Analysis of a Statistical Multiplexer with Heterogenous Markovian On/Off Sources and Applications to Call Admission Control in ATM Networks *

Khaled M. Elsayed Harry G. Perros
Department of Computer Science
and
Center for Communications and Signal Processing
North Carolina State University
Raleigh, NC 27695-8207

Abstract

A computationally efficient algorithm for characterizing the superposition process of N heterogeneous and independent Interrupted Bernoulli Processes is introduced. The algorithm is then used to analyze a statistical multiplexer with finite buffer. Numerical examples highlighting the algorithm accuracy are given. We use the analysis to validate the equivalent capacity and heavy traffic approximation methods which are popular methods for call admission control in an ATM network. We also show the algorithm can be used to handle the case of homogeneous sporadic on/off sources.

1 Introduction

In an ATM environment many types of traffic, such as voice, data, and video, are to be efficiently transported by the same network. An ATM multiplexer receives cells (fixed size packets of 53 octets length) from a number of different incoming links and then transmits them out onto a single outgoing link. A finite buffer is provided in the multiplexer to accommodate the multiple arrivals

*Supported in part by BellSouth, GTE Corporation, and NSF and DARPA under cooperative agreement NCR-8919038 with the Corporation for National Research Initiatives and in part by a gift from BNR INC.

of cells. Each arrival stream is modeled by a bursty and possibly a correlated process. The service time is deterministic and is equal to one slot of the outgoing link which is assumed to be long enough to transmit one cell.

The analysis of such a queueing system is quite complex due to the large number of arrival processes. A possible method for approximately analyzing the queue is to first characterize the superposition process of all arrival processes, and then analyze the queue with a single arrival process. In this paper, we consider the case when the arrival processes are modeled as multiple independent heterogeneous Interrupted Bernoulli Processes (IBP). The IBP and its variants are popular models for bursty traffic sources in an ATM environment.

The problem of characterizing the superposition process of a set of arrival processes has been addressed extensively in the literature. One approach for obtaining the superposition process is to approximate it by a renewal process, see Albin [1], Whitt [27], Sriram and Whitt [22], and also Perros and Onvural [19]. An insightful discussion of the various time-scales affecting the accuracy of the approximate superposition of packet voice sources (modeled as a variant of IBP) was provided in [22]. Heffes and Lucantoni [10] considered the superposition process of packet voice sources modeled as an IBP where arrivals occur periodically. They approximate the superposition by a Markov Modulated Poisson Process (MMPP). The accuracy of the superposition is reasonable when the average delay in the multiplexer is the quantity of paramount importance. However, the superposition does not provide a good estimate for the probability of loss. Several other authors (see [3, 17, 26]) considered the same problem and suggested alternative methods for characterizing the superposition process as an MMPP. The main objective of these papers was to improve the accuracy with regards to calculation of the cell loss probability. Heffes [9] obtained an MMPP approximation to the superposition of different MMPP arrival processes using a set of simple expressions.

An alternative method to model ATM multiplexers is the Uniform Arrival and Service (UAS) model (also known as fluid flow). In this case an on/off process produces a uniform flow of bits when in the on state. Cell departures are modeled as a uniform flow out of the queue. Anick, Mitra, and Sondhi [2] evaluated the system performance using elegant simple expressions for a multiplexer with infinite buffer space and homogeneous arrivals. Tucker[25] considered the finite buffer case. The methodology was generalized to the case of heterogeneous Markov Modulated Rate Processes in a series of papers [24, 6].

The work presented in this paper is related to the methodology presented Hong, Perros, and Yamashita [11] and Makhamreh, McDonald and Georganas [16]. In [11], the authors used an

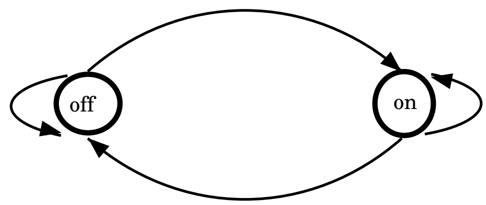


Figure 1: The Interrupted Bernoulli Process

aggregation method for approximating the superposition process. They first obtained the exact probability transition matrix of the Markov chain of the superposition process. This Markov chain has a dimension which is exponential in the number of input sources. Subsequently, the Markov chain was aggregated in order to obtain a superposition process with a small number of states (linear function of the number of sources). The method is limited to a small number of input sources. In [16], the authors used an aggregation method to analyze an output-buffered ATM switch with correlated imbalanced traffic.

In this paper, we present an efficient algorithm for characterizing the superposition process of multiple heterogeneous IBP's. The algorithm is based on a step-wise aggregation scheme. It can be used to construct the superposition of a large number of IBP sources. The space complexity of the algorithm is $O(N)$, while the computational complexity is $O(N^4)$, where N is the number of superimposed IBP's. The superposition algorithm is then used to study the performance of a statistical multiplexer with finite buffer. Finally, we present a brief study of the range of parameters of the superimposed IBPs for which the accuracy of the algorithm is acceptable.

The rest of this paper is organized as follows. In section 2, we briefly describe the IBP model and discuss some of its statistical properties. In section 3, we discuss the step-wise aggregation algorithm for characterizing the superposition process. Section 4 presents the analysis of the statistical multiplexer. Numerical results and a discussion of the algorithm accuracy are presented in section 5. In section 6, we present validation study of the equivalent capacity and the heavy traffic approximation methods for call admission control. We provide an approximate method for handling sporadic sources in section 7. Section 8 concludes the paper.

2 The IBP Source Model

We consider a source that alternates between active and idle states according to a Markov chain (see figure 1). Arrivals occur in a Bernoulli fashion with probability γ only when the source is in the active state. No arrivals occur when the source is in the idle state. The transitions between active and idle states occur in a memoryless fashion. Let us assume that at the end of slot k the process is in the active (or idle) state. Then, in the next slot $k+1$ it will remain in the active (or idle) state with probability α (or β), or it will change to idle (or active) state with probability $1-\alpha$ (or $1-\beta$).

The mean inter-arrival time $\bar{t}$ and coefficient of variation CV^2 are given by:

$$\bar{t} = \frac{(1-\alpha)+(1-\beta)}{\gamma(1-\beta)}, \; CV^2 = 1+\gamma \left[\frac{(1-\alpha)(\alpha+\beta)}{[(1-\alpha)+(1-\beta)]^2} - 1\right].$$

The triplet (α, β, γ) completely characterizes an IBP process.

3 The Approximation Algorithm

We now present an efficient algorithm that can be used to characterize the superposition of $N \geq 2$ heterogeneous IBP sources. The sources are divided into G groups, where the sources within each group are identical. In practice, specially when N is large, this assumption is valid because traffic sources of a specific traffic type (e.g. data or voice source) would tend to have similar or even identical characteristics. The number of sources in group i is given by N_i, where $1 \leq N_i \leq N$ and $\sum N_i = N$.

I- The Superposition Process of Homogeneous IBP Sources: The first step in the approximation algorithm is to characterize the superposition of sources in group i, where $N_i > 1$, in terms of a stochastic process with $N_i + 1$ states. The superposition process is described in terms of the probability transition matrix between the $N+1$ states, and the probability distribution of the number of arrivals per slot at each state.

The superposition process of N similar IBPs with descriptor (α, β, γ) is a Markov chain with state space $\chi = \{n, n = 0, 1, \cdots N\}$ which denotes the number of sources in the active state (or alternatively in the idle state). Let $\boldsymbol{A} = [a(n_1, n_2)]$, $n_1, n_2 \in \chi$ be the probability transition matrix governing the transitions between the states of the superposition. Also, let $b(n, i)$ be the probability that i arrivals occur when the superposition is in state n, where $0 \leq i \leq n$. Consider two states n_1 and n_2. Assume that in slot k the Markov chain is in state n_1. The probability that in the next slot $k+1$, the Markov chain would be in state n_2 is equal to the probability that l, $0 \leq l \leq n_1$, of the

n_1 sources in the active state make a transition to idle state and that m, $0 \leq m \leq N - n_1$, of the sources in the idle state make a transition to active state, and that $n_1 - l + m = n_2$. The probability of the first event to occur is given by the binomial probability density function $\binom{n_1}{l}(1-\alpha)^l \alpha^{n_1 - l}$. Similarly, the probability that the second event occurs is given by $\binom{N-n_1}{n_2-n_1+l}(1-\beta)^{n_2-n_1+l}\beta^{N-n_2-l}$ where $0 \leq n_2 - n_1 + l \leq N - n_1$ and $N - n_2 - l \geq 0$. We then have:

$$a(n_1, n_2) = \sum_{l=0}^{n_1} \Big[\binom{n_1}{l}(1-\alpha)^l \alpha^{n_1-l} \binom{N-n_1}{n_2-n_1+l}(1-\beta)^{n_2-n_1+l}\beta^{N-n_2-l} \\ I(n_2 - n_1 + l \geq 0)\ I(N - n_2 - l \geq 0)\Big]$$

where $I(x) = 1$ iff logical expression x is true and 0 otherwise.

When there are n sources each with probability γ of emitting a cell in a given slot, the the probability of having v arrivals is given by the binomial distribution $b(n, v) = \binom{n}{v}\gamma^v(1-\gamma)^{n-v}$, $0 \leq v \leq n$. For notational convenience, the quantities $b(n, v)$ are referred to by the lower triangular matrix $\boldsymbol{B} = [b(n, v)]$.

II- The Combined Superposition Process of two Groups of Sources: Once the superposition of two or more separate groups of sources is constructed as described above, it is necessary to combine them in order to get the overall superposition process. Let us assume that the superposition process of two arbitrary groups, say groups 1 and 2, has been characterized in terms of matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$, $i = 1, 2$, and that the number of states in process i is equal to $N_i + 1$. The superposition process is described by the state (n_1, n_2), where n_i is the state of component process i. The number of states in the superposition process is equal to $(N_1 + 1)(N_2 + 1)$. The probability transition matrix of the superposition will be given by $\boldsymbol{A} = \boldsymbol{A}_1 \otimes \boldsymbol{A}_2$ where $\otimes$ is the Kronecker product operation of two matrices.

One of the useful properties of the Kronecker product operation is that $\vec{\boldsymbol{\pi}}$, the invariant probability vector of $\boldsymbol{A}$, is given by $\vec{\boldsymbol{\pi}}_1 \otimes \vec{\boldsymbol{\pi}}_2$ where $\vec{\boldsymbol{\pi}}_i$ is the invariant probability vector of $\boldsymbol{A}_i$. This saves us from solving a system of linear equations for finding the invariant of $\boldsymbol{A}$ which is more time consuming than solving for $\vec{\boldsymbol{\pi}}_1$ and $\vec{\boldsymbol{\pi}}_2$ individually.

Note that the above discussion is also valid for more than two groups of sources. However, we focus here on the case of two groups since the approximation algorithm always superposes two processes at a time.

III- Aggregation of the Superposition Process: The above Markov chain which characterizes the superposition process of two groups of IBP sources is exact. However, the dimensionality of the resulting process may not be practical for studying the performance of a multiplexer with a large buffer. The dimensionality of the superposition process can be reduced by lumping all the states (n_1, n_2) of the superposition, where $n_1 + n_2 = n$, to a single state n. The Markov chain resulting from this aggregation represents what we call the *compact* superposition process.

Let the probability transition matrix of the compact process be $\tilde{\boldsymbol{A}} = [\tilde{a}(n, m)]$. We then have

$$\tilde{a}(n', n'') = \begin{array}{l} \left[\sum_{n'_1+n'_2=n'} \sum_{n''_1+n''_2=n''} \pi(n'_1, n'_2) a\left((n'_1, n'_2), (n''_1, n''_2)\right)\right] / \\ \left[\sum_{n'_1+n'_2=n'} \pi(n'_1, n'_2)\right] \end{array} \tag{1}$$

The remaining step is to find the probability distribution function of the number of arrivals in a particular state. This can be calculated from $\boldsymbol{B}_i, i = 1, 2$ as follows:

$$b(n, v) \;=\; \begin{array}{l} \left[\sum_{n_1+n_2=n} \sum_{m=max(0,v-n_2)}^{min(v,n_1)} \pi(n_1, n_2) \left[b_1(n_1, m)\, b_2(n_2, v-m)\right]\right] / \\ \left[\sum_{n_1+n_2=n} \pi(n_1, n_2)\right] \end{array} \tag{2}$$

The resulting superposition process has a characterization identical to that of its component processes. This leads to the following simple iterative scheme.

IV- The Approximation Algorithm: Consider N heterogeneous IBP sources divided into G groups each having $N_i, i = 1, 2, \cdots G$, identical sources. Then, we have the algorithm outline as follows:

- ⋄ Find $\boldsymbol{A}_j$, $\vec{\boldsymbol{\pi}}_j$ and $\boldsymbol{B}_j$ for all groups j, $j = 1, \cdots, G$
- ⋄ Let $\boldsymbol{A}_s = \boldsymbol{A}_1$ and $\boldsymbol{B}_s = \boldsymbol{B}_1$
- ⋄ For $j = 2$ to G do
    - ∘ $\boldsymbol{A}' = \boldsymbol{A}_s \otimes \boldsymbol{A}_j$
    - ∘ $\vec{\boldsymbol{\pi}}' = \vec{\boldsymbol{\pi}}_s \otimes \vec{\boldsymbol{\pi}}_j$
    - ∘ Aggregate $\boldsymbol{A}'$ into $\boldsymbol{A}_s$ (use equation 1)
    - ∘ Calculate $\boldsymbol{B}_s$ (use equation 2)
    - ∘ Find $\vec{\boldsymbol{\pi}}_s$ satisfying $\vec{\boldsymbol{\pi}}_s \boldsymbol{A}_s = \vec{\boldsymbol{\pi}}_s, \sum_i \pi_s(i) = 1$

End

It can be shown that the computational complexity of the algorithm is $O(N^4)$ (The calculation of the Kronecker product of two matrices and the solution of a linear system of equations are the dominant factors.)

Hong, Perros, and Yamashita [11] tackled the same problem. However, their proposed algorithm constructs the complete probability transition matrix of the superposition process and then performs the aggregation. This causes the computational complexity and storage requirements of their method to be $O(2^N)$. This precludes the execution of the algorithm on a conventional computer. Even for a nominal value of N, say $N = 20$, it may not be possible to construct the superposition even using a supercomputer. We conducted a set of experiments where we used identical input mix and multiplexer parameters as the test cases reported in an earlier version of [11] and the achieved level of accuracy relative to simulation is the same. However, our method is considerably computationally more efficient.

4 Analysis of the Multiplexer Model

Consider a FIFO finite buffer multiplexer serving $N \geq 2$ IBP sources. The multiplexer has $S \geq 1$ servers and can accommodate a total of $B \geq S$ cells at any time instant including those in service. The service time for all cells is constant and is equal to one time slot. The multiplexer can serve S cells every time slot. We assume that $N > S$, otherwise no queue will ever form in the multiplexer and the problem will be trivial to handle.

We seek the steady state probabilities, $\pi(n), 0 \leq n \leq B - S$, that there are n cells in the multiplexer's queue. From this, we can obtain other measures of interest such as the mean queue length, the probability of full buffer and the cell loss probability.

Let us first discuss the timing of events in our system. We follow an *early arrival* timing model as defined by Hunter [13]. That is, during an arbitrary time slot, the following sequence of events is possible: potential state transition in the superposition occurs, followed immediately by cell arrivals (if any), which is followed by service of waiting cells if there are any, and finally departure of cells that received service. If an arbitrary cell sees one or more empty servers upon its arrival, it is immediately admitted to one of the available servers without waiting till the next slot. Hence, the multiplexer is effectively of the cut-through type.

The system state is described by the pair (n, h) where n is the number of customers in the system immediately after the end of a slot and h is the state of the superposition process at the current slot.

We first characterize the superposition process of the N IBP sources as described in the previous section. Then, the probability transition matrix of the Markov chain (n, h) is generated. Let $p\left[(n_1, h_1) \rightarrow (n_2, h_2)\right]$ be the probability that the Markov Chain (n, h) makes a transition from state (n_1, h_1) in slot k to state (n_2, h_2) in slot $k+1$. Then

$$p\left[(n_1, h_1) \rightarrow (n_2, h_2)\right] = \sum_{v} b(h_1, v) a(h_1, h_2)$$

where $n_2 = (min(n_1 + v), B) - S)^+$ and v, $0 \leq v \leq h_1$, is the random variable representing the number of arrivals when the superposition process is in state h_1. Finally, we solve for the invariant probability vector of $\boldsymbol{P} = \left[p\left[(n_1, h_1) \rightarrow (n_2, h_2)\right]\right]$ from which various performance metrics can be obtained. We used the Gauss-Seidel method when using a workstation, and the parallelized LAPACK routines for LU decomposition when using a KSR-1 parallel machine at the North Carolina Supercomputer Center.

4.1 Special Case: Single Server Multiplexer

When the number of servers in the multiplexer is one, the the efficient method of Blondia and Casals [4] for analysis of the B-DMAP/$D/1/B$ queue can be used. The B-DMAP is the batch discrete-time Markovian arrival process described as follows. Suppose at time slot k, the process is in some state i, $1 \leq i \leq m$. At the next time instant the k+1, a transition to another or the same state occurs, and a batch arrival may or may not occur. With probability $(d_0)_{i,j}$, $1 \leq i, j \leq m$, there is a transition to state j without an arrival, and with probability $(d_l)_{i,j}$, there is is a transition to state j with a batch arrival of size l. We have that $\sum_{l=0}^{\infty} \sum_{j=1}^{m} (d_l)_(i,j) = 1$. Define the matrices $\boldsymbol{D}_l = [(d_l)_{i,j}]$, $l = 0, 1, \cdots$. The B-DMAP/$D/1/B$ queue is modeled by the M/G/1-type matrix given by:

$$Q = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 & \cdots & \boldsymbol{D}_{B-1} & \sum_{k=B}^{\infty} \boldsymbol{D}_k \\ \boldsymbol{D}_0 & \boldsymbol{D}_1 & \cdots & \boldsymbol{D}_{B-1} & \sum_{k=B}^{\infty} \boldsymbol{D}_k \\ \boldsymbol{O} & \boldsymbol{D}_0 & \cdots & \boldsymbol{D}_{B-2} & \sum_{k=B-1}^{\infty} \boldsymbol{D}_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{D}_0 & \sum_{k=1}^{\infty} \boldsymbol{D}_k \end{bmatrix}$$

The steady state queue length distribution and cell loss probability can be evaluated using the method shown in [4].

In order to use this method in our analysis, we map the superposition process of the N IBP sources into a B-DMAP. Suppose the superposition process of IBP sources, as shown in section 3, is characterized by the probability transition matrix $\boldsymbol{A} = [a(i,j)]$ and the batch arrivals probability distributions $b(i,l)$, the probability of l arrivals per slot at state i, $0 \leq l \leq i$ and $0 \leq i \leq N$. Then, the entries of the matrices $\boldsymbol{D}_l$, $0 \leq l \leq N$ are defined by:

$$\begin{aligned} d_l(i,j) &= 0, & l > j \\ &= a(i,j)b(j,l), & 0 \leq l \leq j. \end{aligned}$$

We have $\boldsymbol{D}_l = 0$ for $l > N$ since no batch size more than N could occur. The methodology of [4] can be applied directly and the time complexity of the solution becomes $O(BN^3)$ instead of $O(B^3N^3)$ if a direct approach is used for solving the linear system. This is particularly useful for systems with large buffer sizes.

5 Validation and Numerical Examples

The approximation algorithm was validated by comparing it against simulation. The confidence intervals of our simulation were very narrow. Most of the figures are plotted without showing the confidence intervals for this reason. We only show the confidence intervals for the most stringent case of a multiplexer with an expected small cell loss probability. This will be addressed in example 4. The effect of the heterogeneity of the sources on the accuracy of the algorithm was investigated and metrics were defined to determine how heterogeneous a given set of sources is. These metrics can be used to determine in advance whether the superposition will be accurate or not.

In all the validation examples (except for example 4), we fix γ_j, the probability that source j emits a cell while in the active state, to be 1. This simplifies the fitting procedure for characterizing an IBP from the first two moments of the inter-arrival time. The following additional notation is used.

Table 1: Parameters of the IBP sources for the second example

i	N_i	r_i	CV_i^2
1	2	0.1	10.00
2	2	0.1	20.00
3	2	0.1	30.00
4	2	0.1	40.00

N_i = Number of sources at group i.

r_i = Source throughput (mean rate) for a source in group i.

R_r = $max\{r_i\}/min\{r_i\}$ is a measure of heterogeneity of the throughput of the input sources.

CV_i^2 = Squared coefficient of variation of inter-arrival time for a source in group i.

R_{cv2} = $max\{CV_i^2\}/min\{CV_i^2\}$ is a measure of heterogeneity of the CV^2 of the input sources.

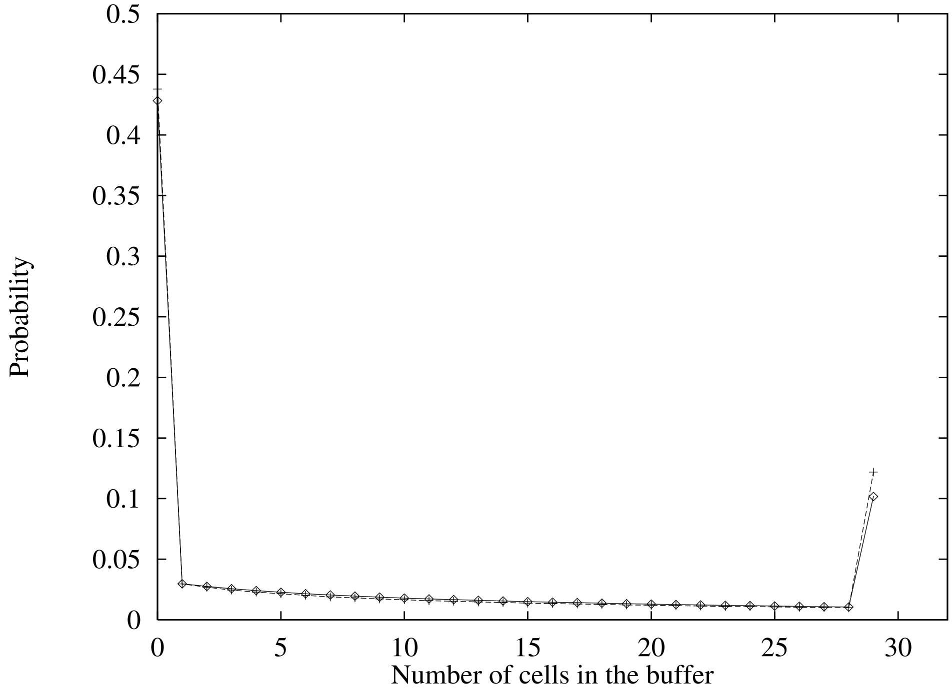
λ = $\sum_{i=1}^{G} N_i r_i$ = Total throughput.

B = Total buffer size.

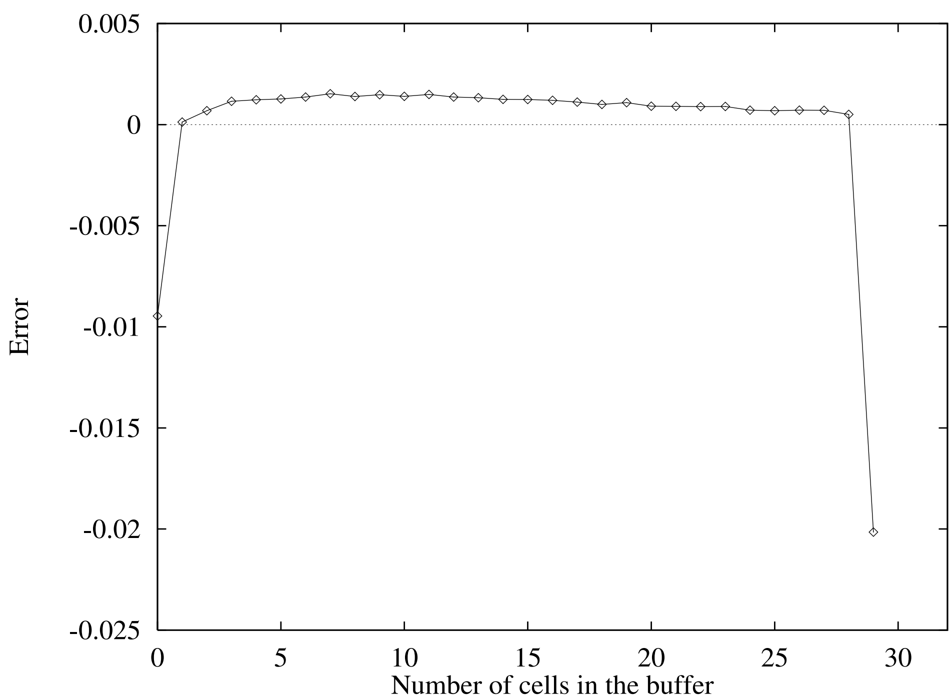
In all examples, the number of servers in the multiplexer was set to 1.

Example 1: In the first example, we have $G = 4$ and $B = 30$ and we use the input sources with parameters given in table 1.

In figure 2(a), we plot the distribution of the number of cells in the queue which is obtained from simulation and the approximation algorithm. The difference between the numerical and simulation values is shown in figure 2(b). The reason why we give the difference and not the relative error is that some of the calculated quantities are very small (in the order of 10^{-3}) which makes the relative error to be large, despite the fact that the overall accuracy is still acceptable. The probability distribution has two peaks, the first is for the probability that queue is full and the second peak for the probability that the queue is empty. This a well-known behavior in queues with bursty correlated input with a moderate to a high input traffic intensity. We note that the magnitude of error of the steady-state probability of a full buffer is high. This causes an underestimation of the cell loss probability. However, it should be stated that for this example the input sources are quite heterogeneous. The value of R_{cv2}, which indicates the achieved accuracy of the algorithm, is 4 a fairly large value as it will be shown below.



(a) Probability distribution of the number of cells



(b) Error

Figure 2: Probability distribution of the number of cells and associated error, $B = 30$

Table 2: Parameters of the IBP sources for the third example

i	N_i	r_i	CV_i^2
1	2	0.1	10.00
2	2	0.1	12.00
3	2	0.1	14.00
4	2	0.1	16.00

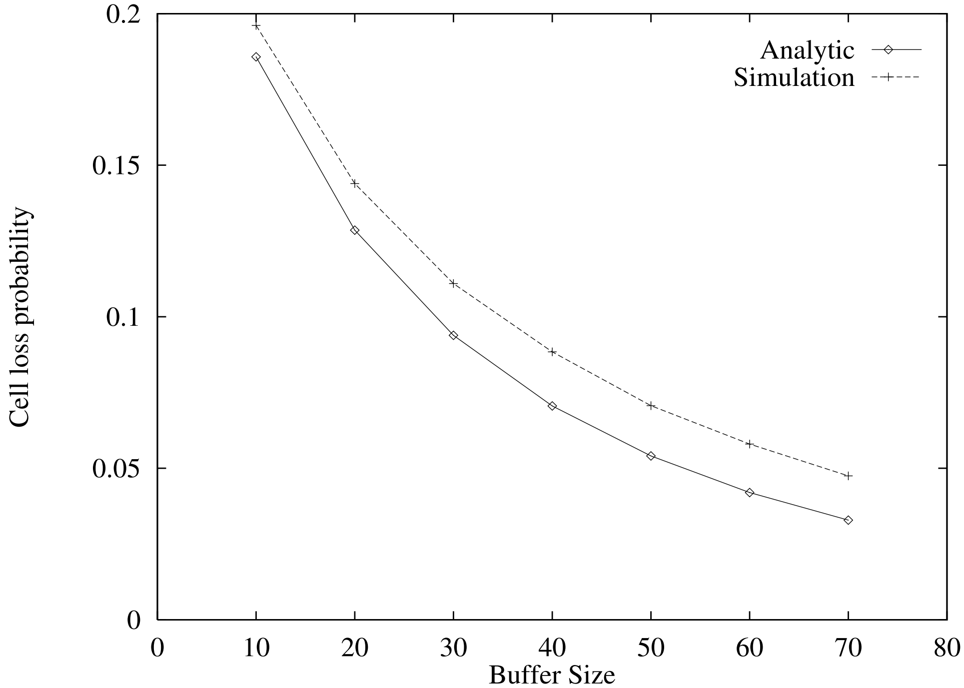
Example 2: In the second example, we use the input sources with parameters given in table 1, with $R_{cv2} = 4$. We first obtain the superposition process, and subsequently use it to generate the probability transition matrix for buffer sizes in the set $\{10, 20, 30, 40, 50, 60, 70\}$.

The cell loss probability and the mean queue length are shown in figures 3(a) and 3(b) respectively. It can be seen that the accuracy of the estimated mean queue length is very good while that of the cell loss probability is less accurate. This is due to the smoothing effect of the aggregation algorithm. It constructs a less bursty and correlated process than the actual superposition, which leads to underestimating the probability of observing a full buffer. However, we noted that the aggregation algorithm provides an upper-bound on the probability of having an empty buffer. Thus a balancing effect takes place and the overall mean queue length predicted by the simulation and numerical solution is close.

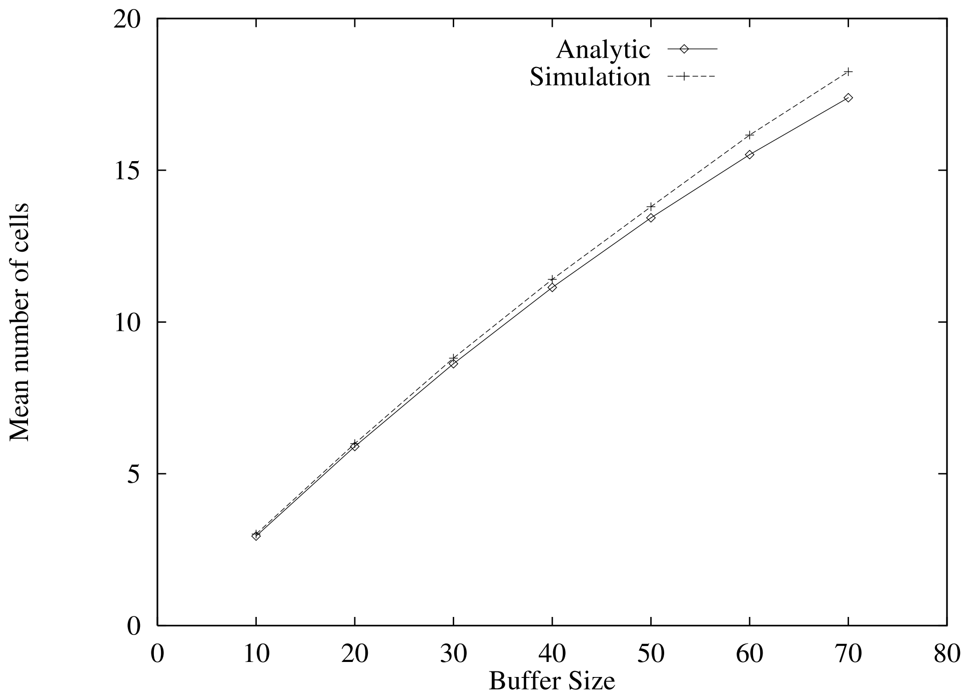
We also investigated the effect of increasing the number of sources in a group while keeping the overall group throughput fixed. The parameters of table 1 were modified as follows. N_i is increased to 4, r_i is halved to 0.05, for all $i = 1, \cdots, 4$ and the values of CV_i remained the same. For this new set of parameters, the cell loss probability and mean queue length slightly increased over the values plotted in figure 3. The accuracy remained the same.

Example 3: The third example is in essence similar to the second one. We let $R_{cv2} = 1.6$ and the parameters of the input sources are given by table 2.

The cell loss probability and the mean queue length are shown in figures 4(a) and 4(b) respectively. We note here that the accuracy of the approximation algorithm is very good. This is due to the fact that the amount of heterogeneity in the input sources is small as indicated by the relatively small value of R_{cv2}.

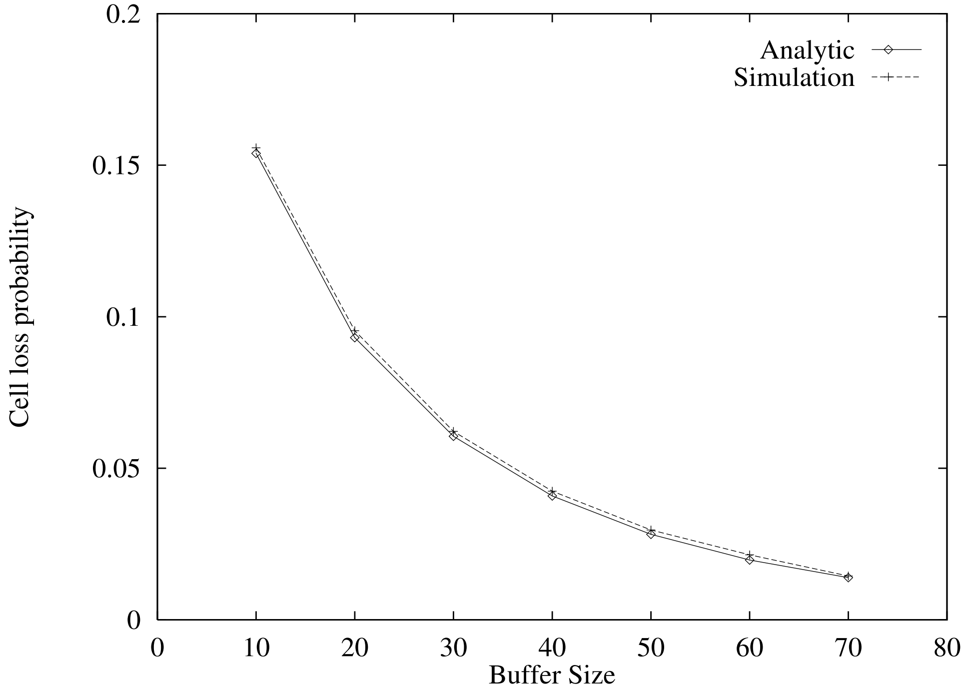


(a) Cell loss probability

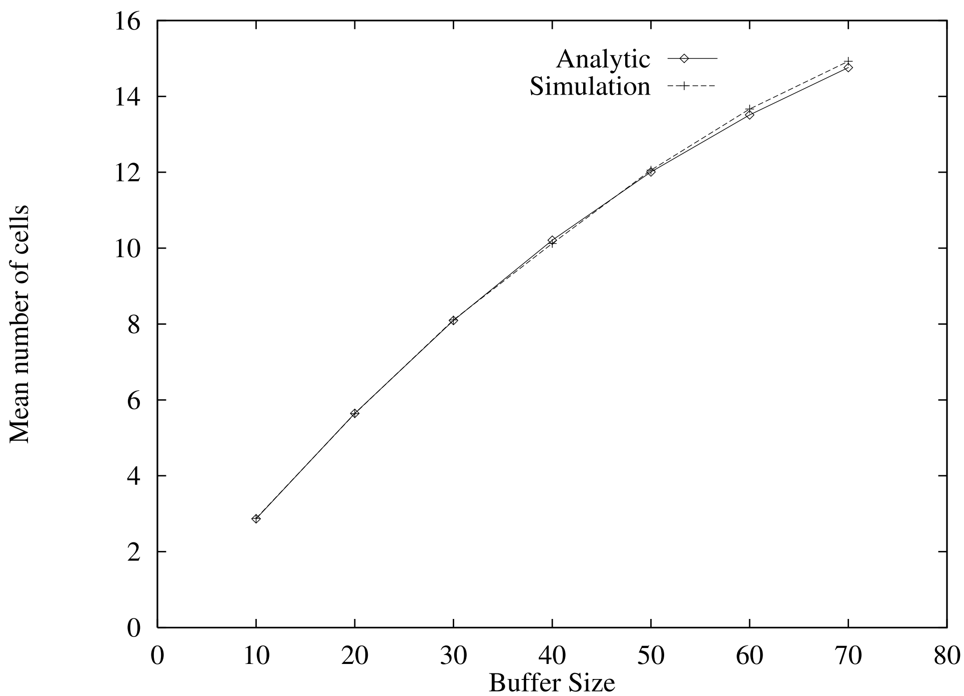


(b) Mean number of cells

Figure 3: Cell loss probability and mean number of cells



(a) Cell loss probability



(b) Mean number of cells

Figure 4: Cell loss probability and mean number of cells

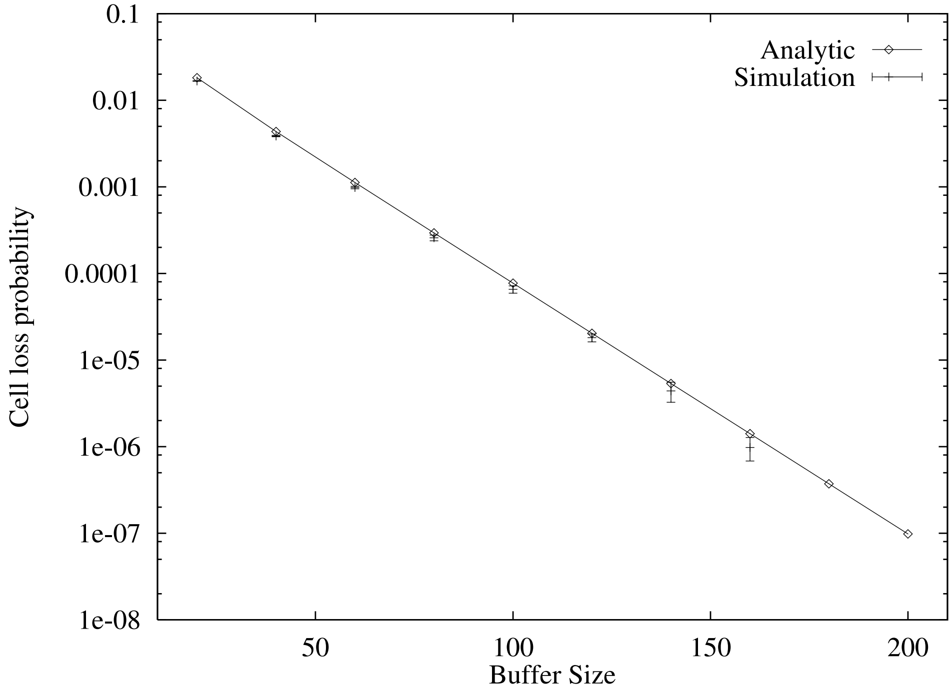
Example 4: We now consider a realistic model of the multiplexer where the cell loss probability is expected to be small. We have two groups of sources. The first group has $N_1 = 17$, and a source in the group is specified by a mean on period of 50 ($\alpha_1 = 0.98$), a mean off period of 200 ($\beta_1 = 0.995$), and the probability of an arrival during a slot in the on period, γ_1, is equal to 0.1. The second group has $N_2 = 9$, and a source in the group is specified by $(\alpha_2, \beta_2, \gamma_2) = (0.96, 0.986666667, 0.2)$. The buffer size is increased from 20 to 200 in increments of 20.

The cell loss probability and the mean number of cells obtained by the approximation algorithm and simulation are shown in figure 5. In this case, we show the confidence intervals for the simulation. We conducted a long simulation with 20 independent replications. A single run would be finished when all of the sources has undergone at least 100,000 state changes. The confidence intervals for the cell loss probability and mean number of cells are then evaluated using standard techniques. The cell loss probability for buffer sizes greater than 150 are not shown since for such small values, it is not possible to obtain a valid estimate from the simulation. The results indicate the applicability of using the algorithm for operating regions with very small expected cell loss probability. In such regimes, simulation fails to provide reliable answers in a reasonable amount of time.

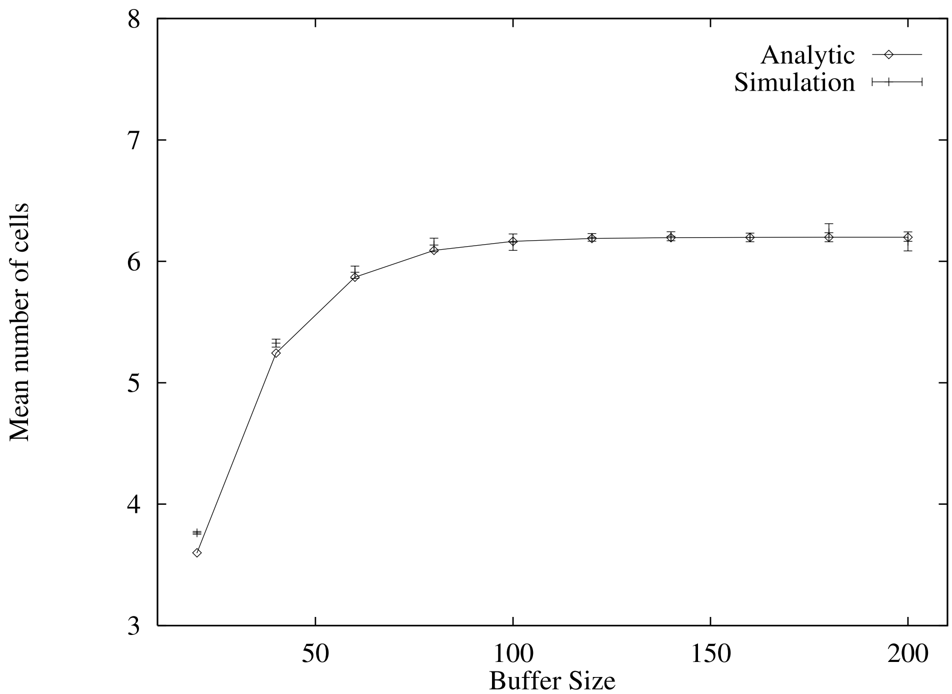
5.1 Effect of the Source Heterogeneity on the Accuracy of the Superposition

We conducted a set of experiments to study the effect of the variation in IBP source throughput and coefficient of variation on the accuracy of the approximation algorithm. We know in advance that the more heterogeneous the sources are, the worse the accuracy of the iterative aggregation algorithm would become. However, we wanted to study the regions for the parameters of the sources in which the accuracy of the algorithm is acceptable. This is explored further in examples 5 and 6. The conclusion of this study was that R_{cv2} is a major deciding factor in affecting the accuracy of the superposition process.

Effect of the variation in the squared coefficient of variation of the inter-arrival time: We study the effect of the variation in the squared coefficient of variation of the inter-arrival time of the input sources on the accuracy. We constructed an input of four groups each having two sources, and fixed the source throughput to be 0.1. We vary the R_{cv2} to take values from the set $\{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ respectively. We assume that $CV_1^2 = 10.0$, and for each value of R_{cv2},



(a) Cell loss probability



(b) Mean number of cells

Figure 5: Cell loss probability and mean number of cells

we find the remaining CV_i^2 by solving the equations:

$$CV_i^2/CV_{i-1}^2 = \theta, \quad i = 2, 3, 4$$

where θ is set equal to $\sqrt[3]{R_{cv2}}$.

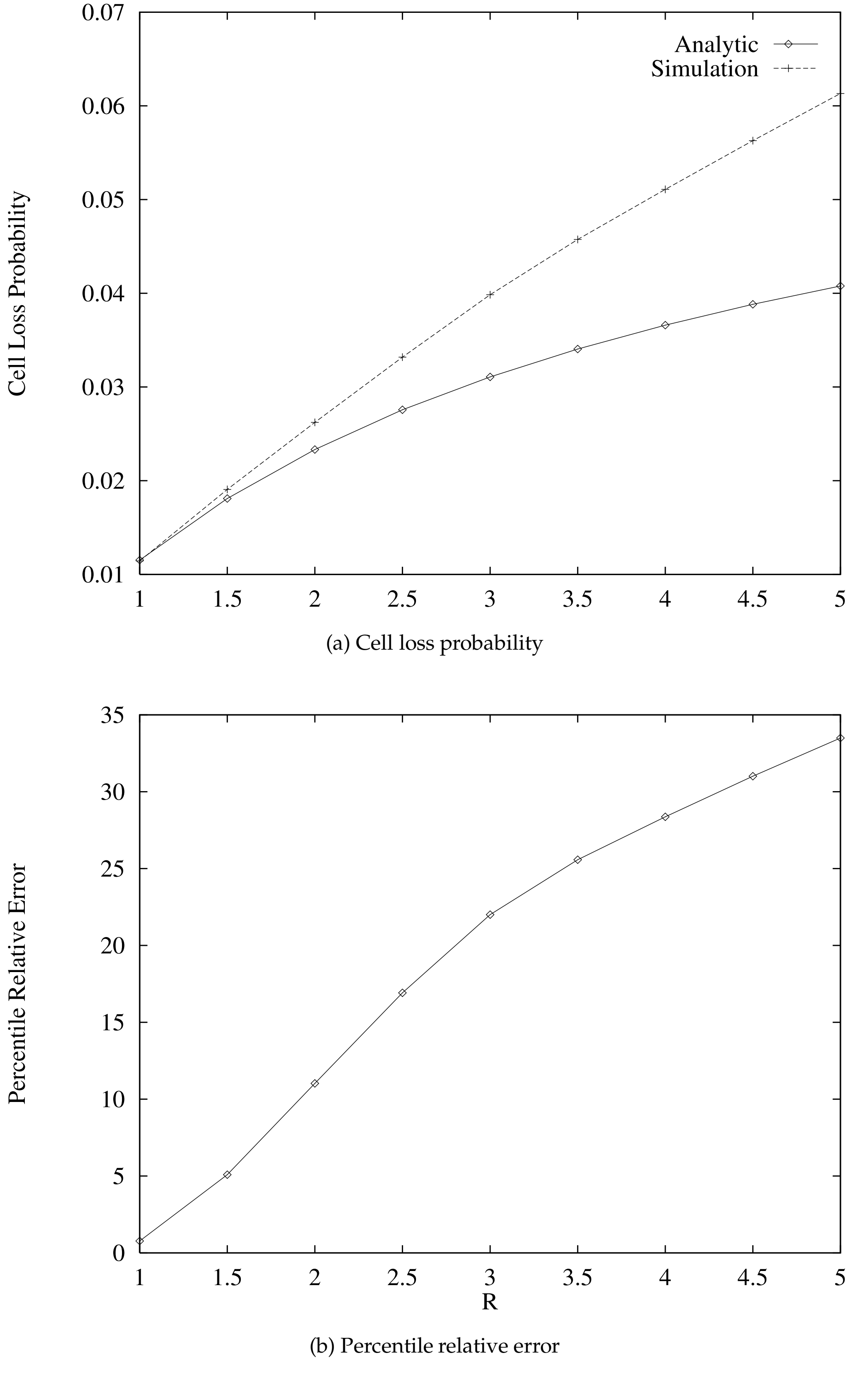
We plotted the cell loss probability, obtained from the approximation algorithm and simulation, and the percentile error as a function of R_{cv2} in figures 6(a) and 6(b) respectively. We note that the relative error is an increasing pseudo-linear function of R_{cv2}. We observed that in general if $R_{cv2} > 3$, then the approximation algorithm does not perform satisfactorily. This seems to hold true independent of the maximum or minimum values of CV_i^2. In some cases, the accuracy actually improves when the minimum and maximum value of the CV_i^2 are increased, while keeping the value of R_{cv2} fixed. This may be in contrast to what one would initially conjecture that the larger the absolute values of CV_i^2 the worse the algorithm would perform.

Effect of the variation in the throughput: We study the effect of variation in the throughput of the input sources. We consider four groups of sources each having a single source. The total throughput is fixed at 0.8 and each source has a squared coefficient of variation equal to 20. We let R_r take values from the set $\{1, 10, 20, 30, 40, 50\}$. The throughput of a single source $r_i, i = 1, \cdots 4$ is calculated by solving the equations:

$$r_i/r_{i-1} = \theta, i = 2, 3, 4, \ \sum_{i=1}^{4} r_i = 0.8$$

where $\theta = \sqrt[3]{R_r}$.

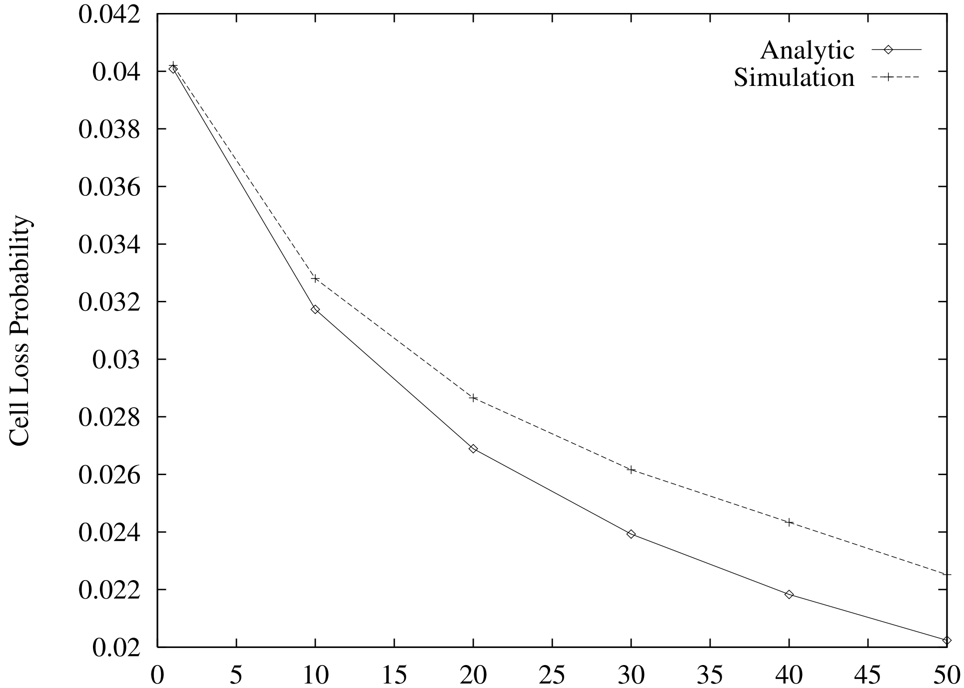
We plotted the cell loss probability, obtained from the approximation algorithm and simulation, and the percentile error as a function of R_r in figures 7(a) and 7(b) respectively. We note here that the accuracy of the algorithm is less sensitive to variations in the throughput by many orders of magnitude than the coefficient of variation. Even for very large values of R_r, say 40, the relative error is about 10%. In addition, we have observed that the algorithm becomes less sensitive to R_r when there are more than one source in a given group. In other words, the case of one source per group is an extreme case in which the effect of heterogeneity in the throughput of the input sources is relatively large.



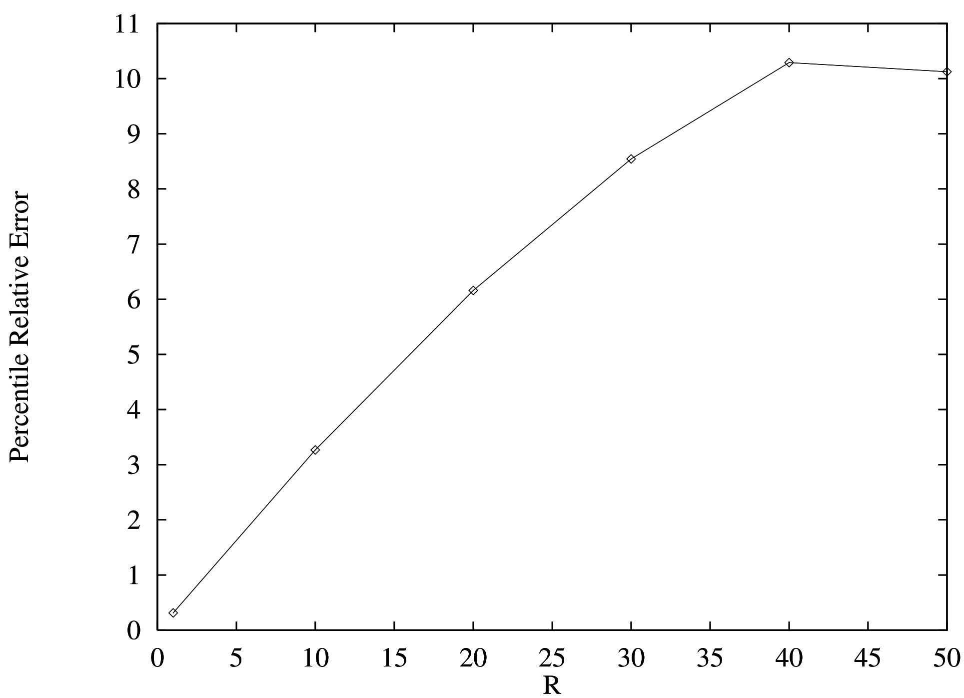
(a) Cell loss probability

(b) Percentile relative error

Figure 6: Cell loss probability and relative error vs. R_{cv2}



(a) Cell loss probability



(b) Percentile relative error

Figure 7: Cell loss probability and relative error vs. R_r

6 Application to Call Admission Control in an ATM Network

The primary role of a network congestion control procedure is to protect the network and the user in order to achieve network performance objectives and optimize the usage of network resources. In ATM-based B-ISDN, congestion control should support a set of ATM quality of service classes sufficient for all foreseeable B-ISDN services.

Call admission control (CAC) is one of the primary mechanisms for preventive congestion control in an ATM network. CAC is one particular type of many possible resource allocation mechanisms performed by the network provider. In an ATM network, resource allocation can be identified on three different levels: call, burst, and cell levels [12].

In the context of ATM networks, the CAC process is described as follows. During the call setup phase, users declare and/or negotiate with the network their connection characteristics and their required quality of service. Some of the parameters that may be used to specify a call characteristics are sustainable (average) bit rate, peak bit rate, cell delay variation tolerance, and maximum burst length, as recommended by the ATM Forum [7]. The required quality of service may include low values of the cell loss probability, maximum delay, and/or jitter. The network provider decides if the new call can be admitted based on the current network load, available resources, and the call's requested quality of service.

In this section we mainly consider loss networks in which the primary quality of service is the cell loss probability. In this context we include two methods that have been reported in the literature for performing call admission control. The two methods are the equivalent capacity method [5, 14, 8] and the heavy traffic approximation method [20, 21]. In these methods traffic sources are characterized by means of 2-state Markovian on/off processes. Hence, we can use our method for the analysis of statistical multiplexers with heterogneous IBP sources to validate the results obtained by the CAC methoods. At the same time, we can also compare the performance of these two CAC methods.

An input source is characterized by the triplet (R, r, b), where R is the peak rate in cells/sec, r is the average rate of the source, and b is the mean burst length of the source. The traffic model assumed by the two methods is a sporadic on/off source with geometrically distributed on and off periods in which arrivals occur with rate R during the on period. The triplet (R, r, b) completely specifies the input sources. We review the equivalent capacity and heavy traffic approximation methods in the following subsections.

6.1 The Equivalent Capacity Method

The equivalent capacity of source reflects the source's characteristics including burstiness and service requirements. The equivalent capacity of a source is independent of traffic submitted by other sources to the multiplexer. Therefore, the complexity of the computation of the equivalent capacity depends only on the source, and not on the overall system. This concept is attractive to network designers because it might serve as a bridge to the familiar trunk reservation methods in circuit switching. The equivalent capacity is defined as follows. Consider a single-channel finite buffer multiplexer model and assume that there is only a single call using the multiplexer. Then, the equivalent capacity of the call is defined as the link capacity that needs to be allocated for the call in order to guarantee a cell loss probability for the call less than a specific value ϵ.

Elwalid and Mitra [5] showed that the equivalent capacity of a Markovian fluid source is (approximately) the maximum real eigenvalue of a matrix derived from source parameters, multiplexer resources, and the cell loss probability quality of service. Using large deviations theory, Kesidis, Walrand and Chang [14] provide a more general framework for various models of Markov modulated sources inlcuding discrete and continuous time models. We provide here the final results without the involved theory.

Consider a traffic source modeled as an L-state source $(\boldsymbol{Q}, \vec{\boldsymbol{\lambda}})$ where $\boldsymbol{Q}$ is the probability transition matrix of the modulating Markov chain and $\vec{\boldsymbol{\lambda}} = (\lambda_0, \lambda_1, \cdots, \lambda_{L-1})$ is the vector of peak arrival rates at the various states. Let $\bar{\lambda}$ be the mean rate and $\hat{\lambda}$ be the maximum rate. The buffer is served by a channel of capacity C ($C = 1$ after normalization). Define $G(B) = Pr[x \geq B]$, where x is the stationary buffer content. Then the grade of service is $G(B) \leq \epsilon$. The equivalent capacity e of a source $(\boldsymbol{Q}, \vec{\boldsymbol{\lambda}})$ was shown to be [14]:

$$e = log(\Omega\{exp(\delta \boldsymbol{\Lambda})Q\})/\delta \tag{3}$$

where $\boldsymbol{\Lambda} = diag(\vec{\boldsymbol{\lambda}})$ and $\delta = -\frac{log(\epsilon)}{B}$, and $\Omega\{M\}$ is the spectral radius of matrix M. The value of e satisfies the relation: $\bar{\lambda} \leq e \leq \hat{\lambda}$. In other words, for our assumed traffic model, $r \leq e \leq R$. When N sources are multiplexed the total equivalent capacity is approximated by $\sum_{i=1}^{N} e_i$.

For a 2-state Markovian source described by the triplet (R, r, b), the equivalent capacity, e, of the source can be shown to be equal to:

$$e = log(\frac{1}{2}\left[Q_{11} + exp(\delta R)Q_{22} + \sqrt{(Q_{11} + exp(\delta R)Q_{22})^2 + 4exp(\delta R)(1 - Q_{11} - Q_{22})}\right])/\delta \tag{4}$$

where $Q_{11} = ((R-r)b - r)/(R-r)b$ and $Q_{22} = 1 - 1/b$.

When N Markovian on/off sources are multiplexed, Guérin, Ahmadi and Naghshineh [8] suggest the following approximation for the total equivalent capacity:

$$e = min(\rho + a'\sigma, \sum_{i=1}^{N} e_i) \quad (5)$$

where e_i is the equivalent capacity of source i (calculated assumed that it is the only source in the system), $\rho = \sum r_i$ is the total average rate, $\sigma = \sum \sigma_i$, where $\sigma_i^2 = r_i(R_i - r_i)$ is the variance of rate of source i, and $a' = \sqrt{-2log(\epsilon) - log(2\pi)}$. This approximation is based on two observations. Firstly, the multiplexed N sources may well correspond to an effective bandwidth which is less then the sum of the individual effective bandwidths $\sum_{i=1}^{N} e_i$. Secondly, the stationary bit rate of N sources has been observed to approximately follow a normal distribution with mean ρ and variance σ. Assuming, therefore, that the multiplexer has no buffer, the probability that the cell loss is ϵ is equivalent to the probability that the aggregate bit rate > C. Hence the equivalent capacity is equal to the aggregate rate past which the area under the normal distribution curve is greater or equal to ϵ. This aggregate rate is approximately $\rho + a'\sigma$, where a', a point in the normal distribution N(0,1) past which the area under the curve is ϵ, is obtained by inverting the normal distribution. This approximation becomes useful for cases where the buffer size is very small in comparison with the mean length of the on period a source. In such cases, losses are bound to occur once the aggregate bit rate exceeds the link capacity. The buffer size would not be sufficient to absorb the bursts and hence the assumption of a buffer-less system is well justified.

6.2 The Heavy Traffic Approximation Method

In a series of papers [20, 21], Sohraby proposed an approximation for the asymptotic behavior of the tail of the queue-length distribution in a statistical multiplexer. The model used is a discrete-time queueing model, which is representative of the ATM environment at the cell level. Using spectral decomposition and asymptotic analysis, he showed that as the mean on period of a source increases, given that its average and peak rate remain constant, the tail behavior of the distribution of the number of cells queued in the multiplexer has a simple characterization.

Consider an infinite capacity queue with constant service times and a Markovian cell arrival process governed by the probability generating matrix $\boldsymbol{P}(z)$. This queueing system has an embedded Markov chain of an $M/G/1$ type and can be fully solved by the techniques introduced by

Neuts [18]. However, the full solution is not a trivial task when the size of the arrival process is large.

It is known that the steady-state queue length distribution exhibits a tail behaviors characterized by z^* the smallest root outside the unit circle of the determinant $|zI - P(z)|$. For sufficiently large i, we have:

$$Pr(\text{queue length} > i) \approx \alpha(\frac{1}{z^*})^i,$$

where α is a constant that is, in general, difficult to obtain.

It can be shown that z^* is the solution of the equation $z = \lambda_{PF}(z)$, where $\lambda_{PF}(z)$ is the Perron-Frobenius eigenvalue of $P(z)$. The following approximation for z^* was suggested [20] for the case when an input source is described by the tuple (T_i, r_i, b_i) where source i emits cells every T_i slots during the on period, and r_i respectively b_i are the mean rate and mean length of the on period of source i. The peak rate R_i is equal to $1/T_i$.

$$z^* \approx 1 + \frac{1-\rho}{\sum_{i=1}^{N} r_i(1 - r_i/R_i)^2 b_i} \tag{6}$$

where $\rho = \sum_{i=1}^{N} r_i$ and N is the number of input sources. The queue length distribution is approximated by the probabilities:

$$Pr(\text{queue length} > i) \approx \rho(\frac{1}{z^*})^i,$$

where ρ, the traffic intensity, is used as an approximation for the value of the unknown constant α. A new call is accepted if $\sum R_i < 1$ or if the resulting $\rho(\frac{1}{z^*})^B < \epsilon$. Sohraby [21] also suggested a real-time version of the algorithm which takes into consideration the effect of adding or removing a call.

6.3 Validation of the Call Admission Control Methods

In order to validate the CAC methods, we compare their admission regions with that obtained by approximate numerical solution of the multiplexer model. Assuming a 2-state Markov modulated on/off model for the input sources, we use the numerical solution method presented in section 3 for approximately analyzing the multiplexer. For a single-server multiplexer, the computational complexity of this numerical method is $O(B \times N^3)$ and the storage is $O(B \times N^2)$, where B is the buffer size and N the total number of input sources. This limits the applicability of the numerical method to about 80 sources from all traffic classes and to buffer sizes not greater than about 400. The total maximum number of sources can be limited by appropriately selecting the values of the mean rates r_i of each class.

It should be noted that the traffic model used in our numerical model for the analysis of the multiplexer is an IBP. In this model, we let the probability of a cell arrival at any time slot in the on period be equal to R. Such an IBP source exhibits more burstiness than a 2-state Markov modulated source where arrivals occurring periodically. Therefore, the analysis of a statistical multiplexer with IBP sources provides an upper bound of the cell loss probability when compared with a multiplexer with input sources with periodic arrivals. This implies that the achievable admission region, when using an IBP as the source model, would be more conservative. For the special case when the peak rate is equal to 1, the IBP model and the periodic arrivals model are equivalent. In this case, we are in a more appropriate position to validate the accuracy of the CAC methods.

6.3.1 Case 1: Peak Rate < 1

We consider a multiplexer with two traffic classes. The values of the system parameters were chosen as follows. We set the required cell loss probability ϵ equal to 10^{-6}, the buffer size B equal to 40, and let class 1 traffic be characterized by $TD1 = (0.1, 0.02, 2)$, and class 2 traffic by $TD2 = (0.1, 0.04, 3)$. The minimum and maximum calls for classes 1 and 2 are (10,50) and (10,25) respectively. The admission region obtained by each of the CAC methods in addition to the region obtained by the numerical solution of the multiplexer model are shown in figure 8. The maximum number of admitted class 1 and class 2 calls, when no calls from the other class are transported, are given by: (46,23), (28,18), and (41,20) for for the equivalent capacity, heavy traffic approximation, and numerical solution methods respectively.

Although the equivalent capacity method gives an admission region that is larger than that of the numerical solution method, it is clearly the closest of the methods to the numerical solution. The question here is how accurate is the equivalent capacity method because it admits 5 more class 1 calls and 3 more class 2 calls than the number of calls from respectively class 1 and class 2 admitted by the numerical solution of the multiplexer. Firstly, as discussed above, the numerical solution gives an upper bound with respect to the cell loss probability. Secondly, the relative error between the equivalent capacity method and the numerical solution is only 9.5% and 15% for the maximum number of admitted class 1 and class 2 calls respectively. Thirdly, the equivalent capacity method actually uses an upper bound of the cell loss probability which is supposed to

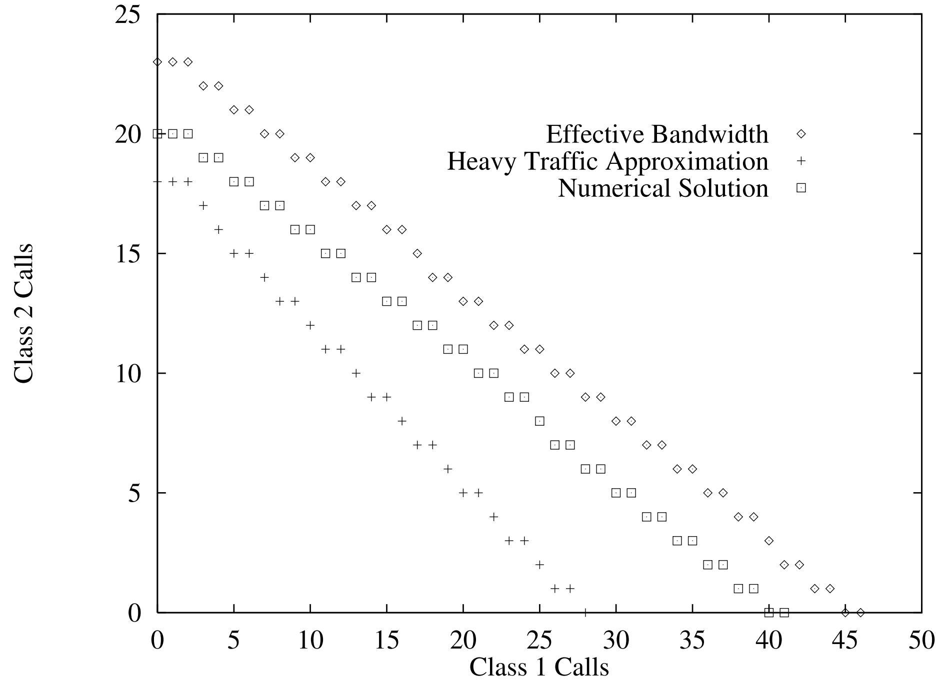


Figure 8: Admission regions for the CAC methods as compared with numerical solution

make the admission region more conservative. Hence, we can argue that in this example, the equivalent capacity gives good results.

The heavy traffic approximation method provides an admission region smaller than that of the numerical solution method and thus causes loss in possible achievable statistical gain. However, the required cell loss probability will be achieved since the heavy traffic admission region is smaller than that of the numerical solution.

6.3.2 Effect of the Buffer Size

We consider a multiplexer with a single class of calls whose traffic is characterized by $(0.1, 0.02, 3.0)$. The maximum number of calls admitted by the CAC methods and the numerical solution was obtained for buffer sizes ranging from 25 to 200. The results are shown in figure 9. As the buffer size increases, the maximum number of calls admitted by the numerical solution, the equivalent capacity, and the heavy traffic approximation methods, all approach the maximum allowable number of 50 calls. Also, considering the numerical solution as a reference, the relative error in the maximum number of admitted calls for both the equivalent capacity and heavy traffic approximation methods decreases as the buffer size increases.

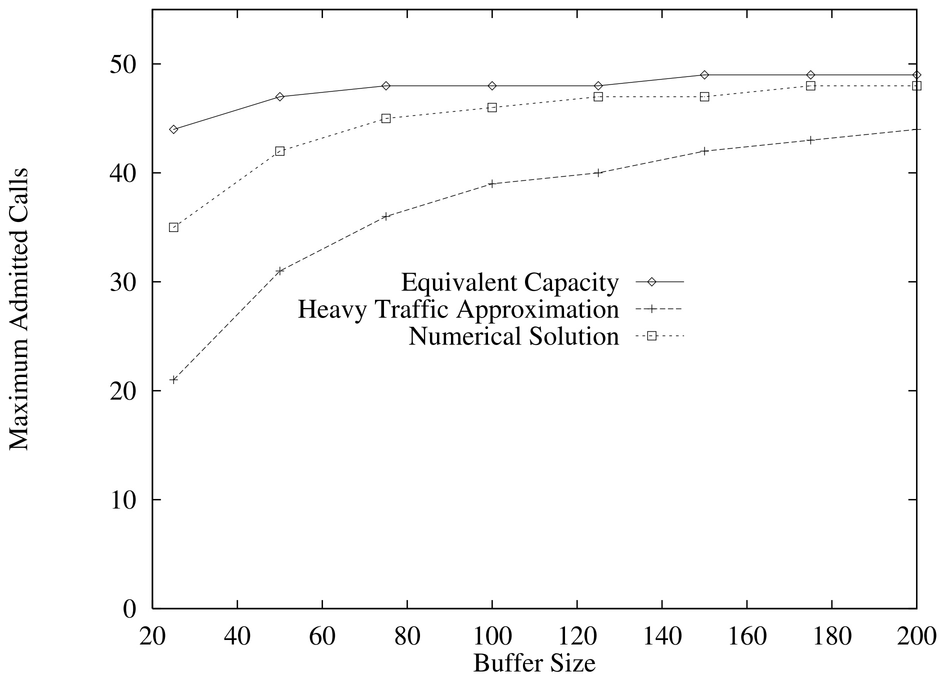


Figure 9: Variation of the maximum number of class 1 calls with buffer size – required cell loss probability = 10^{-6}

6.4 Case 2: Peak Rate < 1

We consider a multiplexer with a single traffic class. The values of the system parameters were chosen as follows. We set the required cell loss probability ϵ equal to 10^{-6}, the buffer size B equal to 40, and let the input traffic be characterized by $TD = (0.2, 0.02, 50.0)$. The minimum and maximum calls (5,50). We vary the input buffer size from 20 to 200 and record the maximum number of calls that can be allocated by the cac methods and the numerical solution of the multiplexer. The results are shown in figure 10. None of the cac methods peform satisfactorily here. Consider for example when the buffer size is 100. The maximum calls admitted by the equivalent capacity method, the heavy traffic approximation method, and the numerical solution method are 12, 8, and 20 respectively. There a large capacity that can go wasted here: the loss in the achievalble utiliztion for the equivalent capacity and heavy traffic approximation methods are 40% and 60% respectively.

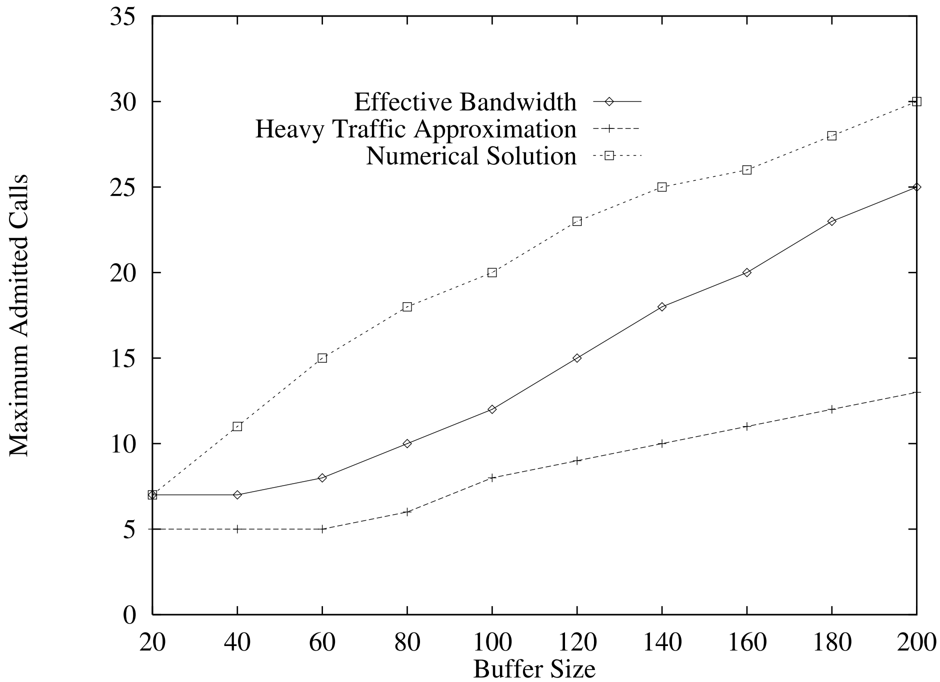
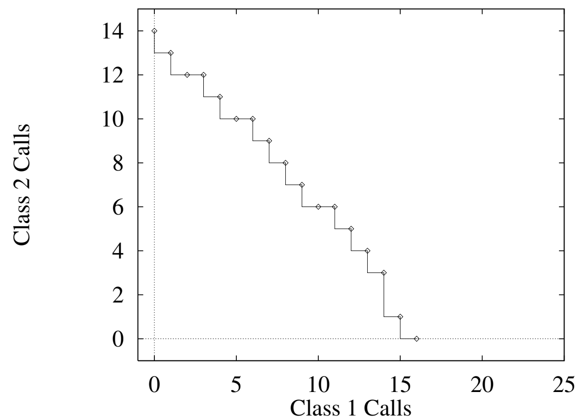


Figure 10: Variation of the maximum number of calls with buffer size – required cell loss probability = 10^{-6}

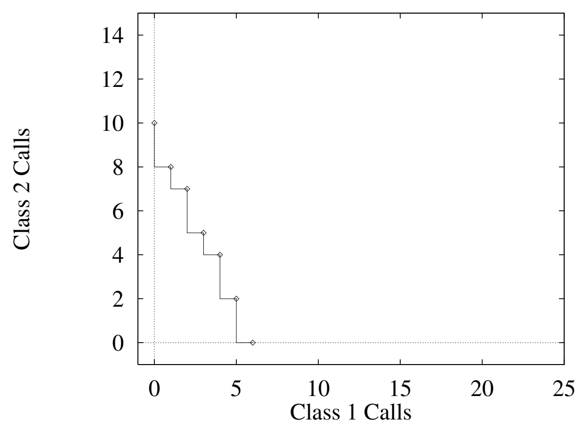
6.4.1 Case 3: Peak Rate = 1

We now study the performance of the methods when the peak rate of calls is equal to one. The values of the system parameters were chosen as follows. We set the required cell loss probability ϵ equal to 10^{-6}, the buffer size B equal to 40, class 1 traffic is characterized by $TD1 = (1.0, 0.02, 3)$ and class 2 traffic by $TD2 = (1.0, 0.04, 2)$. The minimum and maximum calls for classes 1 and 2 are (1,50) and (1,25) respectively. The admission region obtained by each of the CAC methods in addition to the region obtained by the numerical solution of the multiplexer model are shown in figure 11.

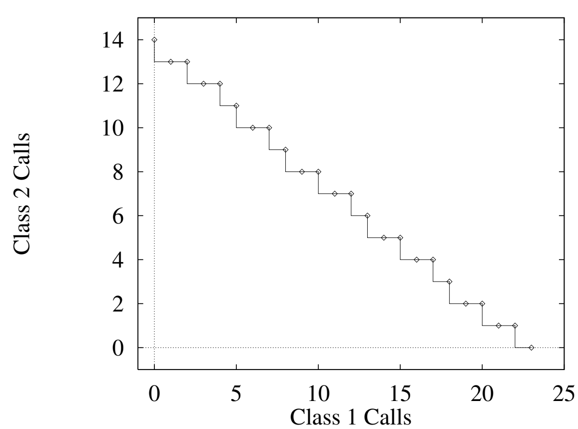
The maximum number of admitted class 1, and class 2 calls, when no calls from the other class are transported, are given by: (6,10), (23,14), and (16,14) for the equivalent capacity, the heavy traffic approximation, and the numerical solution methods respectively. The largest admission region is given by the heavy traffic approximation method. Since for the case when peak rate is equal to 1, the traffic model used by the heavy traffic approximation and the numerical solution methods are the same, we can confidently conclude that the heavy traffic method is over-admitting sources and that the required cell loss probability may not be achieved. On the other hand, the equivalent capacity method is very conservative. The maximum number of class 1 calls is 6 while



(a) Numerical Solution



(b) Equivalent Capacity



(c) Heavy Traffic Approximation

Figure 11: Admission regions for the various CAC methods as compared with numerical solution

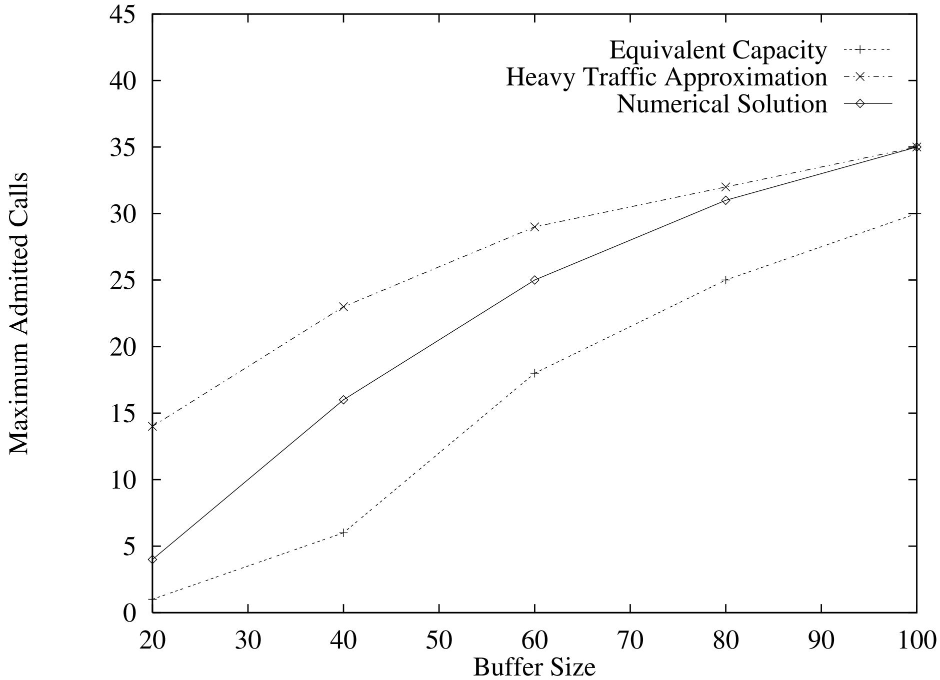


Figure 12: Variation of the maximum number of class 1 calls with buffer size – required cell loss probability = 10^{-6}

---

it can be as high as 16. So, we are losing 62.5% of the achievable statistical gain for class 1. The same is true, but to a less extent, to class 2.

6.4.2 Effect of the Buffer Size

We consider a multiplexer with only class 1 calls as in the previous subsection. The maximum number of calls admitted by the CAC methods and the numerical solution was obtained for buffer sizes ranging from 20 to 100. The results are shown in figure 12.

The accuracy of the heavy traffic approximation method improves as the buffer size increases. The difference between the allocation of the heavy traffic method and that of the numerical solution procedure shrinks as the buffer size increases, and it actually becomes zero when $B = 100$. The performance of the equivalent capacity method is still conservative. As the buffer size increases, the difference between the number of admitted calls by the equivalent capacity method and the numerical solution method approaches a constant value. However, this implies that the relative error gets better when buffer size increases since the maximum number of admitted calls effectively increases.

7 The Sporadic Sources Case

A sporadic source is charactgerized by geometrically distrubuted on and off periods. During the on period, arrivals occur in a deterministic fashion every $T \geq 1$ slots. For our purposes, we assume that the first arrival in an arbitrary on period occurs at the first slot of that period. A sporadic sourec can be characterized by the triplet (α, β, T). Note that the only difference between a sporadic source and an IBP source is the fashion at which arrivals occur during the on period.

The analysis of a statistical multiplexer with sporadic sources is a difficult problem. The numerical complexity of the analysis is proportional to T. When the sources are heterogeneous, the complexity increases dramatically. Kröner [15] provided exact and approximate methods for analyzing a finite buffer ATM multiplexer with a mixture of CBR and sporadic sources. The exact analysis can be used to handle the heterogeneous sources case but the computational complexity becomes enormous. An approximate method based on observing the system state each $k > 1$ time slots instead of at each time slot was devised. Stavrakakis [23] considered a statistical multiplexer fed by homogeneous sporadic sources. The peak arrival rate of a source was assumed to be smaller then the output link speed. The source was modeled as a Markov chain with artificially introduced states to model the periodic arrivals. The whole system was studied as a discrete-time Markov chain. The approach is limited since it is restricted to homogeneous sources and is computationally very intensive. In this section we provide two relatively inexpensive approximations for the analysis of a statistical multiplexer with homogeneous sporadic sources.

7.1 The first approximation method (upper bound)

The first approximation is done by approximating the spoaradic source (α, β, T) by an IBP source with (α, β, γ), where $\gamma = 1/T$. An IBP source is known to provide an upper bound for performance measures, such as the cell loss probability and the mean number of cells, when the actual input is comprised of sporadic sources. This is because an IBP source exhibits more burstiness than a sporadic source with the same mean on and off periods and throughput.

7.2 The second approximation method (lower bound)

The second approximation basically provides a lower bound for the required performance measures. This approximation also uses an IBP as the source model to substitute for the sporadic source. After the superposition process of N IBP sources is obtained, it is smoothed by modifying the batch probability distribution at each state. This provides a superposition process which may well be less bursty than the actual superosition of the sporadic sources.

Smoothing of the Distribution of Batch Arrivals: Consider a probability density function $\{b_i\}, i = 0, 1, \cdots, K$ which represents the distribution of batch arrivals at a generic state of the superposition process at any stage of the aggregation algorithm. Let $\bar{b} = \sum_{i=0}^{K} ib_i$ and $k = \lceil \bar{b} \rceil$. The smoothed batch distribution is given by:

$$\tilde{b}_i = \begin{cases} 1 - \bar{b}/k, & i = 0 \\ \bar{b}/k, & i = k \\ 0, & \text{otherwise} \end{cases}$$

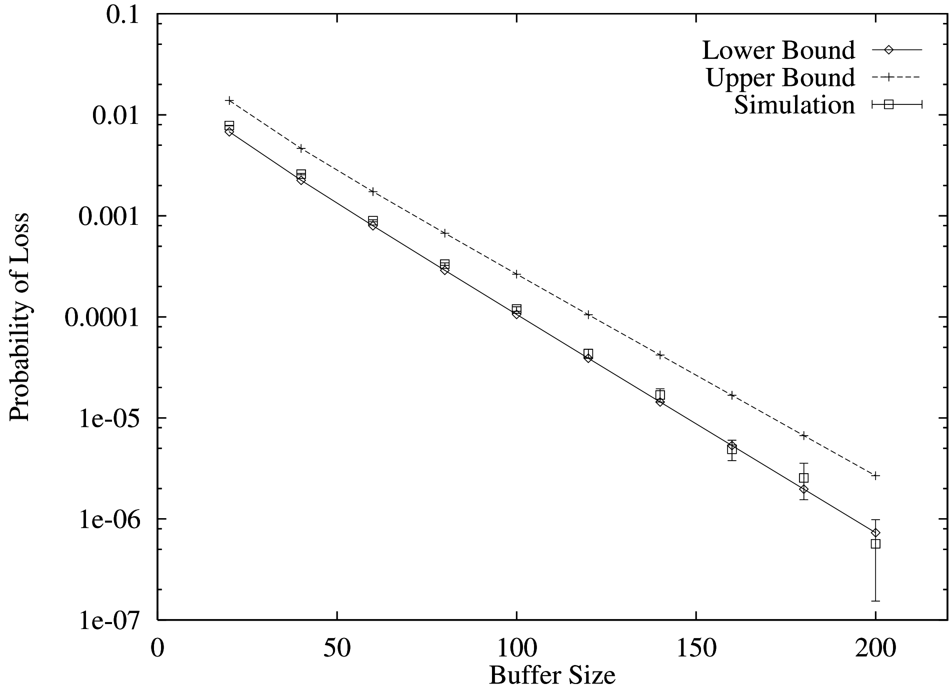
This smoothing of the distribution prserves the mean of the distribution while it reduces the squared coefficient of variation of the batch. For a superposition process of N sources, the smoothing is applied to $b(n, v)$, $0 \leq n \leq N$ and $0 \leq v \leq n$.

In the following section we present results for testing the accuracy of the two approximation methods. The results indicate that the the first method basically provdies an upper bound while the second provides a lower bound.

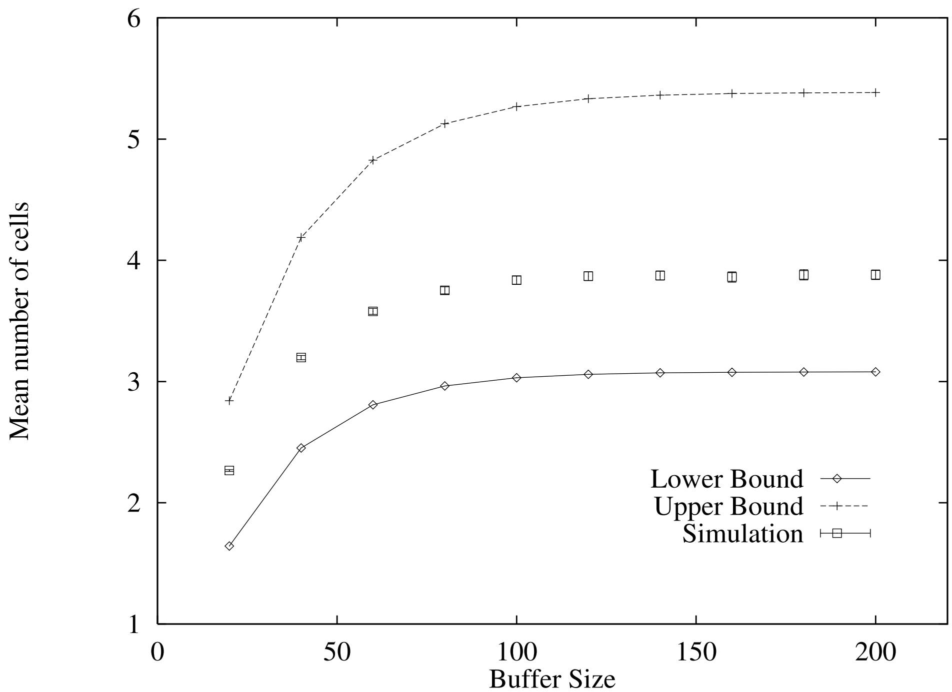
7.3 Validation of the Sporadic Sources Approximations

In this section we provide three numerical examples and discuss their siginificance on the proposed approximation method for the sporadic sources case. In all the examples below, the queue length distribution and the cell loss probability for the multiplexer with sporadic sources is evaluated by means of a discrete-event simulation program with 95% confidence intervals.

Example 1: In this example we have a single group of homogeneous sporadic sources. A source is described by $(0.99, 0.9966666667, 10)$ which gives a a mean on period of 100, a mean off period of 300 and a source throughput of 0.025. We let the total number of source be equal to 30 for a total utilization of 0.75. We evaluate the queue length distribution of the sporadic sources multiplexer using simulation and compare it with the two approximating methods for buffer sizes ranging from 20 to 200. The results for the cell loss probability and the mean number of cells are shown in figure 13. As can be seen from the figure, the lower bound is very close to the actual values.



(a)



(b)

Figure 13: Cell loss probability and mean number of cells vs. buffer size

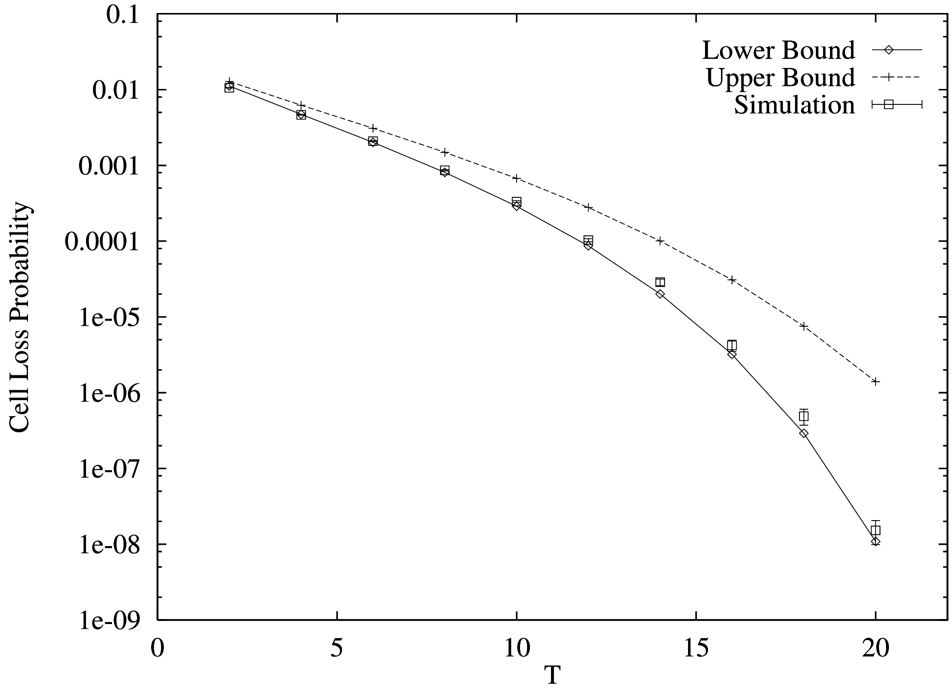
The upper bound method is far from being accurate, however, in case we are solving a buffer dimensioning problem or a call admission control problem, we must take the results from the upper bound into account. We note here that the larger the value of T, the worse the upper bound calculation is. This will be further investigated in the next example.

Example 2: We have a single group of sources with $N = 30$ and let the buffer size be 80. The source throughput r is fixed at 0.025 and the mean number of cells in a burst is fixed to 10. Note that the mean number of cells in a burst is equal to $(\text{mean on period}/T)$. We vary T from 2 to 20 and obtain the parameters α and β of the source from the two invariants r and $(\text{mean on period}/T)$. We evaluate the queue length distribution and the cell loss probability of the multiplexer with sporadic sources using simulation and the two proposed approximations. The results for the cell loss probability and the mean number of cells are shown in figure 13. For the cell loss probaility, we can observe that the lower bound approximation is closer to the simulated cell loss probability than the upper bound approximation. The accuracy of the upper bound method deteriorates with the increase of T. For the mean number of cells, a good approximation would be to take the average of the upper bound and lower bound. We not here that for the mean number of cells, the accuracy of both the upper bound and lower bound deteriorates as T increases.

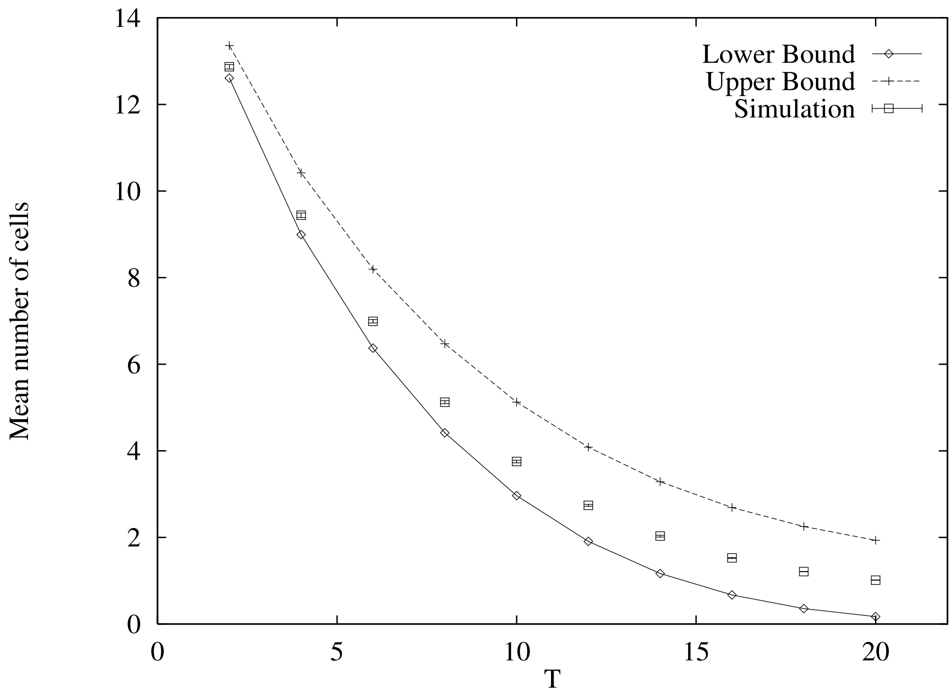
8 Conclusions

A computationally efficient algorithm for characterizing the superposition process of multiple possibly-heterogeneous groups of IBP sources was introduced. The accuracy of the algorithm depends to a large extent on the degree of heterogeneity of the sources. We found out that a good measure of the heterogeneity of the sources is the ratio of the largest to the smallest coefficient of variation of the inter-arrival times. This ratio can be used to forecast the accuracy of the algorithm without the need to compare with simulation. A good rule of thumb is to use the algorithm for cases when this ratio is in the range $[0, 2.5]$. This, however, still needs further testing for large buffer sizes and for various values of the ratio of the buffer size to the largest average on period of the sources (note that this ratio plays a major role in affecting system performance).

We have carried out a numerical study for validating the equivalent capacity and heavy traffic approximation methods for call admission control. Our results indicate a favorable performance for the equivalent capacity method. However, we have observed frequent cases where the allocation of the equivalent capacity method is significantly lower than what can actually be achieved.



(a)



(b)

Figure 14: Cell loss probability and mean number of cells vs. T

A careful study of the regions where the efficiency of the equivalent capacity method is low would be very useful.

An interesting study would be to compare our method with other approaches that characterize the superposition process as a 2-state MMPP, e.g. the work in [3, 17, 26]. Also, a study of an intelligent aggregation scheme would be very desirable since we have seen that aggregation actually distorts the original correlation structure of the superposition process.

References

[1] S. L. Albin. Approximating a Point Process by a Renewal Process, 2: Superposition of Arrival Processes to Queues. *Operations Research*, 32:1133–1162, 1984.

[2] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic Theory of a Data-Handling System with Multiple Sources. *Bell Sys. Tech. J.*, 61:1871–1894, 1982.

[3] A. Baiocchi, N. Bléfari-Melazzi, M. Listani, A. Roveri, and R. Winkler. Loss Performance Analysis of an ATM Multiplexer Loaded with High-Speed On-Off Sources. *IEEE Journal on Selected Areas in Communications*, 9:388–393, 1991.

[4] C. Blondia and O. Casals. Performance Analysis of Statistical Multiplexing of VBR Sources. In *Proceedings of the IEEE Infocom*, pages 828–838, 1992.

[5] A. I. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE Transactions on Networking*, 1:329–343, 1993.

[6] A. I. Elwalid, D. Mitra, and T. E. Stern. Statistical Multiplexing of Markov Modulated Sources: Theory and Computational Algorithms. In *Proceedings of 13th International Teletraffic Congress*, pages 495–500, Copenhagen, June 1991.

[7] ATM Forum. ATM User-Network Interface Specification, Version 2.2, June 1991.

[8] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks. *IEEE Journal on Selected Areas in Communications*, 9:968–981, 1991.

[9] H. Heffes. A Class of Data Traffic Process: Covariance Function Characterization and Related Queueing Results. *Bell Sys. Tech. J.*, 59:897–929, 1980.

[10] H. Heffes and D. M. Lucantoni. A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. *IEEE Journal on Selected Areas in Communications*, 4:856–867, 1986.

[11] S. Hong, H. G. Perros, and H. Yamashita. A Discrete-Time Queueing Model of the Shared Buffer ATM Switch with Bursty Arrivals. *Telecommunication Systems*, 2:1–20, 1993.

[12] J. Hui. Resource Allocation for Broadband Networks. *IEEE Journal on Selected Areas in Communications*, 6:1598–1608, 1988.

[13] J. Hunter. *Mathematical Techniques of Applied Probability. Volume 2: Discrete Time Models: Techniques and Applications*, chapter 9. Academic Press, 1983.

[14] 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.

[15] H. Kröner. Statistical Multiplexing of Sporadic Sources - Exact and Approximate Performance Analysis. In *Proceedings of 13th International Teletraffic Congress (ITC)*, pages 787–793, 1991.

[16] I. I. Makhamreh, D. McDonald, and N. D. Georganas. Approximate Analysis of a Packet Switch with Finite Output Buffering and Imbalanced Correlated Traffic . In *International Conference on Communications (ICC)*, 1994.

[17] R. Nagarajan, J. F. Kurose, and D. Towsley. Approximation Techniques for Computing Packet Loss in Finite-Buffered Voice Multiplexers. *IEEE Journal on Selected Areas in Communications*, 9:368–377, 1991.

[18] M. F. Neuts. *Structured Stochastic Matrices of the M/G/1 Type and Their Applications*. Marcel Dekker Inc., New York, 1989.

[19] H. G. Perros and R. Onvural. On the Superposition of Arrival Processes for Voice and Data. In *Fourth International Conference on Data Communication Systems and Their Performance*, pages 341–357, Barcelona, June 1990.

[20] K. Sohraby. On the Asymptotic Behavior of Heterogeneous Statistical Multiplexer with Applications. In *Proceedings of the IEEE Infocom*, pages 839–847, 1991.

[21] K. Sohraby. On the Theory of General On-Off Sources With Applications in High-Speed Networks. In *Proceedings of the IEEE Infocom*, pages 401–410, 1993.

[22] K. Sriram and W. Whitt. Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data. *IEEE Journal on Selected Areas in Communications*, 6:833–846, 1986.

[23] I. Stavrakakis. Statistical Multiplexing of Correlated Slow Traffic Sources. In *Proceedings of Globecom*, pages 1540–1544, 1992.

[24] T. E. Stern and A. I. Elwalid. Analysis of Separable Markov-Modulated Rate Models for Information-Handling Systems. *Adv. Appl. Prob.*, 23:105–139, 1991.

[25] R. C. F. Tucker. Accurate Method for Analysis of a Packet-Speech Multiplexer with Limited Delay. *IEEE Transactions on Communications*, 36:479–483, 1988.

[26] S. S. Wang and J. A. Silvester. A Fast Performance Model for Real-Time Multimedia Communication. In *Proceedings of the Fifth International Conference on Data Communication Systems and Their Performance*, Raleigh, NC, October 1993.

[27] W. Whitt. Approximation of a Point Process by a Renewal Process, 1: Two Basic Methods. *Operations Research*, 30:125–147, 1982.