

EFFICIENT SIMULATION OF CELL LOSS PROBABILITY USING STANDARDIZED ATM CONNECTION TRAFFIC DESCRIPTORS

JAMES A. FREEBERSYSER¹ J. KEITH TOWNSEND

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, North Carolina 27695-7914

ABSTRACT

In ATM networks, the purpose of connection admission control is to decide if allowing a new connection into the network violates a quality of service measure, such as the cell loss probability. Testing the algorithms that perform connection admission control is difficult because of the complexity of the ATM switch architectures, the low cell loss probabilities required by ATM networks, and the unwieldiness of matching statistical models of the traffic entering the network to the traffic descriptors used by the call admission control algorithm. In this paper, instead of using statistical traffic models, we describe the traffic entering the network by the standardized ATM connection traffic descriptors used by the connection admission control algorithms. We develop a simulation framework for estimating the cell loss probability due to buffer overflow, use importance sampling to increase the efficiency of the simulation, and find an analytical solution for the improvement.

For the experimental examples considered here, we obtained 3 to 9 orders of magnitude improvement in efficiency compared to conventional Monte Carlo simulation. We also derive upper and lower bounds on the cell loss probability that can be used to determine if this method can be applied effectively to a given system before simulation. The efficient simulation method developed here is suitable for the design and testing of the switch architectures and connection admission control algorithms planned for use in ATM networks.²

¹James A. Freebersyser is supported by the U.S. Air Force Palace Knight Program.

²This work was supported in part by the Center for Communications and Signal Processing, North Carolina State University.

1 Introduction

The purpose of Connection Admission Control (CAC) algorithms in Asynchronous Transfer Mode (ATM) networks is to decide in real time whether a new connection can be admitted into the network without violating Quality of Service (QoS) measures, such as the cell loss probability, for the new connection or the existing connections. The decisions made by the CAC algorithm in implementing a given CAC policy are specific to each ATM switch design, and are also used in the negotiation of the traffic descriptors and QoS measures that result in the traffic contract between the user and the network [1]. The complexity of the ATM switch architectures, the low cell loss probabilities required by ATM networks (10^{-6} to 10^{-12}), and the unwieldiness of matching statistical traffic models to the traffic descriptors used by the CAC algorithm combine to make testing of CAC algorithms difficult.

In many cases, the complex ATM switch architectures are simplified so that analysis can be used to develop real time algorithms that approximate performance measures. These approximations are then used to make CAC decisions. However, some method of testing CAC algorithms developed in this manner is necessary to verify that the resulting CAC decisions are correct.

Monte Carlo (MC) simulation can be used to obtain accurate estimates of performance measures of complex ATM switch architectures but is too slow to be used for CAC in real time. Even in non-real time environments, such as using MC simulation to test a CAC algorithm, MC simulation can be too slow for very low cell loss probabilities because of the long run times required to obtain accurate estimates. Importance Sampling (IS) is a technique that, under the proper conditions, can speed up simulations involving rare events of network (queueing) systems (see for example, [2]-[6]). By carefully choosing the modification or bias of the underlying probability measures, large speed-up factors in simulation run time can be obtained.

Most commonly, statistical traffic models have been used in the analysis and MC simulation of ATM switches and networks. Several generations of statistical models have been developed as more insight has been gained into the burstiness and correlation of ATM traffic.

These statistical traffic models have become increasingly accurate, but they have also become more complicated and therefore less tractable to analysis. In addition, it is difficult to establish a relationship between the parameters of the statistical models of the ATM traffic and the traffic descriptors used by the CAC algorithm. [7]

To overcome this problem, we characterize the traffic entering the ATM network by the ATM connection traffic descriptors standardized by the ATM Forum [1]. This approach, where the traffic descriptors used by the CAC algorithm are also used in the analysis of a performance measure, is called the operational approach (versus the statistical approach) in [8]. The operational approach has been previously considered in [8]-[12], where all of the methods cited used analysis to determine approximations or upper bounds of the cell loss probability. However, the low cell loss probabilities which are of interest were not determined for any of the methods cited, and the traffic descriptors standardized by the ATM Forum were not used.

In this paper, we first describe a MC simulation framework using the standardized ATM connection traffic descriptors to characterize the traffic entering the network. We then show how IS is applied to this MC simulation framework. For the experimental systems we consider, our method results in 3 to 9 orders of magnitude increase in efficiency compared to the conventional MC simulation used to estimate the cell loss probability due to buffer overflow. We also derive upper and lower bounds on the cell loss probability that can be used to determine whether applying this method to a particular system will be successful. Our method of efficient simulation is applicable to the design and testing of ATM switches, as well as the CAC algorithms planned for use with these ATM switches.

2 System Description

In this section, we first consider the ATM connection traffic descriptors standardized by the ATM Forum that will be used to characterize the traffic entering the network. We then examine the ATM switch architecture considered and the system model.

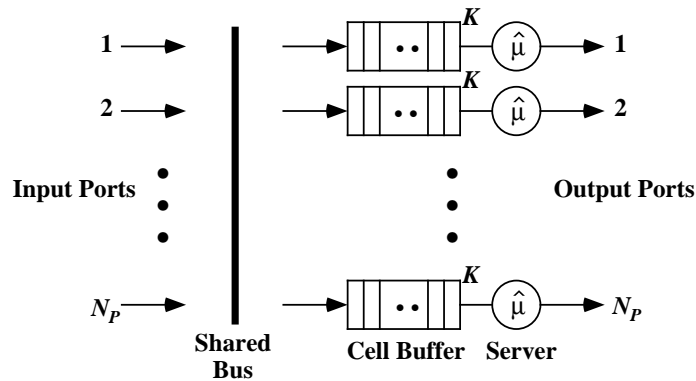


Figure 1: ATM switch architecture.

2.1 ATM Connection Traffic Descriptors

We characterize a connection requesting admission into the ATM network by the following ATM connection traffic descriptors standardized by the ATM Forum User-Network Interface Specification [1]:

$$\hat{\lambda}, \text{ peak cell rate (PCR) in bps} \quad (1)$$

$$\bar{\lambda}, \text{ sustained (average) cell rate (SCR) in bps} \quad (2)$$

$$\hat{B}, \text{ maximum burst length in cells at the peak cell rate} \quad (3)$$

The $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ -triplet contains optional traffic connection descriptors which can be used to derive the remaining required traffic connection descriptors that define the traffic contract, specifically the cell delay variation and burst tolerance [1]. Each connection, which may be an individual source or multiplexed sources, negotiates with the CAC algorithm for admission into the network using these descriptors to describe the traffic.

2.2 ATM Switch Architecture

The model for the ATM switch architecture considered here is shown in Figure 1. The ATM switch has N_P input ports, and each connection in the network routed by the shared bus to the output cell buffer of one of the N_P output ports is described by the same $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ -triplet. The shared bus operates at a speed of $N_P \hat{\lambda}$ so that all N_P input ports are given access to

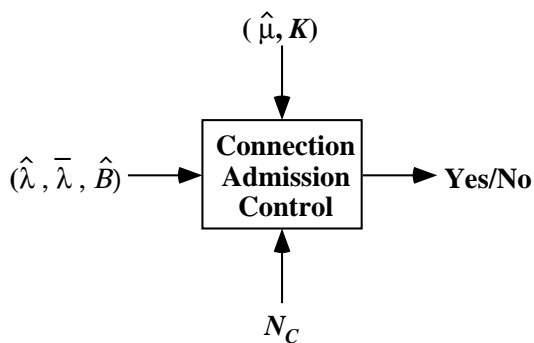


Figure 2: Connection admission control block diagram.

the shared bus during an arrival slot. The output cell buffers use a first-in, first-out (FIFO) queueing discipline. Each output cell buffer has a finite size of K cells, and cells exit the output cell buffer at a service rate of $\hat{\mu}$ bps.

2.3 System Model

Our goal is to determine the cell loss probability due to buffer overflow resulting from the worst-case conforming traffic conditions for a given set of ATM connection traffic descriptors as a function of the number of connections in the network. We consider only homogeneous connection traffic in the network, that is, all N_C connections in the network which are being routed through one of the N_P output cell buffers of size K and service rate $\hat{\mu}$ have identical $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ -triplets. The block diagram of the CAC algorithm for these conditions is shown in Figure 2. We assume ergodicity in the traffic flow routed through each of the N_P output cell buffers, so that by examining one output cell buffer we obtain the performance characteristic of all the N_P output cell buffers, and that the output cell buffer is initially empty, since the steady-state behavior of the system is equivalent whether it is reached by starting $N + 1$ connections with an empty output cell buffer, or by adding an additional connection to the existing N connections and a non-empty output cell buffer.

The worst-case arrival condition which is still compliant according to the Usage Parameter Control (UPC) is when the traffic enters the network at the peak rate, $\hat{\lambda}$, such that the maximum burst length of \hat{B} cells is generated. The CAC algorithm must consider this worst-

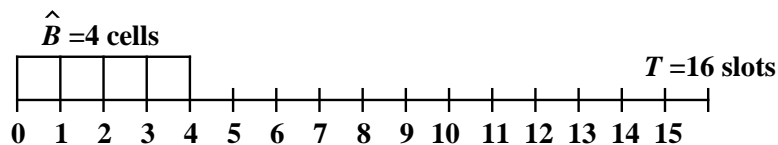


Figure 3: Busy/idle cycle example.

case arrival condition in determining whether or not a QoS measure for cell loss probability is violated by admitting a new connection to the network.

We operate with a slotted-time system where an arrival slot is the amount of time that one cell takes to arrive. The period of the busy/idle cycle of the arriving traffic, T , is the time in slots that the \hat{B} length cell burst at the peak rate must be averaged over so that the sustained rate, $\bar{\lambda}$, is not exceeded, thus

$$T = \hat{B}\hat{\lambda}/\bar{\lambda}, \text{ slots} \quad (4)$$

An example of a burst of length $\hat{B} = 4$ cells is shown in Figure 3. The randomness in the system model described here is found in the connection starting-slot position. A connection can start in any slot within the T length period, so we describe the connection starting-slot position by a uniform probability density.

We let the number of ports on the ATM switch be equal to the number of connections routed through the output cell buffer, $N_P = N_C$, since this condition results in an upper bound of the cell loss probability for any size switch, in terms of the number of input and output ports. The data rate of the shared bus is then N_C cells per arrival slot, as shown in Figure 4, so that all N_C input ports have access to the shared bus in each arrival slot, which results in N_C service slots per arrival slot. When $N < N_C$ input ports are active, we make the worst-case assumption that the cells of the active ports occupy the first N of N_C service slots, rather than operating with some type of round-robin bus access mechanism. We assume that each connection occupies every service slot position with equal probability for each output cell buffer, so that the the cell loss probability for any one connection is equal to the cell loss probability for all connections.

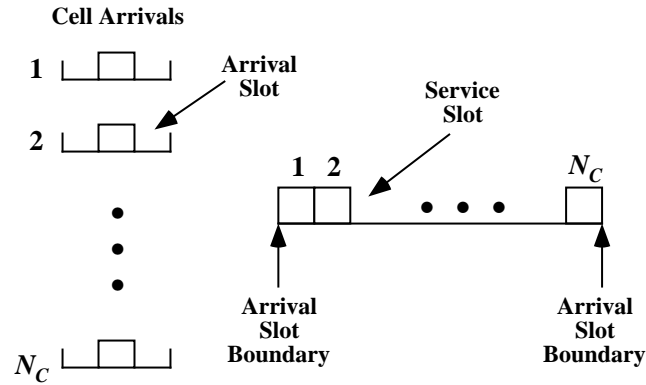


Figure 4: Relationship between arrival slots and service slots.

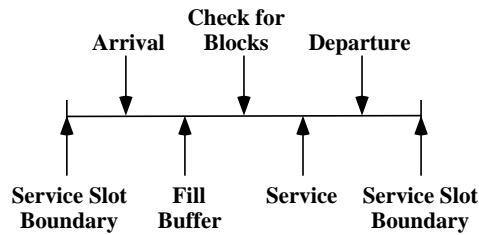


Figure 5: Order of events within a service slot.

The effective service rate, μ , of the server is

$$\mu = \hat{\mu} / \hat{\lambda}, \text{ cells per arrival slot} \quad (5)$$

In each service slot, one cell can be loaded into the output cell buffer if there are arriving cells and μ/N_C cells can be removed from the output cell buffer if the output cell buffer contains greater than μ/N_C cells. If the output cell buffer contains less than μ/N_C cells, the available cells in the output cell buffer are serviced. If there are no cell arrivals, cells in the output cell buffer are still removed according to this policy. Figure 5 shows the order of events within a service slot.

While the ATM switch architecture considered here may seem simplistic, more complicated ATM switch architectures can be considered using this methodology. The main feature of this system model is that the traffic entering the network is defined in terms of the standardized ATM connection traffic descriptors.

3 Closed-Form Analysis of Cell Loss Probability

Consider the set V of all unique connection starting-slot vectors \underline{v}_j for $j = 1 \dots |V|$. A connection starting-slot vector \underline{v}_j is composed of N_C elements, where the l -th element in the vector represents the slot position, from slot 1 to slot T , where the l -th connection starts its \hat{B} length cell burst. Each connection starting-slot vector \underline{v}_j corresponds to the loss of i cells and the arrival of k cells in a steady-state period T_{SS} . Let n_{CL_i} be the total number of connection starting-slot vectors \underline{v}_j that map to i cell losses in a steady-state period T_{SS} , where $|V| = \sum_{i=0}^{L_{\max}} n_{CL_i}$ and L_{\max} is the maximum number of cell losses in a steady-state period T_{SS} . The probability of i cell losses in a steady-state period T_{SS} is described by a multinomial probability distribution with individual probabilities

$$p_{CL_i} = n_{CL_i}/|V| \quad (6)$$

where $\sum_{i=0}^{L_{\max}} p_{CL_i} = 1$.

In a similar fashion, n_{CA_k} is the total number of connection starting-slot vectors \underline{v}_j that map to k cells arriving in a steady-state period T_{SS} . Thus, $|V| = \sum_{k=0}^{A_{\max}} n_{CA_k}$ where A_{\max} is the maximum number of cells that can arrive in a steady-state period T_{SS} . The probability of k cells arriving in a steady-state period T_{SS} is described by a multinomial probability distribution with individual probabilities

$$p_{CA_k} = n_{CA_k}/|V| \quad (7)$$

where $\sum_{k=0}^{A_{\max}} p_{CA_k} = 1$.

Define the cell loss ratio (CLR) or cell loss probability, P_{CL} , as

$$P_{CL} = \frac{\text{Total Number of Cell Losses}}{\text{Total Number of Arriving Cells}} = \frac{\sum_{i=0}^{L_{\max}} i n_{CL_i}}{\sum_{k=0}^{A_{\max}} k n_{CA_k}} \quad (8)$$

We can also define P_{CL} in terms of the average number of cell losses and the average number of cells arrived, \bar{n}_{CL} and \bar{n}_{CA} , respectively.

$$P_{CL} = \frac{\frac{1}{|V|} \sum_{i=0}^{L_{\max}} i n_{CL_i}}{\frac{1}{|V|} \sum_{k=0}^{A_{\max}} k n_{CA_k}} = \frac{\bar{n}_{CL}}{\bar{n}_{CA}} \quad (9)$$

For the systems we consider here, we know that \bar{n}_{CA} is constant for a given $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ -triplet, so we rewrite P_{CL} as

$$P_{CL} = \frac{\bar{n}_{CL}}{\bar{n}_{CA}} = \frac{\sum_{i=0}^{L_{\max}} i p_{CL_i}}{\bar{n}_{CA}} = \sum_{i=0}^{L_{\max}} \left(\frac{i}{\bar{n}_{CA}} \right) p_{CL_i} \quad (10)$$

Thus, P_{CL} is a weighted sum of probabilities from the multinomial distribution of cell loss.

The variance of P_{CL} is then

$$\sigma^2(P_{CL}) = \sum_{i=0}^{L_{\max}} \left(\frac{i}{\bar{n}_{CA}} \right)^2 p_{CL_i} - \left(\sum_{i=0}^{L_{\max}} \left(\frac{i}{\bar{n}_{CA}} \right) p_{CL_i} \right)^2 \quad (11)$$

which will be used later in calculating the improvement in simulation efficiency.

The brute force method of determining P_{CL} is to sum the cell losses and cell arrivals over the entire set V of connection starting-slot vectors \underline{v}_j . A minor simplification is to fix one of the connections at slot zero because the connection starting-slot vectors \underline{v}_j are translationly invariant from this reference point, resulting in

$$P_{CL} = \frac{\left(\frac{1}{T}\right)^{N_C-1} \sum_{i_2=1}^T \sum_{i_3=1}^T \cdots \sum_{i_{N_C}=1}^T \sum_{j=1}^{T_{SS}} \sum_{n=1}^{N_C} I_{CL}(i_2, i_3, \dots, i_{N_C}, j, n)}{\left(\frac{1}{T}\right)^{N_C-1} \sum_{i_2=1}^T \sum_{i_3=1}^T \cdots \sum_{i_{N_C}=1}^T \sum_{j=1}^{T_{SS}} \sum_{n=1}^{N_C} I_{CA}(i_2, i_3, \dots, i_{N_C}, j, n)} \quad (12)$$

where I_{CL} is the binary indicator function of a cell loss in a service slot,

$$I_{CL} = \begin{cases} 1, & \text{if a cell loss occurs} \\ 0, & \text{if a cell loss does not occur} \end{cases} \quad (13)$$

I_{CA} is the binary indicator function of an arriving cell in a service slot,

$$I_{CA} = \begin{cases} 1, & \text{if a cell arrival occurs} \\ 0, & \text{if a cell arrival does not occur} \end{cases} \quad (14)$$

T_{SS} is the steady-state period and has the same length as a period T , i_l indexes the starting-slot of the \hat{B} length cell burst of the l -th connection, and N_C is the number of connections routed through the output cell buffer. However, for even small values of T and N_C , the exponential order of complexity of (12), $O(N_C T^{N_C})$, quickly becomes unmanageable. Therefore, we are interested in using MC simulation to estimate P_{CL} .

4 Estimation Using Simulation

We first consider conventional MC estimation in a network framework. We then discuss how the simulation itself is greatly simplified by the use of ATM connection traffic descriptors, and show how IS is applied within this framework. We also derive upper and lower bounds of the cell loss probability which can be used to determine if simulation of a particular set of ATM connection traffic descriptors and switch parameters using this method is feasible. Lastly, we consider some examples to demonstrate our methodology.

4.1 Monte Carlo Simulation

The brute force method of determining P_{CL} is intractable if V is large, which is the case for many realistic systems of interest, so we use MC simulation to estimate P_{CL} by randomly selecting N_D of the starting-slot vectors \underline{v}_j out of V and form the estimate \hat{P}_{CL}

$$\hat{P}_{\text{CL}} = \frac{\hat{n}_{\text{CL}}}{\hat{n}_{\text{CA}}} = \frac{\frac{1}{N_D} \sum_{i=1}^{N_D} \sum_{j=1}^{T_{SS}} \sum_{n=1}^{N_C} I_{\text{CL}}(i, j, n)}{\frac{1}{N_D} \sum_{i=1}^{N_D} \sum_{j=1}^{T_{SS}} \sum_{n=1}^{N_C} I_{\text{CA}}(i, j, n)} \quad (15)$$

We can rewrite (15) using the multivalued indicator function of a cell loss for all N_C connections in a steady-state period T_{SS}

$$M_{\text{CL}} = \begin{cases} L, & \text{if } L \text{ cell losses occur for } 1 \leq L \leq L_{\max} \\ 0, & \text{if a cell loss does not occur} \end{cases} \quad (16)$$

and the multivalued indicator function of a cell arrival for all N_C connections in a steady-state period T_{SS}

$$M_{\text{CA}} = \begin{cases} A, & \text{if } A \text{ cell arrivals occur for } 1 \leq A \leq A_{\max} \\ 0, & \text{if a cell arrival does not occur} \end{cases} \quad (17)$$

as

$$\hat{P}_{\text{CL}} = \frac{\hat{n}_{\text{CL}}}{\hat{n}_{\text{CA}}} = \frac{\frac{1}{N_D} \sum_{i=1}^{N_D} M_{\text{CL}}(i)}{\frac{1}{N_D} \sum_{i=1}^{N_D} M_{\text{CA}}(i)} \quad (18)$$

where M_{CL} and M_{CA} map each starting-slot vector \underline{v}_j drawn independently from V to L cell losses and A cell arrivals, respectively, in a steady-state period T_{SS} . The multivalued indicators functions are used instead of the binary indicator functions in order to obtain independence between cell loss events, since it is known that the binary indicator function of

a cell loss results in a correlated binomial distribution because the cell losses are the result of bursty events [13].

Since all N_C connections generate a burst of length \hat{B} cells, the number of arrivals per steady-state period T_{SS} is

$$\bar{n}_{CA} = \frac{1}{N_D} (N_D N_C \hat{B}) = N_C \hat{B} \quad (19)$$

As mentioned previously, \bar{n}_{CA} is constant so we are actually concerned with estimating \bar{n}_{CL} , the numerator in (15), the average number of cell losses per steady-state period T_{SS} . By estimating \bar{n}_{CL} we are implicitly estimating the p_{CL_i} of the multinomial distribution of cell loss.

In order to determine the confidence interval of \hat{P}_{CL} , the MC simulation we consider actually estimates \bar{n}_{CL} and therefore P_{CL} by independently drawing N_S of N_R connection starting-slot vectors \underline{v}_j out of the set V .

$$\hat{P}_{CL} = \frac{1}{N_S N_R} \frac{\sum_{i=1}^{N_S} \sum_{j=1}^{N_R} M_{CL}(i, j)}{\bar{n}_{CA}} \quad (20)$$

We determine the confidence interval of \hat{P}_{CL} by letting the estimate of P_{CL} corresponding to the i -th of N_S sets of N_R independent draws be \hat{P}_{CL_i} , where $i = 1, \dots, N_S$, so that overall estimate of P_{CL} is the sample mean

$$\hat{P}_{CL} = \frac{1}{N_S} \sum_{i=1}^{N_S} \hat{P}_{CL_i} \quad (21)$$

We estimate the variance of \hat{P}_{CL} by the sample variance

$$\hat{\sigma}^2(\hat{P}_{CL}) = \frac{1}{N_S(N_S - 1)} \sum_{i=1}^{N_S} (\hat{P}_{CL_i} - \hat{P}_{CL})^2 \quad (22)$$

We do not assume a normal distribution for the confidence interval, but use the method found in [14], which determined that the confidence interval of a weighted sum of individual multinomial probabilities follows a χ^2 distribution. We state the theorem from [14] with the original notation slightly altered for our application:

Theorem: As $N_S N_R \rightarrow \infty$, the lower limit of the probability that P_{CL} satisfies

$$\hat{P}_{CL} - A\hat{\sigma}(\hat{P}_{CL}) \leq P_{CL} \leq \hat{P}_{CL} + A\hat{\sigma}(\hat{P}_{CL}) \quad (23)$$

is at least $1 - \alpha$ where A is the positive square root of the upper $(1 - \alpha)$ -th percentage point of the χ^2 distribution with L_{\max} degrees of freedom.

The multinomial framework used here reduces to the more familiar binomial case with normally distributed confidence intervals when $L_{\max} = 1$. Using the χ^2 distribution results in wider confidence intervals than using the normal distribution, but because we have used the multivalued indicator functions in (18) rather than the binary indicator function in (15), we have independence between cell loss events. This is critical since cell loss events occur in a bursty fashion, which makes the binomial random variable represented by the binary indicator function correlated [13]. Measuring this correlation is difficult when rare events are involved, which in turn makes deriving accurate confidence intervals of our estimates difficult. This problem is solved by the multinomial framework because the selection of each connection starting-slot vector \underline{v}_j from V is done independently.

4.2 Monte Carlo Simulation Framework

As stated previously, the randomness in the system model is found in the position of the connection starting-slots. Once the connection starting-slot vector \underline{v}_j has been drawn, the system model behaves deterministically in that the number of cell arrivals as well as the number of cells serviced is known for each slot.

By observing the performance of this system, several characteristics were determined that assist us in speeding up the simulation without applying IS. First, we are working with a stable system since the size of the output cell buffer is finite. Second, we are able to detect the onset of the steady-state. Last and most important, the system exhibits predictable behavior in the mechanism that causes cell losses, so we can anticipate if a cell loss is going to occur for a given connection starting-slot vector \underline{v}_j . We need only observe the queue length for three periods to determine whether or not a cell loss occurs. We use the following rules to take advantage of these three characteristics:

Define $X[n]$ as the queue length in cells and $A[n]$ as the number of arrivals in cells at the beginning of slot n where n is a non-negative integer.

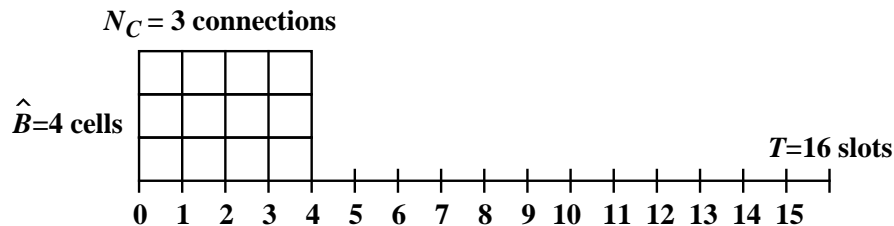


Figure 6: Aligned at zero example.

1. The cell arrivals that occur during the 2nd period constitute the steady-state cell arrivals, i.e. $A[T + n] = A[kT + T + n]$ where k is a non-negative integer.
2. If the output cell buffer size remains unchanged from the end of the second period to the end of the third period, $X[2T] = X[3T]$, and a cell loss does not occur, $M_{CL}(2T, 3T) = 0$, the steady-state has been reached and a cell loss will never occur for this particular connection starting-slot vector \underline{v}_j .
3. If the output cell buffer size increases from the end of the second period to the end of the third period, $X[2T] < X[3T]$, the output cell buffer size keeps increasing until a cell loss occurs.
4. When a cell loss occurs in a period, $M_{CL}(kT, (k+1)T) > 0$ for $k \geq 2$, k an integer, and the output cell buffer size at the end of consecutive periods is the same, $X[kT] = X[(k+1)T]$, the steady-state has been reached.

The connection starting-slot vector \underline{v}_j that results in the maximum number of cell losses in a period corresponds to the situation where all the connections start at slot zero, which we define as the aligned at zero (AAZ) case as shown in Figure 6. The maximum number of cell losses that can occur in a steady-state period T_{SS} , L_{max} , is then

$$L_{max} = \lfloor \hat{B}(N_C - \mu) - K \rfloor + 1 \quad (24)$$

We will consider two types of cell losses: Type A and Type B. Type A cell losses occur when individual connections start so close together that the server can not keep up with the

arriving cells over a time $t < T$ and an output cell buffer overflow results in a cell loss. Type B cell losses occur when the total number of arriving cells in a period T is greater than the service available. We will be mostly interested in Type A cell losses since such cell losses are rare events.

Type A cell losses can not occur until there is a sufficient number of arriving cells over a given time window. Since an arrival must occur for a cell loss to occur, we assume the AAZ case to get the worst-case conditions for cell loss, that is the least number of connections that result in a cell loss requires that

$$\hat{B}N_C > K + \hat{B}\mu \quad (25)$$

Thus the minimum number of connections, $N_{C_{\min}}$, required to cause a cell loss is

$$N_{C_{\min}} = \lceil \mu + K/\hat{B} \rceil \quad (26)$$

where $\lceil x \rceil$ is the ceiling operation that rounds x up to the next highest integer if x is not an integer. For a number of connections $N_C < N_{C_{\min}}$, cell losses at the output cell buffer can never occur.

Type B cell losses, where a cell loss is guaranteed *a priori*, occur when the sum of the average rates of all the connections is greater than the available service, or

$$\mu < \bar{\lambda}N_C \quad (27)$$

Therefore, the number of connections, N_{C_g} , required to guarantee a cell loss at the output cell buffer is

$$N_{C_g} = \lceil \mu/\bar{\lambda} \rceil \quad (28)$$

The performance region we are interested in is over the range of connections $N_{C_{\min}} \leq N_C \leq N_{C_g}$. However, when cell losses are rare events, which is often the case in realistic systems of interest, MC simulation is slow, therefore we use IS to make the simulation more efficient.

4.3 Efficient Simulation Using Importance Sampling

The basic notion behind IS is to modify or bias the original probability density function for the connection starting-slot vector, $f_{\mathbf{v}}(v)$, to a new probability density function, $f_{\mathbf{v}}^*(v)$, that reduces the variance of the estimate of P_{CL} , while weighting each cell loss event so that P_{CL}^* remains an unbiased estimate of P_{CL} . The only required condition on the pdf $f_{\mathbf{v}}^*(v)$ is that $f_{\mathbf{v}}^*(v) > 0$ when the multivalued indicator function $M_{\text{CL}}(\underline{v}_j \in V_{\text{CL}}) > 0$, where $V_{\text{CL}} \subset V$ is the subset of connections starting-slot vectors which cause a cell loss. The pdf $f_{\mathbf{v}}(v)$ we consider here is the product of i.i.d. uniform probability density functions.

The key to using IS is knowing how to bias the pdf $f_{\mathbf{v}}(v)$. Intuitively, we want to use IS to increase the number of cell losses that occur. Type B cell losses result in high values for P_{CL} , so we are most interested in Type A cell losses. We have previously determined that the connection starting-slot has a uniform probability distribution over a period T . Unfortunately, there is no general bias scheme for the uniform probability distribution known to the authors that is effective.

4.3.1 Theory

For this problem, we use *a priori* information about the system model to determine an effective bias scheme. The approach we take is based on the idea that one connection is fixed to start at slot zero and an examination of the restrictions involved with using IS. Assume we know the set V_{CL} of connection starting-slot vectors \underline{v}_j that cause cell losses. Then P_{CL} is

$$P_{\text{CL}} = \frac{\left(\frac{1}{T}\right)^{N_C-1} \sum_{i=1}^{|V_{\text{CL}}|} M_{\text{CL}}(i)}{N_C \hat{B}} \quad (29)$$

IS is applied by reducing the set of connection starting-slot vectors \underline{v}_j from $|V|$ to $|V^*| < |V|$ such that all of the non-zero cell loss events are included in V^* , i.e. $V_{\text{CL}} \subset V^* \subset V$. This results in $n_{\text{CL}_i}^* = n_{\text{CL}_i}$ for $i = 1 \dots L_{\text{max}}$, which implies that $n_{\text{CL}_0}^* < n_{\text{CL}_0}$. By maintaining all the connection starting-slot vectors that cause cell losses in V^* , we do not violate the condition necessary to apply IS that $f_{\mathbf{v}}^*(v) > 0$ when $M_{\text{CL}}(\underline{v}_j \in V_{\text{CL}}) > 0$. Ideally we would

like $V_{\text{CL}} = V^*$, but as long as $|V^*| \ll |V|$ and $V_{\text{CL}} \subset V^*$, the efficiency of the simulation can be increased.

This reduction in connection starting-slot vectors results in the biased individual probabilities from the multinomial distribution for cell loss

$$p_{\text{CL}_i}^* = \frac{n_{\text{CL}_i}^*}{|V^*|} = \frac{n_{\text{CL}_i}}{|V^*|} \quad (30)$$

The relationship between the unbiased p_{CL_i} and the biased $p_{\text{CL}_i}^*$ is based on the the weight function or likelihood ratio, $W = |V^*|/|V|$, so that

$$p_{\text{CL}_i} = \frac{n_{\text{CL}_i}}{|V|} \frac{|V^*|}{|V^*|} = \frac{|V^*|}{|V|} p_{\text{CL}_i}^* = W p_{\text{CL}_i}^* \quad (31)$$

Therefore, IS implicitly biases the p_{CL_i} to $p_{\text{CL}_i}^*$.

The cell loss probability found using IS, P_{CL}^* , is the same as that found without using IS.

$$P_{\text{CL}}^* = \sum_{i=0}^{L_{\text{max}}} \left(\frac{iW}{\bar{n}_{\text{CA}}} \right) p_{\text{CL}_i}^* = \sum_{i=0}^{L_{\text{max}}} \left(\frac{i}{\bar{n}_{\text{CA}}} \right) p_{\text{CL}_i} = P_{\text{CL}} \quad (32)$$

The variance of P_{CL}^* is

$$\begin{aligned} \sigma^2(P_{\text{CL}}^*) &= \sum_{i=0}^{L_{\text{max}}} \left(\frac{iW}{\bar{n}_{\text{CA}}} \right)^2 p_{\text{CL}_i}^* - \sum_{i=0}^{L_{\text{max}}} \left(\frac{iW}{\bar{n}_{\text{CA}}} \right) p_{\text{CL}_i}^* \\ &= W \sum_{i=0}^{L_{\text{max}}} \left(\frac{i}{\bar{n}_{\text{CA}}} \right)^2 p_{\text{CL}_i} - \sum_{i=0}^{L_{\text{max}}} \left(\frac{i}{\bar{n}_{\text{CA}}} \right) p_{\text{CL}_i} \end{aligned} \quad (33)$$

which will be used later in calculating the improvement in simulation efficiency.

4.3.2 Application

We mentioned previously that Type A cell losses occur when individual connections start so close together that the server can not keep up with the arriving cells over a time $t < T$ and an output cell buffer overflow results in a cell loss. Assume that one connection is fixed to start at slot zero. Then for a given number of connections, we can determine the distance in slots c from the connection fixed to start at slot zero, so that a cell loss does not occur if any connection starts its \hat{B} length cell burst more than c slots away from slot zero. The

multi-valued indicator function M_{CL} is always zero outside this region and the support of the uniform pdf that originally covered the entire interval T can be reduced around both sides of slot zero.

Define c_{\min} as the distance in slots from the start of the \hat{B} length cell burst of the connection furthest from slot zero for a connection starting-slot vector \underline{v}_j that still causes a cell loss for $N_{C_{\min}}$ connections, and define $c_{\min+1}$ similarly for $N_{C_{\min+1}}$ connections. The equations used to determine c_{\min} and $c_{\min+1}$ are a function of the ATM connection traffic descriptors and the switch parameters.

For $\mu \geq 1.0$ and $N_{C_{\min}} - 1 < \mu$

$$c_{\min} = \lfloor \hat{B} - \frac{K}{N_{C_{\min}} - \mu} \rfloor + 1 \quad (34)$$

unless $\hat{B} - \frac{K}{N_{C_{\min}} - \mu} = 0$ then $c_{\min} = 0$, and

$$c_{\min+1} = \hat{B} + c_{\min} \quad (35)$$

For $\mu \geq 1.0$ and $N_{C_{\min}} - 1 \geq \mu$

$$c_{\min} = \lfloor \hat{B} (N_{C_{\min}} - \mu) - K \rfloor + 1 \quad (36)$$

unless $\hat{B} (N_{C_{\min}} - \mu) - K = 0$ then $c_{\min} = 0$, and

$$c_{\min+1} = \hat{B} + c_{\min} \quad (37)$$

For $\mu < 1.0$

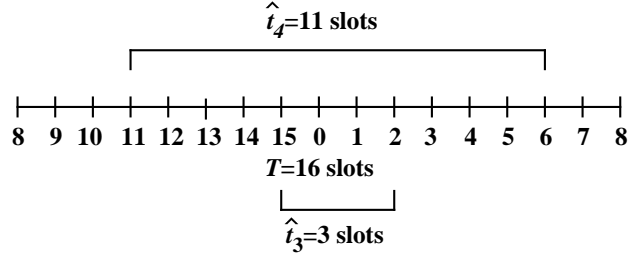
$$c_{\min} = \lfloor \hat{B}/\mu - \hat{B} \rfloor + 1 \quad (38)$$

and

$$c_{\min+1} = \lfloor 2\hat{B}/\mu - \hat{B} \rfloor + 1 \quad (39)$$

The number of slots covered by the biased uniform pdf is

$$\hat{t} = 2c + 1 \quad (40)$$

Figure 7: Example of \hat{t} distances.

compared to the unbiased uniform pdf that covers the full T slot period. An example of \hat{t} slot distances is shown in Figure 7.

This quasi-analytic form of IS fixes the weight within the simulation according to how much the support of the original uniform pdf is reduced around the connection fixed to start at slot zero. Each connection starting-slot \underline{v}_j is drawn independently from the identical biased uniform distribution, which results in the weight function

$$W = f_{\mathbf{v}}(v)/f_{\mathbf{v}}^*(v) = \prod_{j=2}^{N_C} f_j(v)/f_j^*(v) = (\hat{t}/T)^{N_C-1} \quad (41)$$

The expression of the weight function in (41) is the same as that in (31) where $|V| = T^{N_C-1}$ and $|V^*| = \hat{t}^{N_C-1}$.

4.3.3 Simulation Using IS

The MC simulation using IS estimates P_{CL}^* by independently drawing N_S sets of N_R connection starting-slot vectors \underline{v}_j out of the set V^*

$$\hat{P}_{\text{CL}}^* = \frac{1}{N_S N_R} \sum_{i=1}^{N_S} \sum_{j=1}^{N_R} M_{\text{CL}}^*(i, j) \quad (42)$$

where $M_{\text{CL}}^*(i, j)$ takes on the values 1 to L_{max} and indicates the number of cell losses for the (i, j) -th connection starting-slot vector \underline{v}_j drawn from the biased distribution, $f_{\mathbf{v}}^*(v)$.

Now for $N_{C_{\text{min}}}$, we observe that $P_{\text{CL}}(N_{C_{\text{min}}}) = P_{\text{CL}}(\hat{t}_{\text{min}})$, but for $P_{\text{CL}}(N_{C_{\text{min}+1}})$ this is not the case. Rather we have

$$P_{\text{CL}}(N_{C_{\text{min}+1}}) = P_{\text{CL}}(\hat{t}_{\text{min}+1}) + P_{\text{CL}}(\hat{t}_{\text{min}+1}^c) \quad (43)$$

We first find the cell losses involving the connection fixed at zero, or $N_{\text{CL}}(\hat{t}_{\min+1})$, and then add in the cell losses not involving the connection fixed at zero, $N_{\text{CL}}(\hat{t}_{\min+1}^c)$. We account for the cell losses outside the $\hat{t}_{\min+1}$ range by sliding the \hat{t}_{\min} range over the $T - \hat{t}_{\min+1}$ slots outside $\hat{t}_{\min+1}$ and subtracting the AAZ case at the boundary so that

$$\begin{aligned} N_{\text{CL}}(N_{C_{\min+1}}) &= N_{\text{CL}}(\hat{t}_{\min+1}) + C_{N_{C_{\min}}}^{N_{C_{\min+1}}} (T - \hat{t}_{\min+1}) N_{\text{CL}}(N_{C_{\min}}) \\ &\quad + N_{\text{CL}}(N_{C_{\min}}) - N_{\text{CL}}(N_{C_{\min}} \text{ AAZ}) \end{aligned} \quad (44)$$

where $C_{N_{C_{\min}}}^{N_{C_{\min+1}}} = N_{C_{\min+1}}$ is the combination operator and this factor accounts for the different combinations of $N_{C_{\min}}$ connections that can occur outside the $\hat{t}_{\min+1}$ range. We divide by a factor of $1/T^{N_{C_{\min+1}}-1}$ and use (44) to find

$$\bar{n}_{\text{CL}}(N_{C_{\min+1}}) \leq \bar{n}_{\text{CL}}(\hat{t}_{\min+1}) + \left(\frac{N_{C_{\min+1}}(T - \hat{t}_{\min+1}) + 1}{T} \right) \bar{n}_{\text{CL}}(N_{C_{\min}}) \quad (45)$$

or in terms of P_{CL} we have

$$\begin{aligned} P_{\text{CL}}(N_{C_{\min+1}}) &\leq \frac{\bar{n}_{\text{CL}}(\hat{t}_{\min+1})}{\bar{n}_{\text{CA}}(N_{C_{\min+1}})} \\ &\quad + \frac{\left(\frac{N_{C_{\min+1}}(T - \hat{t}_{\min+1}) + 1}{T} \right) \bar{n}_{\text{CL}}(N_{C_{\min}})}{\bar{n}_{\text{CA}}(N_{C_{\min+1}})} \end{aligned} \quad (46)$$

$$\begin{aligned} &= P_{\text{CL}}(\hat{t}_{\min+1}) \\ &\quad + \left(\frac{N_{C_{\min}}}{N_{C_{\min+1}}} \right) \left(\frac{N_{C_{\min+1}}(T - \hat{t}_{\min+1}) + 1}{T} \right) P_{\text{CL}}(N_{C_{\min}}) \end{aligned} \quad (47)$$

which is a very tight bound since the AAZ case constitutes such a small portion of the total number of cell losses. In finding $P_{\text{CL}}(N_{C_{\min+1}})$, we actually violate the IS rule that $f_{\mathbf{v}}^*(v) > 0$ when $M_{\text{CL}}(\underline{v}_j \in V_{\text{CL}}) > 0$ in a known way, and correct for this by replacing the correct number of cell losses using $P_{\text{CL}}(N_{C_{\min}})$. The simulation uses the estimates of $P_{\text{CL}}(\hat{t}_{\min+1})$ and $P_{\text{CL}}(N_{C_{\min}})$ to estimate $P_{\text{CL}}(N_{C_{\min+1}})$ in (47).

The variance of $\hat{P}_{\text{CL}}^*(N_{C_{\min}})$ is the sample variance

$$\hat{\sigma}^2(\hat{P}_{\text{CL}}^*(N_{C_{\min}})) = \hat{\sigma}^2(\hat{P}_{\text{CL}}^*(\hat{t}_{\min})) = \frac{1}{N_S(N_S - 1)} \sum_{i=1}^{N_S} (\hat{P}_{\text{CL}_i}^* - \hat{P}_{\text{CL}}^*)^2 \quad (48)$$

The variance of $\hat{P}_{\text{CL}}^* (N_{C_{\min+1}})$ is the weighted sum of the variances $\hat{\sigma}^2 (\hat{P}_{\text{CL}}^* (\hat{t}_{\min}))$ and $\hat{\sigma}^2 (\hat{P}_{\text{CL}}^* (\hat{t}_{\min+1}))$

$$\begin{aligned} \hat{\sigma}^2 (\hat{P}_{\text{CL}}^* (N_{C_{\min+1}})) &= \hat{\sigma}^2 (\hat{P}_{\text{CL}}^* (\hat{t}_{\min+1})) \\ &+ \left(\left(\frac{N_{C_{\min}}}{N_{C_{\min+1}}} \right) \frac{(N_{C_{\min+1}} (T - \hat{t}_{\min+1}) + 1)}{T} \right)^2 \hat{\sigma}^2 (\hat{P}_{\text{CL}}^* (\hat{t}_{\min})) \end{aligned} \quad (49)$$

The quantities in (48) and (49) are used in the calculation of the confidence intervals of the estimates for $\hat{P}_{\text{CL}}^* (N_{C_{\min}})$ and $\hat{P}_{\text{CL}}^* (N_{C_{\min+1}})$

4.3.4 Calculation of the Improvement Factor

The improvement, R_{net} , that results from using IS describes the factor by which an IS estimator that uses the given method is more accurate than a conventional MC estimator based on the same sample size, or, equivalently, how many fewer steady-state periods must be simulated to obtain a given accuracy. The improvement, R_{net} , is found analytically by taking the ratio of the variances, $\sigma^2 (P_{\text{CL}})$ and $\sigma^2 (P_{\text{CL}}^*)$ from (11) and (33), respectively.

$$R_{\text{net}} = \frac{\sigma^2 (P_{\text{CL}})}{\sigma^2 (P_{\text{CL}}^*)} = \frac{\sum_{i=0}^{L_{\max}} \left(\frac{i}{\bar{n}_{\text{CA}}} \right)^2 p_{\text{CL}_i} - P_{\text{CL}}^2}{W \sum_{i=0}^{L_{\max}} \left(\frac{i}{\bar{n}_{\text{CA}}} \right)^2 p_{\text{CL}_i} - P_{\text{CL}}^2} \rightarrow \frac{1}{W} \text{ as } P_{\text{CL}} \rightarrow 0 \quad (50)$$

As $P_{\text{CL}} \rightarrow 0$, the improvement obtained by using IS approaches the inverse of the weight from above, so a lower bound on the improvement, $\lfloor R_{\text{net}} \rfloor$, is

$$\lfloor R_{\text{net}} \rfloor = 1/W = (T/\hat{t})^{N_C - 1} \quad (51)$$

From (51), we see that improvement always results if $\hat{t} < T$ for any value of P_{CL} if $N_C > 2$.

4.4 Upper and Lower Bounds of the Cell Loss Probability

The distance c_{\min} can be used to derive an upper bound for P_{CL} for $N_{C_{\min}}$ connections by assuming that L_{\max} cell losses from (24) occur whenever one or more cell losses occur. For at least one cell loss to occur, all $N_{C_{\min}}$ connections must start within c_{\min} slots of each other. There will be $c_{\min} + 1$ cases where this can occur within \hat{t}_{\min} slots since $\hat{t}_{\min} = 2c_{\min} + 1$ (see

Figure 7). If we assume that one connection is fixed at slot zero, the probability of the remaining $N_{C_{\min}} - 1$ connections starting within c_{\min} slots of each other is $(c_{\min}/T)^{N_{C_{\min}}-1}$.

Therefore an upper bound for P_{CL} is

$$\begin{aligned}
P_{CL}(N_{C_{\min}}) \leq [P_{CL}(N_{C_{\min}})] &= \frac{[\bar{n}_{CL}]}{\bar{n}_{CA}} \\
&= \frac{P(\text{any cell loss}) L_{\max}}{\bar{n}_{CA}} \\
&= \frac{\left(\frac{c_{\min}}{T}\right)^{N_{C_{\min}}-1} (c_{\min} + 1) \left(\lfloor \hat{B}(N_{C_{\min}} - \mu) - K \rfloor + 1\right)}{N_{C_{\min}} \hat{B}}
\end{aligned} \tag{52}$$

For the experimental systems considered later, the upper bound in (52) is one or two orders of magnitude high.

An upper bound for P_{CL} for $N_{C_{\min+1}}$ connections can be derived using the same method used to find the $\hat{P}_{CL}(N_{C_{\min+1}})$ in (47), except that we substitute the bounds for $P_{CL}(\hat{t}_{\min+1})$ and $P_{CL}(\hat{t}_{\min+1}^c)$ into (47) rather than the estimates. We upper bound $P_{CL}(\hat{t}_{\min+1})$ using (52) as

$$\begin{aligned}
P_{CL}(\hat{t}_{\min+1}) \leq [P_{CL}(\hat{t}_{\min+1})] &= \frac{[\bar{n}_{CL}]}{\bar{n}_{CA}} \\
&= \frac{P(\text{any cell loss}) L_{\max}}{\bar{n}_{CA}} \\
&= \frac{\left(\frac{c_{\min+1}}{T}\right)^{N_{C_{\min+1}}-1} (c_{\min+1} + 1) \left(\lfloor \hat{B}(N_{C_{\min+1}} - \mu) - K \rfloor + 1\right)}{N_{C_{\min+1}} \hat{B}}
\end{aligned} \tag{53}$$

We use $[P_{CL}(N_{C_{\min}})]$ in the expression for $P_{CL}(\hat{t}_{\min+1}^c)$ so that

$$\begin{aligned}
P_{CL}(N_{C_{\min+1}}) &\leq [P_{CL}(N_{C_{\min+1}})] \\
&= [P_{CL}(\hat{t}_{\min+1})] + [P_{CL}(\hat{t}_{\min+1}^c)] \\
&= [P_{CL}(\hat{t}_{\min+1})] \\
&\quad + \frac{N_{C_{\min}}}{N_{C_{\min+1}}} \left(\frac{N_{C_{\min+1}} (T - \hat{t}_{\min+1}) + 1}{T} \right) [P_{CL}(N_{C_{\min}})]
\end{aligned} \tag{54}$$

A lower bound for P_{CL} for a number of connections less than that guaranteed to cause an error, $N_C < N_{C_g}$, is found by assuming that the only time cells losses occur is for the AAZ

case shown in Figure 6, where L_{\max} cell losses occur. If we assume that one connection is fixed at slot zero, the probability of the remaining $N_C - 1$ connections starting at slot zero, is $(1/T)^{N_C-1}$. Therefore, a lower bound for P_{CL} is

$$[P_{\text{CL}}(N_C < N_{C_g})] = \frac{\lfloor \bar{n}_{\text{CL}} \rfloor}{\bar{n}_{\text{CA}}} = \frac{\left(\frac{c_{\min}}{T}\right)^{N_C-1} (\lfloor \hat{B}(N_C - \mu) - K \rfloor + 1)}{N_C \hat{B}} \quad (55)$$

For combinations of $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ -triplets and $(\hat{\mu}, K)$ -pairs that result in $c_{\min} = 0$, the lower bound is exactly P_{CL} for $N_{C_{\min}}$ connections.

A lower bound for P_{CL} for a number of connections greater than or equal to that guaranteed to cause an error, $N_C \geq N_{C_g}$, is found by assuming that for every connection starting-slot vector \underline{v}_j , the L_g cell losses associated with the overflow in a period T occur so that

$$P_{\text{CL}}(N_C \geq N_{C_g}) \geq [P_{\text{CL}}] = \frac{\lfloor \bar{n}_{\text{CL}} \rfloor}{\bar{n}_{\text{CA}}} = \frac{L_g}{\bar{n}_{\text{CA}}} = 1 - \mu T / N_C \hat{B} \quad (56)$$

where $L_g = \bar{n}_{\text{CA}} - \mu T$. This is a lower bound because it is possible to lose more than L_g cells, such as in the AAZ case where $L_g \leq L_{\max}$. In (56), we also see that as N_C increases, $P_{\text{CL}} \rightarrow 1$ as expected.

4.5 Feasibility of System Simulation

We can determine whether or not a given system can be simulated effectively before simulation by using the upper bound of P_{CL} and the lower bound of R_{net} . We define \tilde{P}_{CL} as the equivalent cell loss probability that must be simulated using conventional MC simulation after the improvement has been taken into account. We will see later that the upper bound of $P_{\text{CL}}(N_{C_{\min}})$ is roughly two orders of magnitude high for the experimental examples we consider here, so to find \tilde{P}_{CL} we divide $[P_{\text{CL}}(N_{C_{\min}})]$ by a factor of 100 as in (57)

$$\tilde{P}_{\text{CL}}(N_{C_{\min}}) = ([P_{\text{CL}}(N_{C_{\min}})]/100) \left(\frac{T}{\hat{t}_{\min}}\right)^{N_{C_{\min}}-1} \quad (57)$$

We define the simulation effort, E , as

$$E = T N_C / \tilde{P}_{\text{CL}} \quad (58)$$

noting that the effort required by any simulation is inversely proportional to P_{CL} and proportional to T and N_C . Based on the results of the experimental examples we consider later, E can be used to predict whether or not the system is amenable to simulation.

System	$\hat{\lambda}$, Mbps	$\bar{\lambda}$, Mbps	\hat{B} , cells	$\hat{\mu}$, Mbps	K , cells
D1	500	50	2	200	4
D2	8	2	4	13	5

Table 1: ATM connection traffic descriptors and switch parameters for the simple examples.

System	T , slots	μ , cells/slot	$N_{C_{\min}}$, conns.	c_{\min} , slots	$L_{\max}(N_{C_{\min}})$, cells
D1	20	0.8	3	1	1
D2	16	1.625	3	1	2

Table 2: Derived simulation parameters for the simple examples.

4.6 Simulation Methodology

In order to demonstrate and verify our methodology, we considered the two simple examples with the ATM connection traffic descriptors and switch parameters given in Table 1. The parameters for the simple examples were selected so that the P_{CL} and p_{CL_i} could be determined by brute force as in (12). The service rates were selected so that the different formulas for determining c_{\min} and $c_{\min+1}$ could be tested depending on whether $\mu \geq 1$ or $\mu < 1$.

The ATM connection traffic descriptors and switch parameters in Table 1 are substituted into the appropriate equations in Section 4.3.2 to find the derived simulation parameters for the simple examples shown in Table 2. The derived simulation parameters are used as inputs to the simulation.

The exact results for $N_{C_{\min}}$ and $N_{C_{\min+1}}$ connections for the simple examples are shown in Table 3. We performed conventional MC simulation and simulation using IS for $N_{C_{\min}}$ and $N_{C_{\min+1}}$ connections using $N_S = 20$ of N_R steady-state periods shown in Table 3 in order to compare MC and IS simulation results. The estimates of P_{CL} from the MC and

System	N_C , conns.	N_R	P_{CL}
D1	3	10,000	2.917×10^{-3}
D1	4	10,000	9.975×10^{-3}
D2	3	40,000	2.604×10^{-3}
D2	4	40,000	1.811×10^{-2}

Table 3: Exact results for the example systems.

System	N_C , conns.	\hat{P}_{CL}	95% CI	\hat{P}_{CL}^*	95% CI
D1	3	2.956×10^{-3}	$(2.850 \times 10^{-3}, 3.062 \times 10^{-3})$	2.915×10^{-3}	$(2.908 \times 10^{-3}, 2.922 \times 10^{-3})$
D1	4	9.975×10^{-3}	$(9.773 \times 10^{-3}, 1.022 \times 10^{-2})$	9.981×10^{-3}	$(9.941 \times 10^{-3}, 1.004 \times 10^{-2})$
D2	3	2.621×10^{-3}	$(2.578 \times 10^{-3}, 2.666 \times 10^{-3})$	2.604×10^{-3}	$(2.600 \times 10^{-3}, 2.608 \times 10^{-3})$
D2	4	1.827×10^{-2}	$(1.795 \times 10^{-3}, 1.859 \times 10^{-3})$	1.814×10^{-2}	$(1.798 \times 10^{-3}, 1.832 \times 10^{-3})$

Table 4: MC and IS estimates and confidence intervals for the simple examples.

System	Exact		Observed		1/W	
	R_{net}	$N_{C_{min}}$	R_{net}	$N_{C_{min}}$	for $N_{C_{min}}$	for $N_{C_{min+1}}$
D1	196.5	226.0	44.4	26.0	23.7	11.0
D2	96.0	123.7	28.4	4.3	4.0	3.1

Table 5: Improvements for the simple examples.

IS simulations and the resulting confidence intervals of the simple examples are shown in Table 4. Confidence intervals for the estimates of P_{CL} for $N_{C_{min}}$ and $N_{C_{min+1}}$ connections are determined using the method from [14] given in (23). As expected, the estimates of P_{CL} found using IS are closer to the exact value for P_{CL} than the MC estimates, and also have tighter confidence intervals.

Shown in Table 5 are the improvement factors obtained for the simple examples using several different methods. The exact improvement is found by substituting the p_{CL_i} found by brute force into (50). The observed improvement is found by taking the ratio of $\hat{\sigma}^2(\hat{P}_{CL})$ and $\hat{\sigma}^2(\hat{P}_{CL}^*)$ since the MC and IS simulations were performed for an identical number of N_R steady-state periods. Finally, the value of $1/W$ is given, and this quantity is indeed a lower bound for R_{net} . Clearly, the computational effort expended in estimating P_{CL} by simulation for the simple examples is greater than that required by brute force, but the purpose of this exercise was to verify the methodology itself rather than estimate P_{CL} .

5 Experimental Examples

We considered the following experimental examples, where P_{CL} and p_{CL_i} are unknown, with the ATM connection traffic descriptors and switch parameters shown in Table 6. The derived simulation parameters for the experimental examples are shown in Table 7. The cell loss probability estimates for the experimental examples are shown in Tables 8-11 and Figures

System	$\hat{\lambda}$, Mbps	$\bar{\lambda}$, Mbps	\hat{B} , cells	$\hat{\mu}$, Mbps	K , cells
1	15	1	10	120	100
2	15	1	200	120	100
3	50	1	10	120	100
4	50	1	200	120	100
5	75	1	200	120	100
6	15	5	25	120	100
7	50	5	25	120	100
8	500	2	10	120	200
9	500	10	25	120	200
10	500	1	100	120	200
11	500	2	50	120	200
12	500	2	100	120	200
13	2.1	0.031	148	120	100
14	32	1	200	120	100

Table 6: ATM connection traffic descriptors and switch parameters for the experimental examples.

System	T , slots	μ , cells/slot	$N_{C_{\min}}$, conns.	c_{\min} , slots	$L_{\max}(N_{C_{\min}})$, cells
1	150	8.0	18	0	1
2	3000	8.0	9	101	101
3	500	2.4	13	7	7
4	10000	2.4	3	34	21
5	15000	1.6	3	129	181
6	75	8.0	12	0	1
7	250	2.4	7	15	16
8	2500	0.24	21	32	8
9	1250	0.24	9	80	20
10	50000	0.24	3	317	77
11	12500	0.24	5	159	39
12	25000	0.24	3	317	77
13	10025	57 1/7	58	32	27
14	6400	3.75	5	121	151

Table 7: Derived simulation parameters for the experimental examples.

N_C , conns.	\hat{P}_{CL}	Comment
120	1.38×10^{-2}	$\epsilon = 0.02$
110	1.96×10^{-3}	$\epsilon = 0.04$
100	2.49×10^{-4}	$\epsilon = 0.1$
90	2.08×10^{-5}	$\epsilon = 0.29$
80	2.03×10^{-6}	$\epsilon = 0.2$
19	3.92×10^{-22}	Upper Bound
18	5.64×10^{-40}	Exact
17	0	Exact

Table 8: Cell loss probability as a function of the number of connections for system 1.

N_C , conns.	\hat{P}_{CL}	Comment
120	2.10×10^{-2}	$\epsilon = 0.02$
110	3.68×10^{-3}	$\epsilon = 0.07$
100	5.50×10^{-4}	$\epsilon = 0.10$
90	6.78×10^{-5}	$\epsilon = 0.12$
80	7.36×10^{-6}	$\epsilon = 0.10$
14	7.17×10^{-20}	Upper Bound
13	2.88×10^{-24}	Upper Bound
11	0	Exact

Table 9: Cell loss probability as a function of the number of connections for system 3.

8-16. The estimates of P_{CL} for systems 1, 3, 8, and 13 are shown in tables because we were unable to generate complete performance curves for reasons which we discuss later. However, P_{CL} for these systems is extremely low and in such systems we give the upper bound. For systems 1 and 6, we are able to determine the exact P_{CL} using the lower bound in (55) since $c_{\min} = 0$, which results in $L_{\max} = 1$ cell losses for the AAZ case for $N_{C_{\min}}$ connections. Note that a $P_{CL} \approx 10^{-14}$ corresponds to 1 cell loss per year for a service rate on the order of 1 Gbps.

N_C , conns.	\hat{P}_{CL}	Comment
61	1.64×10^{-2}	Exact
60	1.96×10^{-3}	$\epsilon = 0.01$
22	4.84×10^{-32}	Upper Bound
21	1.75×10^{-38}	Upper Bound
20	0	Exact

Table 10: Cell loss probability as a function of the number of connections for system 8.

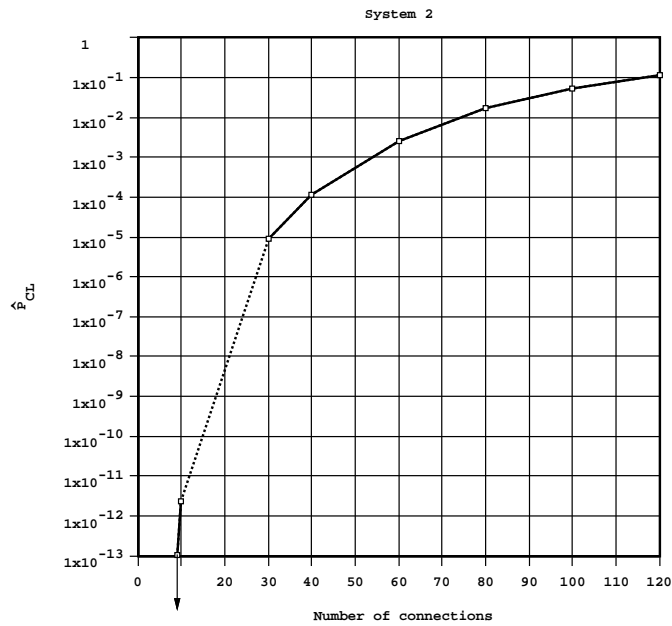


Figure 8: Cell loss probability as a function of the number of connections for system 2 (The dashed portion of the curve is a straight line interpolation).

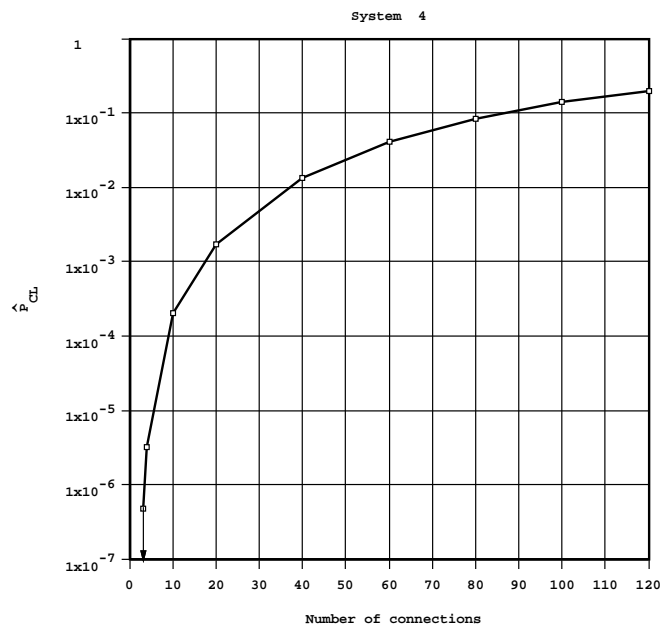


Figure 9: Cell loss probability as a function of the number of connections for system 4.

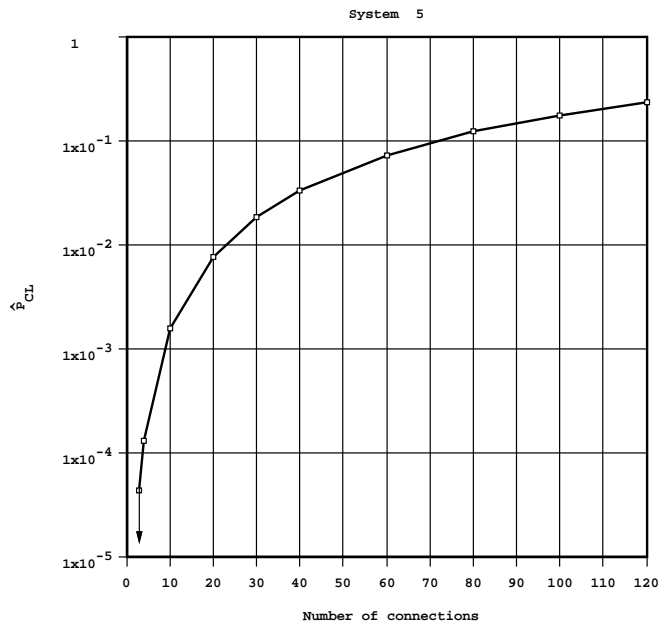


Figure 10: Cell loss probability as a function of the number of connections for system 5.

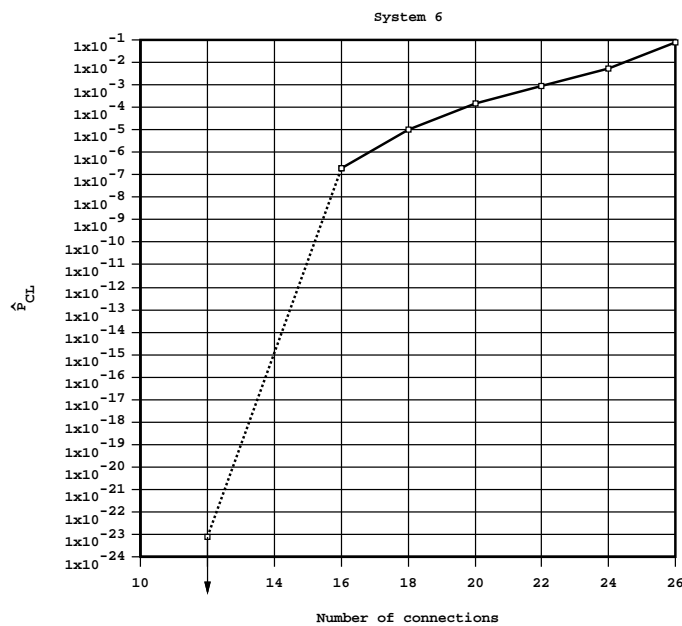


Figure 11: Cell loss probability as a function of the number of connections for system 6 (The dashed portion of the curve is a straight line interpolation).

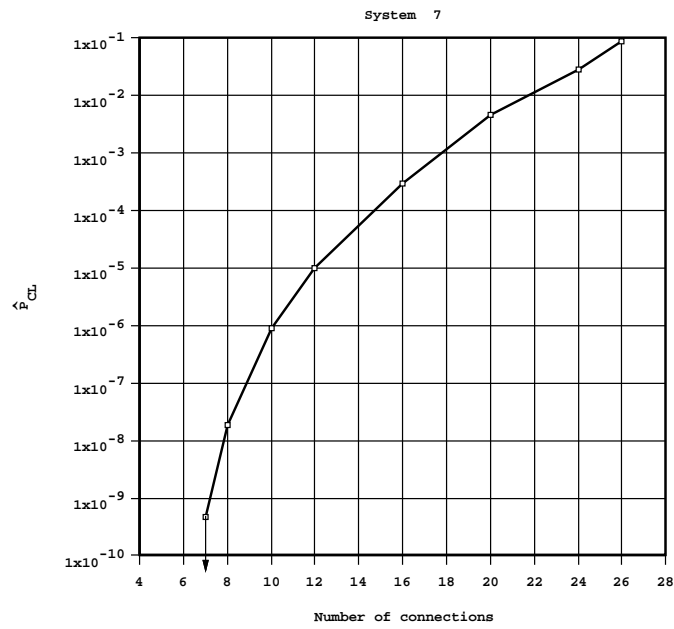


Figure 12: Cell loss probability as a function of the number of connections for system 7.

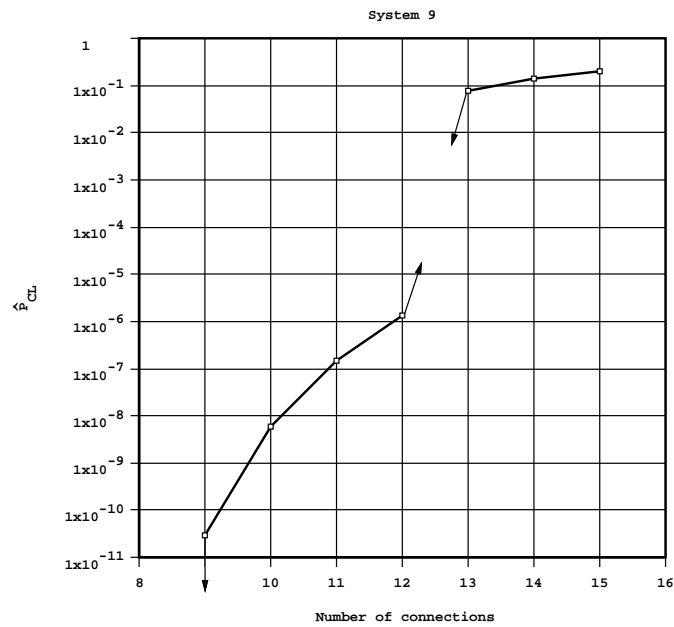


Figure 13: Cell loss probability as a function of the number of connections for system 9.

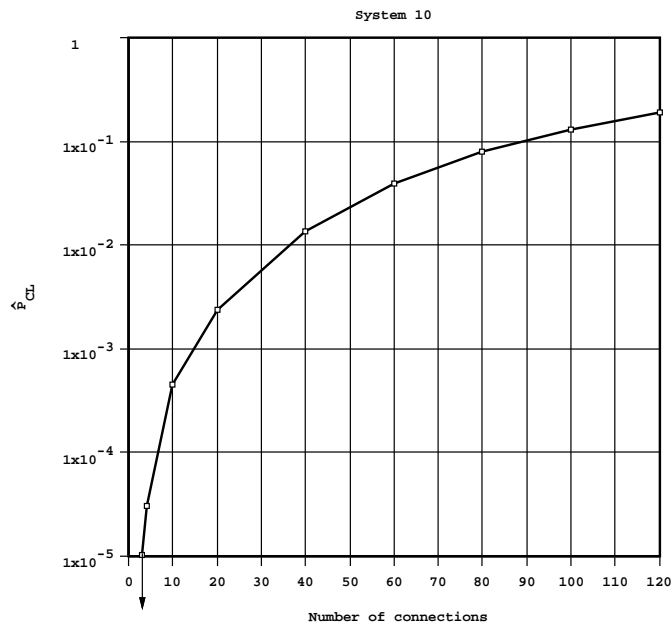


Figure 14: Cell loss probability as a function of the number of connections for system 10.

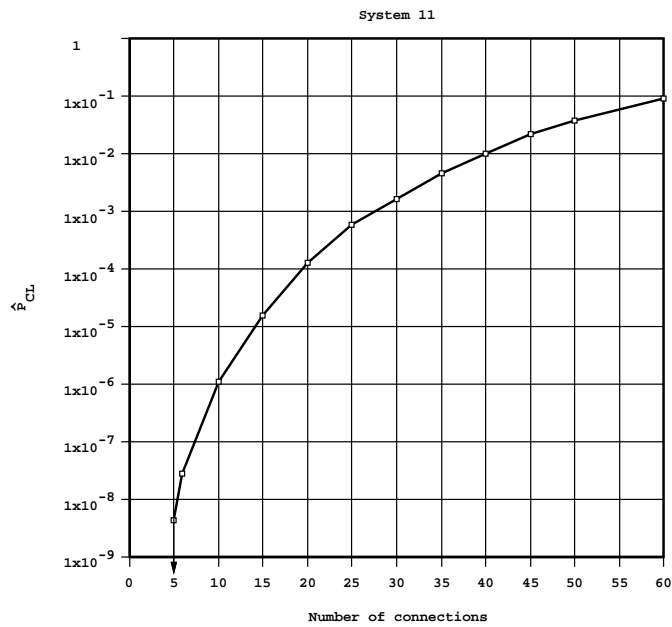


Figure 15: Cell loss probability as a function of the number of connections for system 11.

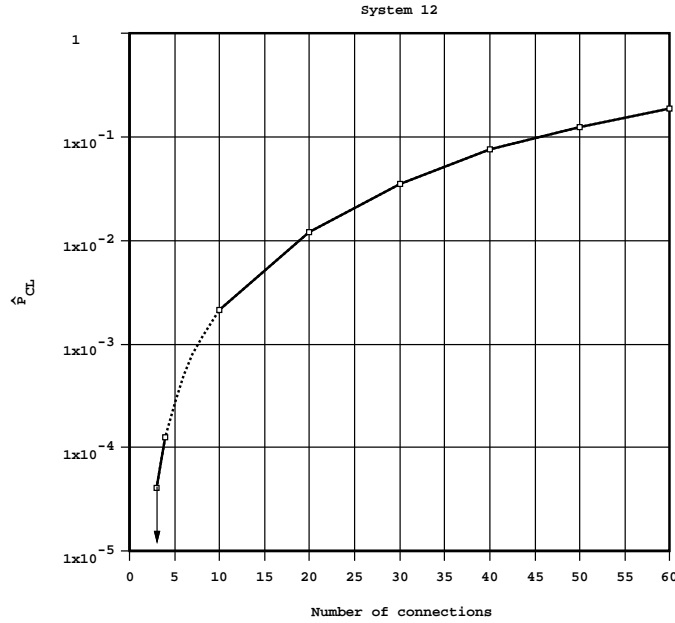


Figure 16: Cell loss probability as a function of the number of connections for system 12 (The dashed portion of the curve is a cubic spline interpolation).

N_C , conns.	\hat{P}_{CL}	Comment
3870	4.91×10^{-2}	$\epsilon = 0.01$
3500	1.57×10^{-2}	$\epsilon = 0.04$
3200	4.47×10^{-3}	$\epsilon = 0.05$
2900	7.11×10^{-4}	$\epsilon = 0.14$
2600	6.96×10^{-5}	$\epsilon = 0.37$
2400	6.56×10^{-6}	$\epsilon = 0.30$
2200	4.22×10^{-7}	$\epsilon = 1.00$
59	2.01×10^{-101}	Upper Bound
58	5.60×10^{-144}	Upper Bound
57	0	Exact

Table 11: Cell loss probability as a function of the number of connections for system 13.

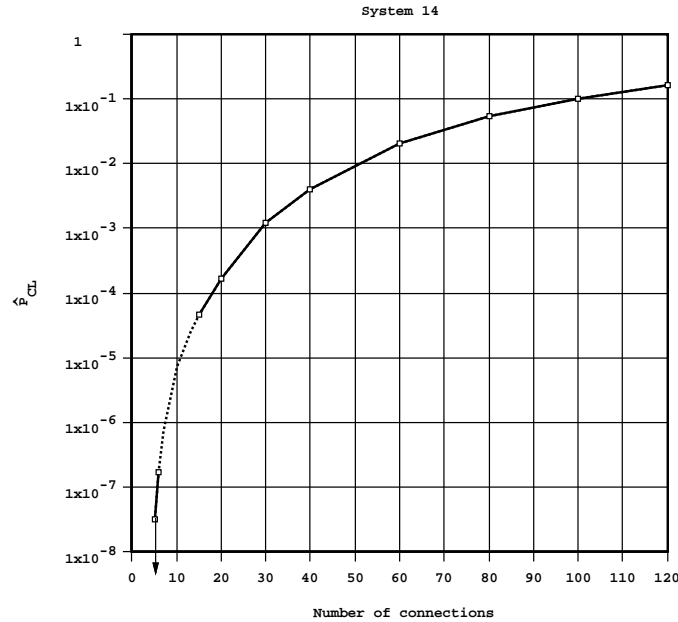


Figure 17: Cell loss probability as a function of the number of connections for system 14 (The dashed portion of the curve is a cubic spline interpolation).

For each performance curve, the bottom two points were obtained by simulation using IS, and the remaining points were obtained by conventional MC simulation. The accuracy of the MC estimates is measured using the precision, ϵ , which we define as

$$\epsilon = \frac{\hat{\sigma}(\hat{P}_{CL})}{\hat{P}_{CL}} \quad (59)$$

and is typically $\epsilon \leq 0.1$ for the systems where performance curves could be obtained.

Two types of discontinuities occur in the performance curves. In (26), we determined the minimum number of connections, $N_{C_{min}}$, needed for at least one cell loss at the output cell buffer due to overflow. Thus, for $N_C < N_{C_{min}}$, $P_{CL} = 0$. The downward pointing arrow at the bottom of each performance curve is meant to signify that P_{CL} is zero below $N_{C_{min}}$ connections, thus there is a point on the performance curve where cell losses begin to occur and below this point cell losses do not occur. This is an important result because the CAC algorithm can quickly determine that a new connection can be admitted to the network since no cell losses will occur, thus satisfying any QoS standard.

System	$\hat{P}_{CL}(N_{C_{\min}})$	95% CI	N_R	$\hat{P}_{CL}(N_{C_{\min+1}})$	95% CI	N_R
2	1.00×10^{-13}	$(7.55 \times 10^{-14}, 1.25 \times 10^{-13})$	4000	2.09×10^{-12}	$(0.0, 6.41 \times 10^{-12})$	20000
3	NS			NS		
4	4.75×10^{-7}	$(3.68 \times 10^{-7}, 5.83 \times 10^{-7})$	20	3.25×10^{-6}	$(1.64 \times 10^{-6}, 4.87 \times 10^{-6})$	40
5	4.36×10^{-5}	$(3.23 \times 10^{-5}, 5.39 \times 10^{-5})$	50	1.31×10^{-4}	$(8.88 \times 10^{-5}, 1.73 \times 10^{-4})$	50
6	7.89×10^{-24}	CFS		NS		
7	4.61×10^{-10}	$(3.85 \times 10^{-10}, 5.36 \times 10^{-10})$	10000	1.85×10^{-8}	$(8.39 \times 10^{-9}, 2.86 \times 10^{-8})$	10000
8	NS			NS		
9	2.96×10^{-11}	$(2.09 \times 10^{-11}, 3.84 \times 10^{-11})$	1000	6.09×10^{-9}	$(1.95 \times 10^{-9}, 1.02 \times 10^{-8})$	1000
10	1.01×10^{-5}	$(5.35 \times 10^{-6}, 1.49 \times 10^{-5})$	20	3.01×10^{-5}	$(1.05 \times 10^{-5}, 4.98 \times 10^{-5})$	50
11	4.23×10^{-9}	$(3.83 \times 10^{-9}, 4.63 \times 10^{-9})$	1000	2.70×10^{-4}	$(2.39 \times 10^{-8}, 3.00 \times 10^{-8})$	1000
12	4.06×10^{-5}	$(3.30 \times 10^{-5}, 4.82 \times 10^{-5})$	200	1.27×10^{-4}	$(9.36 \times 10^{-5}, 1.60 \times 10^{-4})$	200
13	NS			NS		
14	3.18×10^{-8}	$(1.07 \times 10^{-8}, 5.29 \times 10^{-8})$	50	1.68×10^{-8}	$(2.59 \times 10^{-8}, 3.10 \times 10^{-8})$	100

Table 12: IS estimates and confidence intervals for the experimental examples (CFS = Closed Form Solution, NS = Not Simulated).

The guaranteed cell loss condition in (28) also creates a discontinuity in that there is a point on the curve where a ‘jump’ in the cell loss probability is possible. The discontinuity related to the load being greater than the available service is only noticeable for system 9 in Figure 13, where we see a ‘jump’ of 5 orders of magnitude in \hat{P}_{CL} between $N_C = 12$ and $N_C = 13$ connections which is associated with the guaranteed cell loss floor that begins at $N_C = 13$ connections.

Confidence intervals for the estimates of P_{CL} for $N_{C_{\min}}$ and $N_{C_{\min+1}}$ connections for the experimental examples using the method from [14] given in (23) are shown in Table 12. Because the experimental examples we considered are capable of large values for the degrees of freedom, L_{\max} , the χ^2 distribution can give negative values for the lower confidence interval, such as in system 2. In such cases, we use zero as the lower confidence interval.

For systems 2, 6, 12, and 14 in Figures 8, 11, 16, and 17, the use of this method resulted in a ‘gap’ between the points found by simulation using IS and the points found by conventional MC simulation. The existence of this ‘gap’ is a present limitation of this method which is still under study. In systems where this ‘gap’ is small, such as systems 12 and 14 in Figures 16 and 17, it can be bridged using cubic spline interpolation, since we expect a smooth curve in the ‘gap’ region. Nevertheless, the method still provides a means of estimating P_{CL} in the range 10^{-6} to 10^{-12} for systems 4, 7, 9, and 11 in Figures 9, 12, 13, and 15.

System	$\lfloor R_{\text{net}} \rfloor, N_{C_{\min}}$	$\lfloor R_{\text{net}} \rfloor, N_{C_{\min+1}}$
1	CFS	NS
2	2.3×10^9	1.9×10^6
3	NS	NS
4	2.1×10^4	9.7×10^3
5	3.4×10^3	1.2×10^4
6	CFS	NS
7	2.8×10^5	2.7×10^3
8	NS	NS
9	1.3×10^7	5.9×10^4
10	6.2×10^3	3.9×10^4
11	2.4×10^6	1.4×10^6
12	1.5×10^3	4.9×10^3
13	NS	NS
14	4.8×10^5	9.8×10^4

Table 13: Lower bound of the improvement obtained from $1/W$ for the experimental examples (CFS = Closed Form Solution, NS = Not Simulated).

The lower bound on the improvement, $\lfloor R_{\text{net}} \rfloor$, obtained by using IS for $N_{C_{\min}}$ and $N_{C_{\min+1}}$ connections for the experimental examples is shown in Table 13. The improvement obtained for the experimental systems that could be simulated ranges from 3 to 9 orders of magnitude.

From the first column of Table 14 we see that $\lfloor P_{\text{CL}} \rfloor$ is typically two orders of magnitude above \hat{P}_{CL} , as previously mentioned. The \hat{P}_{CL} values for the experimental systems are shown in Table 14, along with the simulation effort, $E(N_{C_{\min}})$, ordered from smallest to largest, and the simulation time required for $N_R = 100$ steady-state periods for $N_{C_{\min}}$ connections on a DECStation 5000/25. Unlike other applications of IS in a network framework [6], the application of IS here does not increase the period length or the number of arrivals in a period, therefore the length of the simulation time of a single steady-state period for conventional MC simulation is the same as it is for simulation using IS.

By examining the value of E , we can see that if $N_{C_{\min}}$ and T are large and the improvement does not significantly increase P_{CL} to \hat{P}_{CL} , then this method is unable to estimate P_{CL} . There is a gap of two orders of magnitude between the system with the highest E that we were able to simulate, system 2, and the system with the lowest E that we were unable to simulate, system 3, where systems 1 and 6 were not considered because closed form solutions existed for these systems. Thus, computation of E can indicate how amenable a particular system is

System	$[P_{CL}(N_{C_{\min}})]$	$\tilde{P}_{CL}(N_{C_{\min}})$	$E(N_{C_{\min}})$	$t_{N_R=100}(N_{C_{\min}})$, seconds
12	1.29×10^{-2}	2.0×10^{-1}	3.8×10^5	200.36
5	2.88×10^{-3}	9.7×10^{-2}	4.7×10^5	124.82
10	3.24×10^{-3}	2.0×10^{-1}	7.5×10^5	360.58
14	2.34×10^{-6}	1.1×10^{-2}	2.8×10^6	78.33
11	6.37×10^{-7}	1.5×10^{-2}	4.2×10^6	149.15
7	6.39×10^{-8}	1.8×10^{-4}	9.9×10^6	4.79
4	1.35×10^{-5}	2.8×10^{-3}	1.1×10^7	80.52
9	1.92×10^{-9}	2.5×10^{-4}	4.4×10^7	28.41
2	9.35×10^{-12}	2.1×10^{-4}	1.3×10^8	58.25
6	7.89×10^{-24}	CFS	CFS	2.58
1	5.64×10^{-40}	CFS	CFS	5.71
3	2.09×10^{-23}	4.0×10^{-7}	1.7×10^{10}	13.58
8	1.75×10^{-38}	8.8×10^{-9}	6.0×10^{12}	110.03
13	5.60×10^{-144}	3.0×10^{-21}	2.0×10^{26}	1093.69

Table 14: Upper bound of cell loss probability, equivalent MC simulation of cell loss probability, simulation effort, and time in seconds for $N_R = 100$ steady-state periods T_{SS} for the experimental examples (CFS = Closed Form Solution).

to simulation, once a baseline has been determined for a particular ATM switch architecture and the computer on which the simulation is being performed.

6 Conclusion

We have developed a simulation framework that describes the traffic entering the network by the standardized ATM connection traffic descriptors used to make connection admission control decisions. We have also described how IS can be applied within this framework to the problem of estimating the cell loss probability due to buffer overflow, including an analytical solution for the improvement. For the experimental systems considered, this method resulted in 3 to 9 orders of magnitude improvement in efficiency compared to conventional Monte Carlo simulation. In addition, we have derived upper and lower bounds on the cell loss probability that can be used to determine in advance of simulation whether this method can be applied effectively to a given system. This method of efficient simulation is suitable to the design and testing of even the most complicated ATM switches and connection admission control algorithms planned for use in ATM networks.

7 Acknowledgements

The authors would like to thank Brad Makrucki of BellSouth for suggesting this problem and his helpful feedback while researching this topic, as well as Ahmet A. Akyamac of the Center for Communications and Signal Processing at North Carolina State University for his many helpful discussions of this material.

References

- [1] ATM User-Network Interface Specification, Version 3.0, September, 10 1993. The ATM Forum.
- [2] S. Parekh and J. Walrand. A Quick Simulation Method for Excessive Backlogs in Networks of Queues. *IEEE Trans. Automat. Contr.*, AC-34(1):54–66, Jan. 1989.
- [3] J. S. Sadowsky and J. A. Bucklew. On Large Deviation Theory and Asymptotically Efficient Monte Carlo Estimation. *IEEE Trans. Inform. Theory*, IT-36(3):579–588, May 1990.
- [4] Q. Wang and V. S. Frost. Efficient Estimation of Cell Blocking Probability for ATM Systems. *IEEE/ACM Trans. on Networking*, 1(2):230–235, 1993.
- [5] P. Heidelberger. Fast Simulation of Rare Events in Queueing and Reliability Models. In *Proc. of Performance '93*, Rome, Italy, October 1993.
- [6] M. Devetsikiotis, W. Al-Qaq, J. A. Freebersyser, and J. K. Townsend. Fast Simulation of Queueing Networks Using Stochastic Gradient Techniques and Importance Sampling. Submitted to *IEEE/ACM Trans. on Networking*.
- [7] H. Perros. A Literature Review of Call Admission Algorithms. Technical Report TR-94/8, Center for Comm. and Signal Processing, North Carolina State University, April 1994.

- [8] A. W. Berger and A. E. Eckberg. A B-ISDN/ATM Traffic Descriptor, and its Use In Traffic and Congestion Control. In *Proc. IEEE Global Telecom. Conf., GLOBECOM '91*, Phoenix, AZ, May 1991.
- [9] T. Murase, H. Suzuki, S. Sato, and T. Takeuchi. A Call Admission Control Scheme for ATM Networks Using A Simple Quality Estimate. *IEEE J. Select. Areas in Commun.*, 9(9):1461–1470, Dec. 1991.
- [10] Y. Miyao. A Call Admission Control Scheme in ATM Networks. In *Proc. IEEE Int. Conf. Commun., ICC '91*, Denver, CO, Dec. 1991.
- [11] B. Jabbari and F. Yegenoglu. An Upper Bound for Cell Loss Probability of Bursty Sources in Broadband Packet Networks. In *Proc. IEEE Int. Conf. Commun., ICC '91*, Denver, CO, Dec. 1991.
- [12] H. Saito. Call Admission Control in an ATM Network Using Upper Bound of Cell Loss Probability. *IEEE Trans. on Commun.*, 40(9):1512–1521, Sept. 1992.
- [13] M. C. Jeruchim. Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems. *IEEE J. Select. Areas Commun.*, SAC-2(1):153–170, Jan. 1984.
- [14] R. Z. Gold. Tests Auxiliary to χ^2 Tests in a Markov Chain. *Ann. Math. Statist.*, Vol. 34:56–74, 1963.