show that the performance of the proposed model appears as good as the Heffes and Lucantoni’s one for the multiplexed voice sources, and is much better for the integrated sources including video signals. Therefore, the MMPP model derived by this approach is a more suitable model for the integrated traffic of BISDN. Given the above MMPP model as the input process, let's consider an ATM statistical multiplexer with deterministic service rate and finite capacity. Hence, this multiplexer can be modeled as an M M P P / D / 1 / K queue. The details of the analysis of the M M P P / D / 1 / K queue are given in Appendix C, which leads to the evaluation of the cell loss probability. The effect of buffer size on the cell loss probability is shown in Figure 3. It is observed, as expected, that the cell loss probability is inversely proportional to buffer size. In Figure 4, Figure 5 and Figure 6, we show the cell loss probabilities as a function of traffic intensity. A buffer with finite capacity K of 32 cells is selected. We can observe that the traffic parameters ν, α and τ have a strong influence on the queueing behaviour.

\[ \bar{R}_1 = n_0 + \sqrt{\bar{q}_1 q_1}, \quad \bar{R}_2 = n_0 - \sqrt{\bar{q}_2 / q_2}, \quad \bar{q}_1 = q_1 / q_2, \quad \bar{q}_2 = (1 - q_1) / q_2 \]  

---

Fig. 4: Cell loss probability vs. traffic intensity for different values of variance, with α = 2.0, T = 10.0. The effect of buffer size on the cell loss probability is shown in Figure 3. It is observed, as expected, that the cell loss probability is inversely proportional to buffer size. In Figure 4, Figure 5 and Figure 6, we show the cell loss probabilities as a function of traffic intensity. A buffer with finite capacity K of 32 cells is selected. We can observe that the traffic parameters ν, α and τ have a strong influence on the queueing behaviour.

3 The Proposed C A C Scheme

Although the above model gives elegant formulas for computing cell loss probability in the M M P P / D / 1 / K queue, the applicability of this method seems limited to off-line buffer dimensioning (due to its computational complexity) rather than on-line performance evaluation (e.g., in CAC for determining if a specified QOS can be satisfied). Actually, in order to use the analytical result of the M M P P / D / 1 / K queue, the authors of [2] employ the effective bandwidth concept along with table-lookup procedure to perform CAC. The problem with this approach is that a huge table has to be maintained and it is not flexible enough to accommodate new
services and possible network changes. It also becomes computationally infeasible for large buffer sizes. Therefore, in this section, we propose a novel effective bandwidth approach to perform real-time CAC. The effective bandwidth of a time-varying source is the minimum amount of bandwidth required to satisfy its QOS.

Fig.5: Cell loss probability vs. traffic intensity for different values of peak-to-mean ratio, with $\nu = 0.1$, $\tau = 10.0$

Fig.6: Cell loss probability vs. traffic intensity for different values of autocovariance time coefficient, with $\nu = 0.1$, $\alpha = 2.0$

Fig.7: Effective bandwidth vs. buffer size width
The call admission problem is then solved by checking whether the effective bandwidth of the aggregate user population, including the new user, exceeds the service capacity. This value should be easily computable so that online computations may be carried out. Note that the notion of effective bandwidth is based on large deviations asymptotics for the tail probabilities of large buffer queues, which will be discussed in detail in the next chapter. Here we take the buffer overflow probability in an infinite buffer as an approximation to the cell loss ratio (QOS) in a finite buffer. For a very small overflow probability of $\epsilon$ and a large buffer size of $B$, we define $\theta$ as

$$\theta = -\log(\epsilon)/B.$$  

Then the effective bandwidth $\epsilon$ of the two-state MMPP can be calculated as [5]

$$\epsilon = \frac{1}{2\theta} \left( -a(\theta) + \sqrt{a^2(\theta) - 4b(\theta)} \right)$$  

(4)

where $a(\theta) = \varphi_1 + \varphi_2 - (e^\theta - 1)(R_1 + R_2)$ and $b(\theta) = (e^\theta - 1)^2 R_1 R_2 - (e^\theta - 1)(\varphi_1 R_2 + \varphi_2 R_1)$.

Figure 7 shows the effective bandwidth for 10 video sources (whose model will be described below) versus buffer size, with $\epsilon = 10^{-5}$. It is clear that effective bandwidth is a monotonically decreasing function of buffer size. Figure 8 shows the effective bandwidth per video source versus the number of multiplexed sources.

In this case, $\epsilon = 10^{-5}$ and the buffer size is 5 Mb. In the figure we include both our novel effective bandwidth calculation method and a conventional one proposed by Guerin et al. [6], which simply sums up individual effective bandwidths to estimate the total bandwidth required by the aggregate traffic. Obviously, statistical multiplexing gain is achieved in the proposed method.

4 Simulation Results

We have compared the efficiency of our approach with three other methods, i.e., peak rate allocation, Gaussian approximation and sum of individual effective bandwidths. The peak rate allocation scheme is the simplest one and it accepts or rejects calls on the basis of their peak bit rates. In this scheme, the QOS is always guaranteed because the aggregate bit rate will never exceed the link rate of the system. However, it leads to low utilisation of network resources. The Gaussian approximation of the required bandwidth is based on the assumption that the distribution of the stationary bit rate is Gaussian and is given by [6]:

$$c_g = \bar{m} + \beta \sigma$$  

(5)

where $\bar{m}$ and $\sigma$ represent the mean and standard deviation of the total arrival rate, respectively, and $\beta = \sqrt{-2 \log(\epsilon) - \log(2\pi)}$. For $N$ fluid flow ON/OFF sources, the sum of individual effective bandwidths is defined as [6]:
Fig.9: State transition diagram of the birth-death process

\[
c_s = \sum_{i=1}^{N} c_i = \frac{1}{2 \theta} \sum_{i=1}^{N} \left( -\frac{\lambda_i - \mu_i}{\theta i + \sqrt{(\lambda_i + \mu_i - \theta i)^2 + 4 \theta i \lambda_i \mu_i} } \right)
\]

where \( \frac{1}{\lambda_i} \) and \( \frac{1}{\mu_i} \) are the mean OFF and ON periods of source \( i \), respectively, and \( \tau_i \) is the corresponding peak rate.

We consider a single server queue (equivalent to an ATM link) with a deterministic service rate of 150 Mb/s.

The buffer size is taken to be 5 Mb and the desired overflow probability is set at \( 10^{-5} \). Two types of traffic sources are used. The first type represents a model for voice calls and the second for video telephony. A voice source can be characterized by the ON/OFF model we mentioned earlier \[4\]. In \[7\], Maglaris et al. consider the problem of modeling videotelephone scenes with a uniform activity level, e.g., showing a person talking. In their model, the arrival rate \( \tau_i \) is quantized into finite discrete levels. Transitions between levels are assumed to occur with exponential transition rates that may depend on the current level. Thus, the rate variations over time are approximated by a continuous-time process with discrete jumps at random Poisson times. This model is a discrete finite-state, continuous-time Markov process. Its state space is the set of the quantized levels up to a maximum level. This model is used in \[7\] to analyze the statistical multiplexer queue as a fluid flow reservoir that is filled from \( N \) variable rate sources each with rate \( \tau_i \).

It has been shown a birth-death Markov model as shown in Figure 9 will accurately describe the aggregate video source bit rate. The rate \( \tau_N(t) \) of the process in Figure 9 represents the quantized level of the aggregate bit rate of \( N \) sources. We assume uniform quantization step \( A \) bits/pixel, and \( M + 1 \) possible levels, \( \{0, A, \ldots, MA\} \). It is easy to see that the rate \( \tau_N(t) \) can be thought of as the aggregate rate from \( M \) independent minisources, each alternating between transmitting 0 bits/pixel (the OFF state) and \( A \) bits/pixel (the ON state) according to a Bernoulli distribution. A minisource turns ON with exponential rate \( \lambda \) and OFF with rate \( \mu \). Thus this model is almost the same as that used in analyzing statistical multiplexing of \( M \) voice sources.

The quantization step, the number of states, and the transition rates can be tuned to fit the mean, variance and autocovariance function of the measured data. The results are:

<table>
<thead>
<tr>
<th>Traffic source</th>
<th>( M )</th>
<th>( \lambda(1/s) )</th>
<th>( \mu(1/s) )</th>
<th>( r(Mb/s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>voice</td>
<td>1</td>
<td>0.65</td>
<td>0.352</td>
<td>0.064</td>
</tr>
<tr>
<td>video</td>
<td>10</td>
<td>3.078</td>
<td>2.5922</td>
<td>1.163</td>
</tr>
</tbody>
</table>

\[
A = \frac{C_N(0)}{E(\tau_N)} + \frac{E(\tau_N)}{M} \tag{7}
\]

\[
\mu = 3.9/(1 + \frac{E^2(\tau_N)}{MC_N(0)}) \tag{8}
\]
\[ \lambda = 3.9 - \mu \quad \text{(9)} \]

where \( E(r_N) \) and \( C_N(0) \) are the average and the variance of the aggregate arrival process from \( N \) identical and independent sources. They are given by 
\[ E(r_N) = 0.52N \text{ bits/pixel} \]and
\[ C_N(0) = 0.0536N(\text{bits/pixel})^2, \]respectively. Since there are about 250000 pixels per frame and 30 frames/s, 1 bit/pixel corresponds to 7.5 Mbits/s. The number of states \( M \) is set to be \( 10N \). It is found that this value of \( M \) yields reasonable results that are close to the measured data. The parameter values used for simulation are summarized in Table 1. In the table, we denote by \( M \) the number of ON/OFF sources required for characterizing one traffic source. In Figure 10, we plot the acceptable numbers of voice sources and video sources. It is clear that the proposed approach can accept more calls than other methods, while all cases meet the QOS requirement. This is because the Gaussian approximation assumes zero buffer and the conventional effective bandwidth approach ignores the statistical multiplexing effect between multiple sources. As expected, the peak rate allocation yields the poorest performance. Table 2 compares different utilizations obtained by four CAC schemes. We see that our CAC method can increase revenues by at least 10% over the other schemes.

![Fig.10: Comparison of acceptable number of sources by different CAC approaches.](image)

Solid line: proposed scheme; dashed line: Gaussian approximation; dotted line: sum of individual effective bandwidths; circled line: peak rate allocation

5 CONCLUSION

The CAC scheme proposed in this chapter is based on the concept of effective bandwidth. In contrast to conventional methods, we model the aggregate traffic from heterogeneous sources as a two-state MMPP by a novel matching technique and then estimate the required bandwidth. Since this approach takes the statistical multiplexing effect into account, it can lead to higher resource utilisation. The theory of effective bandwidth is an active area of research and is being applied to the design, simulation and analysis of high-speed ATM networks. In the next chapter, we address this topic and related issues in more detail. We use the theory of large deviations to provide a unified description of effective bandwidths for various traffic sources and the associated ATM multiplexer queueing performance approximations, illustrating their strengths and limitations. On the basis of this discussion, we propose a more accurate estimation method for ATM QOS parameters, which is a refinement of the original effective bandwidth approximation.

6. REFERENCES


**Author’s Profile**

Dr. M. Usha Rani is an Associate Professor in the Department of Computer Science and HOD for MCA, Sri Padmavati Mahila Visvavidyalayam (SPMVV Womens’ University), Tirupati. She did her Ph.D. in Computer Science in the area of Artificial Intelligence and Expert Systems. She is in teaching since 1992. She presented many papers at National and Internal Conferences and published articles in national & international journals. She also written 4 books like Data Mining - Applications: Opportunities and Challenges, Superficial Overview of Data Mining Tools, Data Warehousing & Data Mining and Intelligent Systems & Communications. She is guiding M.Phil. and Ph.D. in the areas like Artificial Intelligence, DataWarehousing and Data Mining, Computer Networks and Network Security etc.

K. Sailaja is a Research scholar in the Department of Computer Science (SPMVV) and working as assistant Professor in the Department of MCA, Chadalawada Ramanamma Engineering College, Tirupati. She is in teaching since 1999. She did her M.Phil in computer science in the area of WDM networks. She presented papers at National and Internal Conferences and published articles in national & international journals.