Approximate Analysis of an ATM Switching System with Bursty Arrivals and Finite Capacity

Y. Frank Jou

Arne A. Nilsson

Fuyung Lai

Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University

TR-92/15
September 1992

Approximate Analysis of an ATM Switching System with Bursty Arrivals and Finite Capacity

Y. Frank Jou and Arne A. Nilsson
Center for Communications and Signal Processing
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, N.C. 27695-7914

Fuyung Lai
IBM
V57/B660 P.O. Box 12195
Research Triangle Park, N.C. 27709

Abstract

This paper is concerned with both the mean delay and the probability of cell loss that bursty arrivals incur in an ATM switching system which can be modeled as a finite capacity polling system with nonexhaustive cyclic service. The arrival process to each input port of the system is assumed to be bursty and is modeled by an Interrupted Bernoulli Process(**IBP**). A practical polling system with finite capacity as the one we are dealing here does not lend itself to an exact solution. In this paper, we introduce a simple yet effective approach to provide an analytical approximation. We demonstrate that the cell loss probabilities obtained from this model follow the results from simulation closely when the loss rate is heavy. It is shown that the cell loss rate from the analytical model gives us a conservative estimate and can serve as an upper bound to the loss probability obtained from simulation. We also conclude that this model provides a highly accurate estimate for cell delay under the ATM environment where the cell loss rate is required to be less than 10^{-9}.

1 Introduction

The emerging needs for high speed communications and the promise of the technologies to support these services in an integrated fashion has generated a lot of interest in research, development, and standardization of broadband integrated networks. Among the transport and switching techniques for B-ISDN, the Asynchronous Transfer Mode (ATM) technique has shown to be the most promising solution. ATM is a packet oriented transfer mode based on statistical multiplexing in which the information is transported in short, fixed length blocks, referred to as cells, composed of a header and an information field. ATM provides the means for transporting different types of highly bursty traffic such as voice, video images and bulk files. The bandwidth flexibility, the capability to handle all services in a uniform way, and the possible use of statistical multiplexing are advantageous features of ATM. Furthermore, ATM enables a better utilization of the network given the existence of bursty sources.

Most of the architectures which have been proposed for an ATM switch are based on multi-stage interconnection networks. The switching elements in a multi-stage interconnection network may be buffered or unbuffered. In the unbuffered case, there may be buffers at the input ports of the switch, or at the output ports, or at both input and output ports. There have been several performance studies of such switch architectures (see for instance [1], [2], [3]). For the performance analysis of switches with buffered switching elements see [4], [5], and the references therein. The architecture of the ATM switch which we consider is shown in figure 1 [6]. It is constructed by connecting self-routing switching modules (SRMs) in a three-stage link configuration which is called a multi-stage self-routing network (MSRN). Each stage of MSRN consists of eight self-routing switching modules. Each module is an 8×8 crossbar switch which has a finite buffer associated with each crosspoint. The cells in each buffer will be transmitted in a cyclic order.

To study the performance of this SRM, we model it by a polling system with cyclic service. In the literature, multiqueue systems served by a single server have been the subject of numerous investigations (see [7], [8] and references therein). Various polling strategies

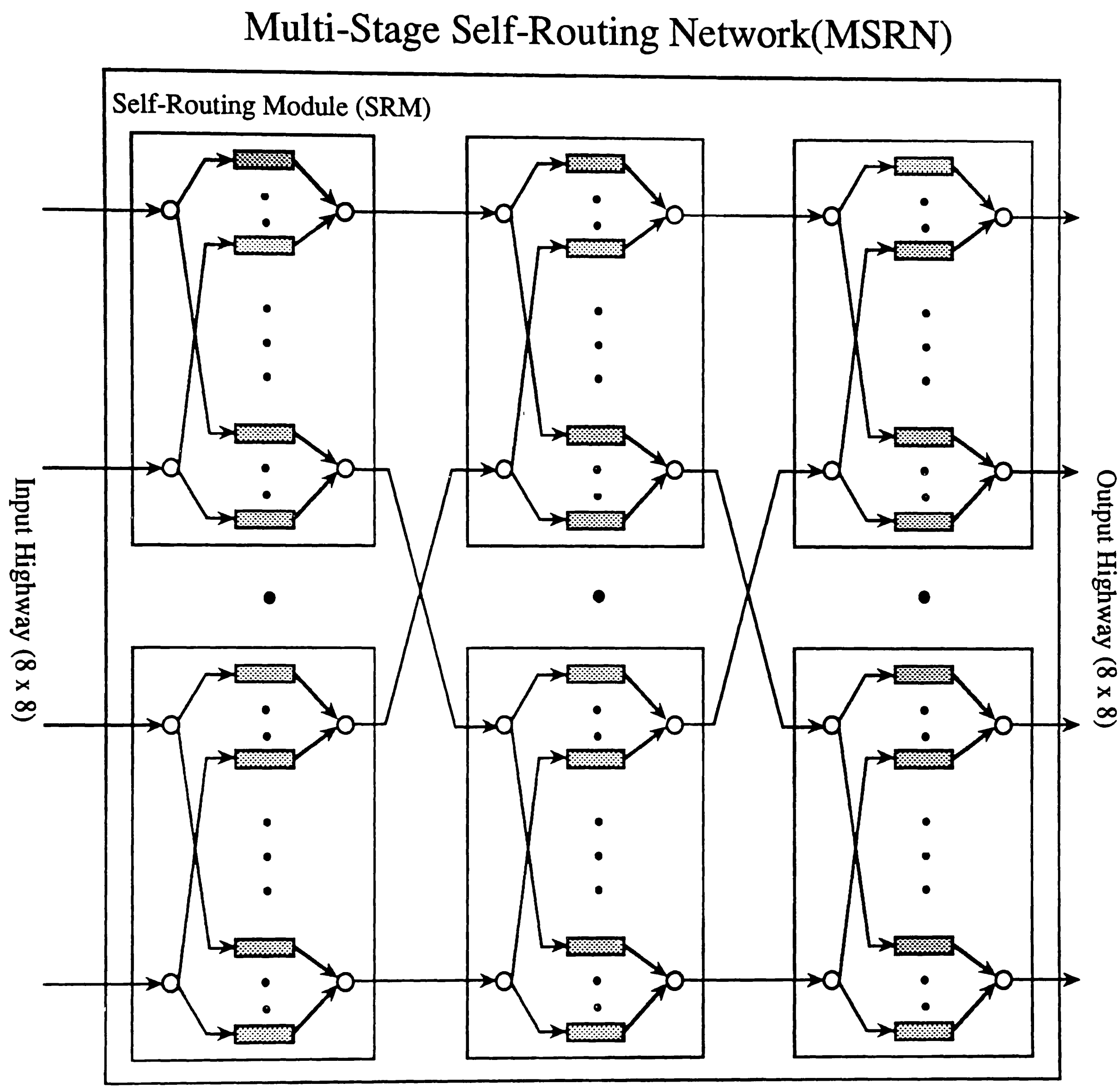


Figure 1: ATM Switching Architecture.

like cyclic or priority service and different types of service disciplines, e.g. exhaustive, gated, or limited service, have been considered. In most of these investigations, the input processes are assumed to be Poisson, and the queues of the polling system to have infinite capacity. In order to include more realistic modeling elements in the class of polling systems, we consider bursty arrival processes as inputs, and finite buffer capacity in the polling system.

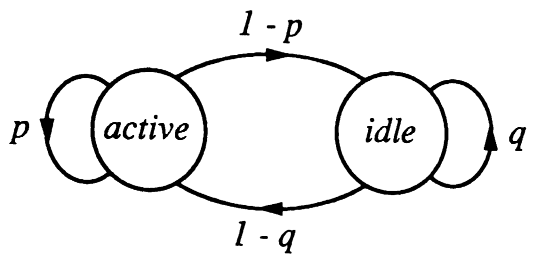
In this paper, we present a queueing model to compute the mean delay and cell loss probability incurred in the finite capacity polling system. This model will assume symmetric traffic load, zero switchover time, and 'limited - 1' service [9]. In Section 2 we describe in detail the model we propose. This approach will require analysis of the queue length distribution of the polling system and a multiple urn model with uniform occupancy distribution which are presented in Section 3 and Section 4, respectively. Section 5 describes how to match the output process in order to obtain the mean delay and loss probability that cells incur at the second and third stages of MSRN. Some numerical results validated by computer simulations are given in Section 6. Finally, Section 7 presents our conclusions.

2 Model description

In this section we describe in detail the switch architecture, the arrival process, and the queueing models which we propose.

2.1 Switch architecture

Figure 1 shows the configuration of the MSRN. The MSRN is constructed by connecting SRMs in a three-stage link configuration. In this configuration, there are multiple paths between a first-stage SRM and a third- stage SRM. This configuration allows the traffic flow to be routed efficiently between the input highway and the output highway, and reduces the delay in the switching network. Also, this configuration is inherently reliable because a faulty second stage SRM can be bypassed. Each stage of MSRN consists of eight self-routing switching modules. Each module is an 8×8 crossbar switch which has a finite buffer associated with each crosspoint. The SRMs consist of a cell distributor at each inlet, and FIFO buffers for temporarily storing cells in order to resolve outlet contention at each outlet. Cells are assigned to paths (links between SRMs) so that each link carries an equal amount of traffic.



$$\pi_A = \frac{1-q}{2-p-q}$$

$$\pi_I = \frac{1-p}{2-p-q}$$

Figure 2: The Markov chain for an IBP

2.2 Arrival process

Since most of the traffic sources that an ATM network supports are bursty, a Poisson process may no longer be suitable for describing the network traffic. For instance, interactive data and compressed video generate cells at a near-peak rate for a very short period of time. Immediately, following a near peak rate such a source may become inactive, thus generating no cells. With this scenario, the usual approximation of arrival process by a Poisson process will fail to capture the bursty nature of input traffic and may result in a quite dramatic error in the performance estimation. Kuehn[10] has shown that the system behavior is much more sensitive to arrival processes than to service process. Therefore, we propose to use the IBP which is the generalization of the Interrupted Poisson Process in a discrete time system.

A simple IBP is governed by a Markov chain with two states, an active state and an idle state. The duration of stay in these two states are geometrically distributed. Arrivals occur in a Bernoulli fashion with parameter α when the process is in the active state. No arrivals occur if the process is in the idle state. Given that the process is in the active state (or idle state) at slot i, it will remain in the same state in the next slot $i+1$ with probability p (or q), or will change to the idle state (or active state) with probability $1-p$ (or $1-q$). The transitions between the active and idle states are shown in figure 2, where π_A and π_I are the probabilities that the Markov chain is in the active and idle states, respectively. During the active state, a slot contains a cell with probability α. Here we assume α equals 1 which will generate the most bursty traffic.

Letting t be the interarrival time of a cell, it can be shown [11] that the z-transform of the probability distribution of the interarrival time $A(z) = E\{z^t\}$ is

$$A(z) = \frac{z\alpha[p + z(1-p-q)]}{(1-\alpha)(p+q-1)z^2 - [q + p(1-\alpha)]z + 1}.$$

The mean interarrival time $E\{t\}$ and the squared coefficient of variation of the time between successive arrivals, C^2 are as follows:

$$\begin{aligned} E\{t\} &= \frac{2-p-q}{\alpha(1-q)} \\ C^2 &= \frac{Var(t)}{E\{t\}^2} \\ &= 1 + \alpha\left[\frac{(1-p)(p+q)}{(2-p-q)^2} - 1\right]. \end{aligned}$$

The average arrival rate, i.e. the probability that a slot contains a cell, λ is

$$\lambda = \frac{\alpha(1-q)}{2-p-q}.$$

By varying p and q, we can have different traffic loads and at the same time change the burstiness of the arrival process.

2.3 Queueing models

Based on the structure of SRM, we evaluate its performance by using a multiqueue system as shown in figure 3. Under the assumption of symmetric traffic, the arrival processes to the multiqueue system will be characterized also as IBPs with the same parameters of p and q as of the original arrival processes. However, since every original arrival process branches to eight possible outlets, the parameter α of the arrival processes to the polling system will only be one eighth.

Instead of considering the queue lengths of the multiqueue system individually, we will call the distribution of the total number of cells in the polling system as the aggregate

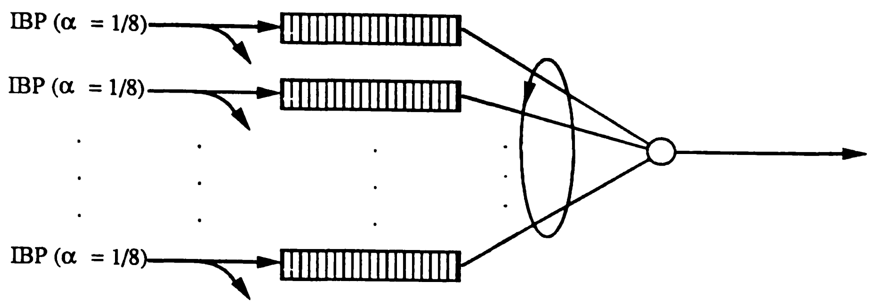


Figure 3: Multiqueue System Served by a Single Server.

queue length distribution. In order to obtain this aggregate queue length distribution, it is necessary to consider the blocking effect due to finite buffer space.

We model the queues in the multiqueue system as multiple urns which have the same limited capacity. Given the number of cells waiting in the system, it is assumed that the occupancy of each queue is independent from each other and the cells are uniformly distributed in any queue, i.e., each position in a queue is equally likely to be occupied. From this model we can compute the weighting of the cell occupancy configuration which could cause cell loss, and the cell loss probability will follow. Notice that given R cells in the system, the occupancy of these cells in reality will more likely be evenly distributed among these queues because that in general the server will visit the queues with longer queue sizes more frequently than the queues with shorter queue sizes. Therefore, given the number of cells in the polling system exceeding a single queue capacity, the occupancy configuration which has at least one full queue is less likely to occur. Hence, the assumption of uniform occupancy will give us a conservative estimate which can serve as an upper bound for cell loss probability of the polling system.

With this model, we will be able to obtain the weighting of the cell occupancy configuration which could cause cell loss. From this weighting, we can establish the transition matrix which allows us to compute the aggregate queue length distribution. The mean delay can therefore be found through Little's result [12].

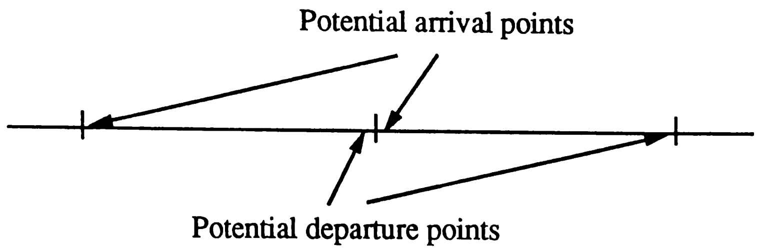


Figure 4: Potential arrival and departure points.

3 Aggregate queue length distribution

In our model, we assume that arrivals can only occur at the beginning of each slot and departures leave the system at the end of each slot. This arrangement, as illustrated in figure 4, is called an early arrival system according to Hunter [13]. During a slot period, one cell may arrive on each input link, and one cell may be transmitted given that the system is not empty. The state change, from active to idle or vice versa, only occurs at the slot point.

Next, the cell arrival process is analyzed for a cell arriving from all inputs in a slot time. We assume that there are N queues in the multiqueue system. The number of active input links K and arrival cells M in a unit time can both vary from zero to N. The probability that the number of active input links is K is

$$P_K(k) = \binom{N}{k} \pi_A^k \pi_I^{N-k},$$

where π_A and π_I denote the probability that an input link is in an active or idle state.

If the number of active input links is k, then m cells will arrive in a unit time with probability

$$P_{M|K}(m|k) = \binom{k}{m} \alpha^m (1-\alpha)^{k-m}.$$

The state transition probability of having k' active lines in a slot given k active lines in the previous slot is given by

$$P_{K'|K}(k'|k) = \sum_{j=0}^{k} \binom{k}{j} p^j (1-p)^{k-j} \binom{N-k}{k'-j} q^{N-k-k'+j}(1-q)^{k'-j}. \quad (1)$$

In order to describe the blocking effect, we need to define a conditional probability $P(m"|m', qsize)$ as

$P\{$ *m" cells accepted* $|$ *m' cells arrived and qsize cells in the system before arrivals* $\}$.

This probability will be further described and computed by a multiple urn model which is discussed in the next section.

Now, we define a two dimensional state variable (K, Q) such that the queue length becomes Q as the result of having M' cells arrive and $M"$ cells accepted in a slot given that K input lines are active. The state probability $P_{K,Q}(k, qsize)$ can be obtained by a numerical solution of the following steady state equations:

$$P_{K,Q}(k', qsize') = \sum_{k=0}^{N} \sum_{m'=0}^{k'} \sum_{qsize=0}^{Q_m} P_{K,Q}(k, qsize) P_{K'|K}(k'|k) P_{M|K}(m'|k') P(m"|m', qsize"), \quad (2)$$

$$\sum_{k=0}^{N} \sum_{qsize=0}^{Q_m} P_{K,Q}(k, qsize) = 1,$$

where Q_m is the total capacity of the multiqueue system. In equation (2), $qsize"$ and $qsize'$ are given by $qsize" = max(qsize - 1, 0)$ and $qsize' = m" + qsize"$, respectively. From $P_{K,Q}(k, qsize)$, we can sum over K and find the queue length distribution $P_Q(qsize)$. Since we have assumed zero switchover time in the system, the mean output rate λ_{out} can be determined as

$$\begin{aligned} \lambda_{out} &= 1 - P_Q(0) \\ &= \lambda_{in}\,(1 - P_{loss}). \end{aligned}$$

Therefore, the cell loss probability P_{loss} is obtained as

$$P_{loss} = 1 - \frac{\lambda_{out}}{\lambda_{in}}.$$

After we compute the mean queue length L, the mean delay W can be determined by using Little's result as

$$W = \frac{L}{\lambda_{out}}.$$

4 Multiple urn model

In order to compute the conditional probability $P(m"|m', qsize)$ defined in the last section, we propose a multiple urn model to find the total number of ways to place r indistinguishable balls into n distinguishable boxes given that the capacity of each box is limited to l. It is shown [14] that given $r \leq l$ (i.e. no capacity limit), the number of distinguishable distributions is

$$A_{n,r} = \binom{n+r-1}{r} = \binom{n+r-1}{n-1}.$$

Define $B_{n,r,k}$ as the number of ways of having at least k_i balls in i^{th} box (no limit). Then $B_{n,r,k}$ can be easily found as the following

$$B_{n,r,k} = \binom{n+r-k-1}{r-k} = \binom{n+r-k-1}{n-1}, \qquad k = \sum_{i=1}^{n} k_i.$$

From the principle of inclusion and exclusion we can obtain the total number of different configurations in this multiple urn model as

$$\begin{aligned} T_{n,r,l} &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{n+r-i(l+1)-1}{n-1} \\ &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} B_{n,r-i(l+1),0}, \end{aligned}$$

where $B_{n,r,0}$ is equal to $A_{n,r}$ which denotes the total number of solutions given no limit imposed on the capacity of boxes. $B_{n,r-(l+1),0}$ represents the number of configurations in which at least one of the n boxes contains no less than $l+1$ balls. This condition violates the limit and should be subtracted from $B_{n,r,0}$. The rest of the terms is just to compensate the over subtraction that $B_{n,r-(l+1),0}$ introduces.

We now proceed to find $C_{n,r,l}$, the number of ways of having at least one full box for this multiple urn model. If the total number of balls r is less than l, $C_{n,r,l}$ equals zero. For $l \leq r < 2l$, it is easy to verify that

$$C_{n,r,l} = nB_{n-1,r-l,0}.$$

Likewise, when $2l \leq r < 3l$

$$\begin{aligned} C_{n,r,l} &= \binom{n}{2} B_{n-2,r-2l,0} + \binom{n}{1}\left[T_{n-1,r-l,l} - \binom{n-1}{1} B_{n-2,r-2l,0}\right] \qquad (3) \\ &= \binom{n}{2} T_{n-2,r-2l,l} + \binom{n}{1}\left(T_{n-1,r-l,l} - C_{n-1,r-l,l}\right), \end{aligned}$$

where the first term on the right hand side of equation (3) denotes the number of ways to have two full boxes, and the second term gives the number of ways of having one full box. Proceeding in the same manner, we conclude that for $kl \leq r < (k+1)l$

$$C_{n,r,l} = \sum_{i=1}^{k} \binom{n}{i} \left(T_{n-i,r-i*l,l} - C_{n-i,r-i*l,l}\right), \qquad (4)$$

where $C_{n,r,l}$ is obtained recursively. Notice that the expression in the summation in equation (4) denotes the number of ways to have exactly i full boxes.

According to the definition of $P(m"|m', qsize)$, we now obtain the conditional probability as

$$P(m"|m',r) = \frac{\sum_{i=0}^{\lfloor \frac{r}{l} \rfloor} \frac{\binom{N-i}{m"}\binom{i}{m'-m"}}{\binom{N}{m'}} \binom{N}{i} \left(T_{N-i,r-i*l,l} - C_{N-i,r-i*l,l}\right)}{T_{N,r,l}}.$$

5 Matching the output process

For the analysis of the second and the third stages of MSRN, we need to characterize the departure process from the previous stage. This departure process will in turn become the arrival process to the next stage. We will use another IBP to match the output process where the parameter α is again assumed to be one. This assumption is reasonable due to the fact that we have zero switchover time in the system.

An IBP is defined by three parameters: α, p, and q. By assuming α as one, what we need to decide are the two transition probabilities, p and q. Since the switchover time is zero in the polling system, the output process stays in its busy state as long as the aggregate queue length is greater than zero. On the other hand, the output process becomes idle when the system is empty. Under this observation, we will be able to determine the transition probabilities p and q of the IBP if we can obtain the mean busy period and mean idle period of the output process.

Again we define a two dimensional state variable (K, Q) where K denotes the number of active input lines, and Q denotes the total number of cells in the system observed right after the arrival point. We are mainly interested in the states where the system is empty. The rest of the states where the queue length is greater than zero are aggregated and considered as one absorbing state. Hence, given N input lines in this multiqueue system, there are $N+1$ transient states and one absorbing state. This setup allows us to find the idle period by using the concept of random walk.

Consider the Markov chain X_n which denotes the state of the system at the time slot n given that the chain starts from a empty system. The transition probability matrix $\mathbf{P}$ is defined as follows:

$$
\begin{aligned}
P_{i,j} &= P_{K'|K}(j|i)\,(1-\alpha)^j, && 0 \le i,j \le N,\\
P_{i,N+1} &= 1 - \sum_{j=0}^{N} P_{i,j}, && 0 \le i \le N,\\
P_{N+1,j} &= 0, && 0 \le j \le N,
\end{aligned}
$$

$$P_{N+1,N+1} = 1,$$

where $P_{K'|K}$ was given in equation (1). We further define a submatrix of $\mathbf{P}$, $\mathbf{P_t}$ which consists of all the transition probabilities among the transient states (i.e. $\{P_{i,j}\}, 0 \leq i, j \leq N$). Let

$$\begin{aligned}
T &\equiv min\{n \geq 1, X_n = N+1\} \\
\nu_i &\equiv E\{T|X_0 = i\}, \qquad\qquad 0 \leq i \leq N \\
V &\equiv [\nu_0, \nu_1, \cdots, \nu_N] \\
\underline{\mathbf{e}} &\equiv [1, 1, \cdots, 1].
\end{aligned}$$

From the first step analysis, we have

$$V^T = \underline{\mathbf{e}}^T + \mathbf{P_t} V^T,$$

and therefore

$$V^T = (\mathbf{I} - \mathbf{P_t})^{-1} \underline{\mathbf{e}}^T.$$

The mean idle period I of the output process is then given as

$$I = E[T] = \sum_{i=0}^{N} \nu_i \; P(X_0 = i \mid qsize = 0).$$

Let B denote the mean busy period of the output process. We have

$$\lambda_{out} = \frac{B}{B + I},$$

and hence

$$B = \frac{\lambda_{out} \, I}{1 - \lambda_{out}}.$$

Finally, the two transition probabilities p and q can be easily obtained as

$$p = 1 - \frac{1}{B},$$

$$q = 1 - \frac{1}{I}.$$

6 Numerical results

In this section, we examine several configurations where the presented approximations are compared against the simulation results. The performance measures, mean delay, and cell loss probability are affected by the buffer capacity of the multiqueue system and the burstiness of the arrival processes. In terms of queue capacity, we present three cases where the buffer sizes are of 4, 8, and 16, respectively. As mentioned in section 2.1, there are 8 queues in the multiqueue system. In order to show the effect of the arrival burstiness, we vary the squared coefficient of variation (C^2) of the arrival processes from 1, 20, to 200.

These three kinds of burstiness represent three typical cases. When C^2 equals 1, we can regard this arrival process as Bernoulli. The burstiness of voice is represented by the case where C^2 equals 20. We use $C^2 = 200$ to show the burstiness of data traffic. Figure 5 shows the influence of arrival burstiness toward mean delays at the first stage of MSRN when each buffer capacity is 16. This figure clearly demonstrates the necessity of considering the bursty effect of the arrival processes under the ATM environment.

The results regarding cell loss probability are shown in figures 6 to 9. We do not include the cases which C^2 equals 1 because no cell loss was recorded in the simulation except for cases with heavy traffic loads. We observe that the cell loss incurred at the first stage of MSRN is much higher than at the second stage. This is especially the case when the buffer size or arrival burstiness is increased. This phenomenon is due to the filtering effect of the buffers at the first stage which absorbs and smoothes out the burstiness of the arrival traffic. It is observed from the simulation results that cells experience about the same delay and loss rate at the second and third stages of MSRN. Therefore, for the purpose of clarity, we only show the measures from the second stage and compare them against the results from the first stage.

From figures 6 to 9, we see that the analytical results for cell loss follow the results from simulation closely. As we discussed in section 2.3, the assumption of equal probability among the configurations of occupancy guarantees us a conservative approximation which can serve as an upper bound. It is observed that the higher the cell loss rate, the better the approximation. This is because that our assumption of equal probability is getting closer to the reality when the queue capacity is smaller or the traffic load is heavier.

Figures 10 to 13 show the mean delay time that the arrivals incur under different setups. It is recognized that the difference of delay between the approximation and simulation results is affected interactively by two factors, the discrepancy of cell loss between the two models and the level of cell loss probability itself. As we discussed previously, when the buffer capacity gets smaller, the cell loss rate gets higher but the relative error of cell loss between two models gets smaller. Compare figures 11, 13, and 5 where the queue capacities are 4, 8, and 16, respectively. If we only look at the worst cases where C^2 equals 200 and arrival rate is 0.9 at the first stage, the relative errors between these two models are 11.0%, 12.5%, and 9.4%. When the arrival rate drops to 0.8, the errors are 5.7%, 4.6% and 0.8%, respectively. This shows that the effects of the two factors mentioned above balance each other when the loss rate is relatively high.

On the other hand, even though the relative error of cell loss gets higher when the loss rate decreases, the level of the absolute error of loss nevertheless becomes insignificant. As we observe from figure 5, the effect of the level of cell loss dominates and the relative error of mean delay becomes smaller when the cell loss probability gets lower. Comparing among the same cases mentioned above with arrival rate down to 0.7, then the relative errors between these two models are 3.5%, 2.2%, and 0.5%, respectively. In fact, the analytical results in figure 5 fall in the confidence intervals of simulation results for different arrival rates up to 0.9 when C^2s are 1 and 20. We found out by experiment that the relative error of delay is less than one percent as long as the cell loss probability from simulation is smaller than 10^{-5}. Therefore, as far as the mean delay is concerned, our approach is well applicable in the ATM environment where the cell loss probability is required to be less than 10^{-9}

For both delay and cell loss, we show that the comparison between the analytic and simulation results at the second stage of MSRN follows the same trend as we discussed above for these two measures at the first stage. This comparison justifies the approach we used to match the output process.

7 Conclusion

A realistic polling system with finite capacity does not lend itself to an exact analysis. In this paper, we presented a simple yet effective approach to provide an analytical approximation. As for the arrival processes, we also take account of the bursty effect which is an essential feature under ATM environment. The strength of this analytical model, as discussed in the last section, is such that it is able to give good estimates when either the cell loss probability or mean delay is a major concern. The results from this approximation asymptotically approach the results from simulation when we consider either the cell loss rate with smaller queue capacity or the mean delay with bigger buffer capacity. The cell loss probability from this analytical model can serve as an upper bound to the simulation results. We also show that this model provides a highly accurate estimate for cell loss under the ATM environment where the required cell loss rate is much less than 10^{-5}.

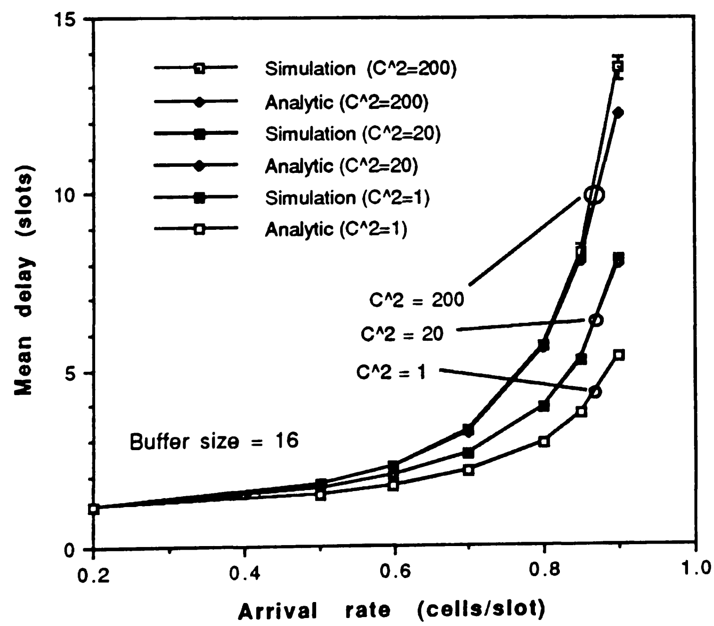


Figure 5: Mean delays w.r.t. different burstiness at the first stage of MSRN.

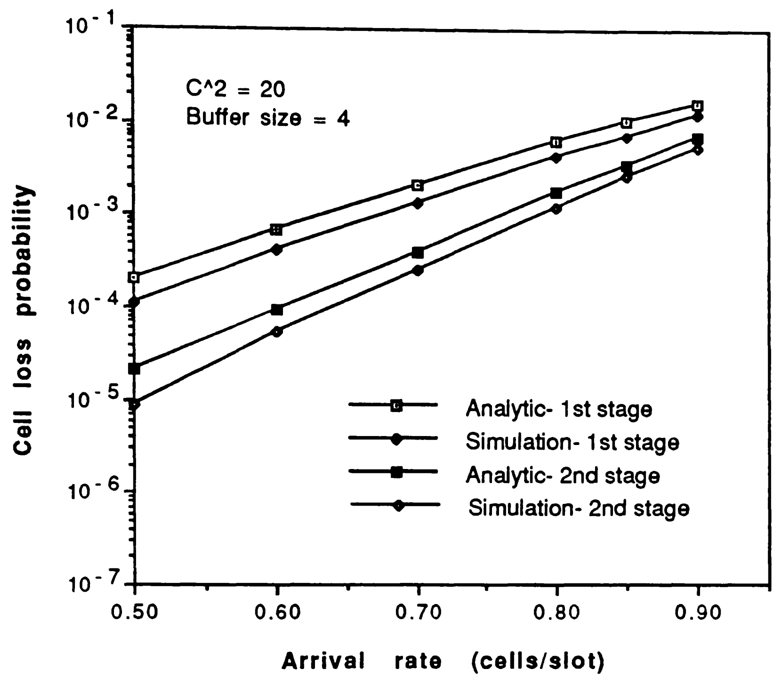


Figure 6: Cell loss probabilities incurred at the first and second stages of MSRN when $C^2 = 20$, queue capacity = 4.

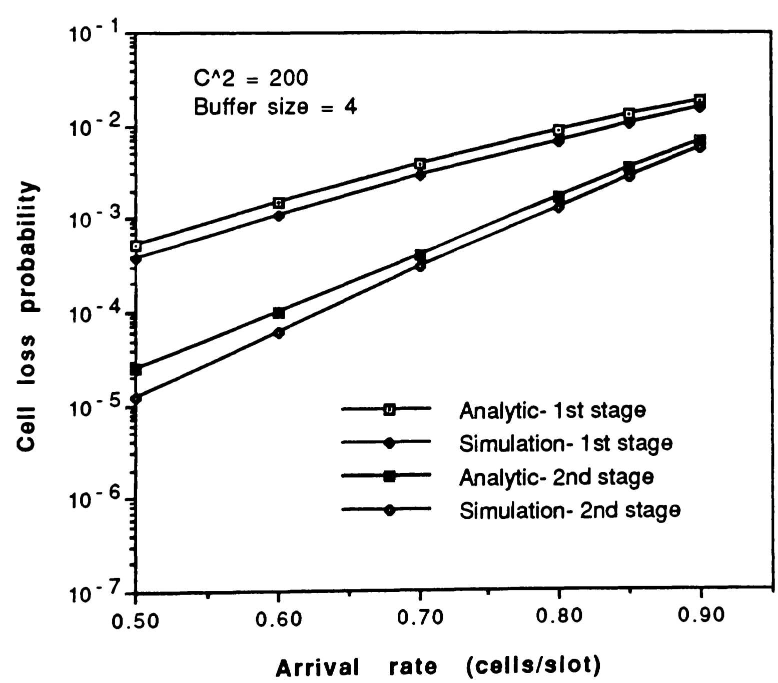


Figure 7: Cell loss probabilities incurred at the first and second stages of MSRN when $C^2 = 200$, queue capacity = 4.

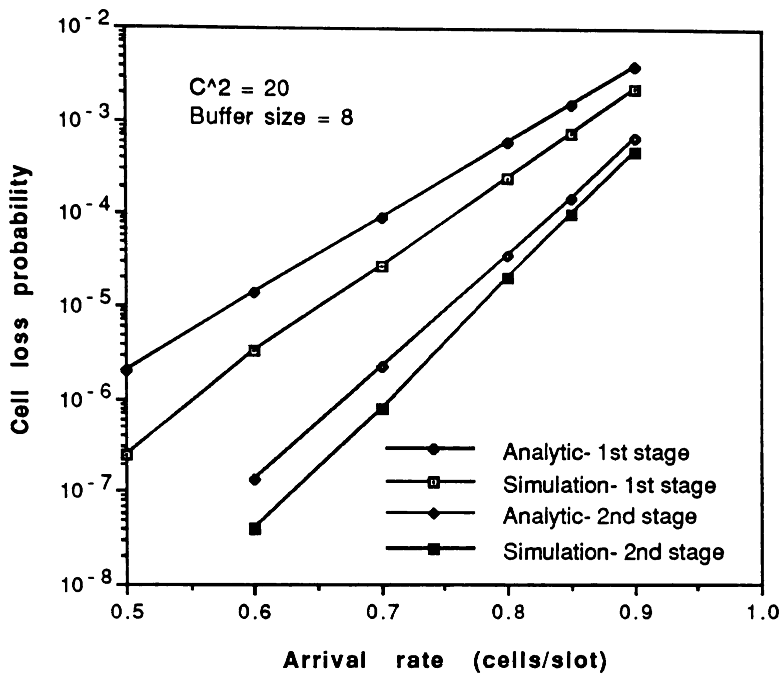


Figure 8: Cell loss probabilities incurred at the first and second stages of MSRN when $C^2 = 20$, queue capacity = 8.

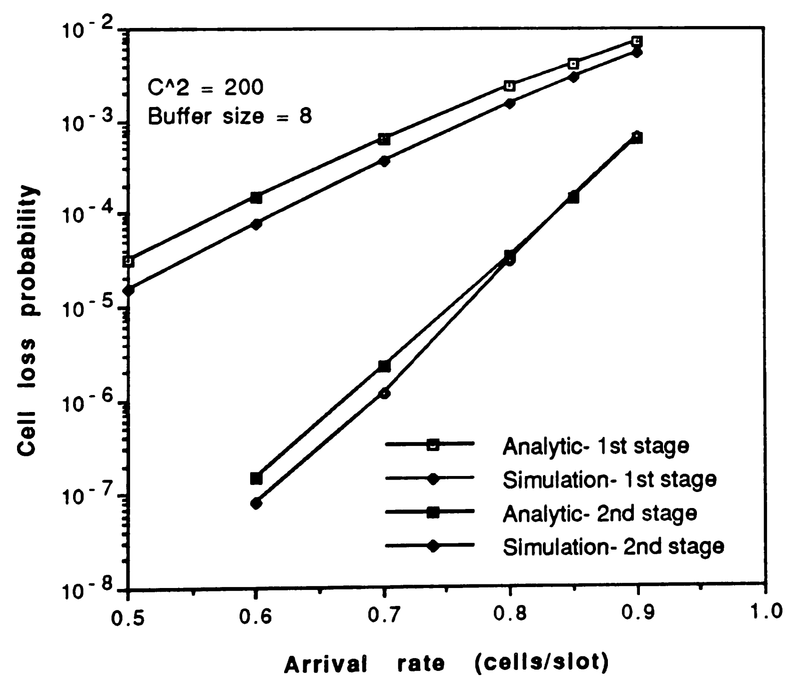


Figure 9: Cell loss probabilities incurred at the first and second stages of MSRN when $C^2 = 200$, queue capacity = 8.

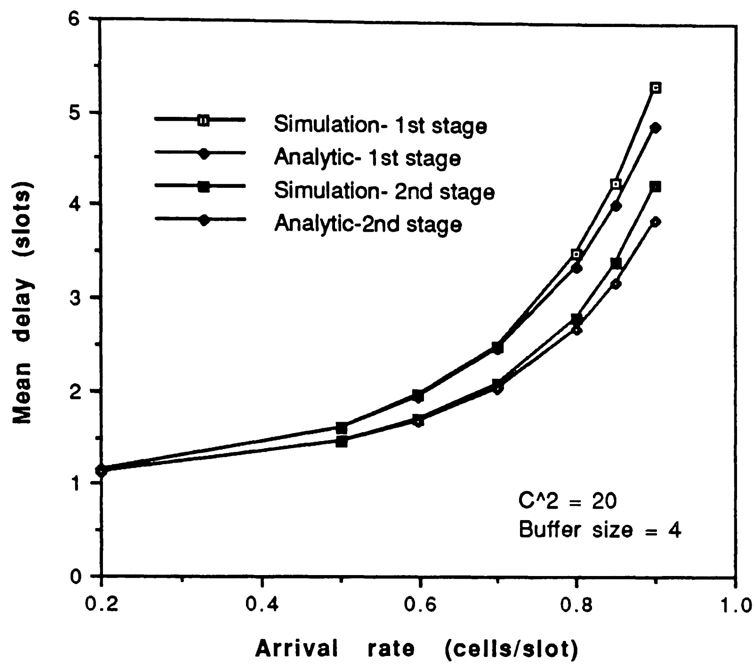


Figure 10: Mean delays incurred in the first and second stages of MSRN when $C^2 = 20$, queue capacity = 4.

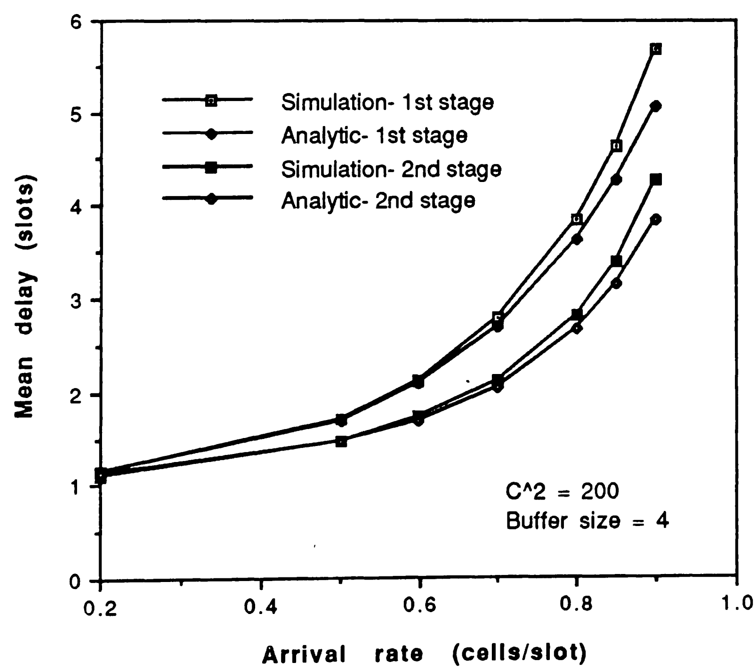


Figure 11: Mean delays incurred in the first and second stages of MSRN when $C^2 = 200$, queue capacity = 4.

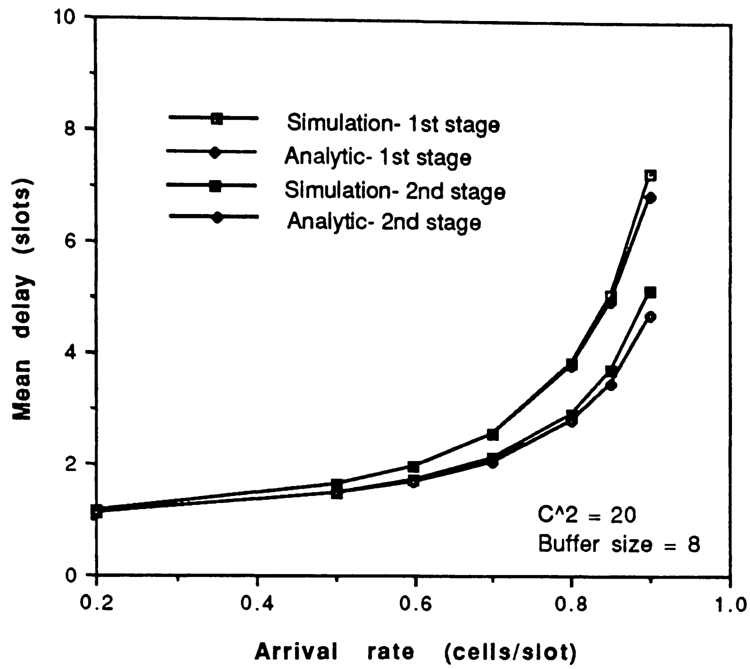


Figure 12: Mean delays incurred in the first and second stages of MSRN when $C^2 = 20$, queue capacity = 8.

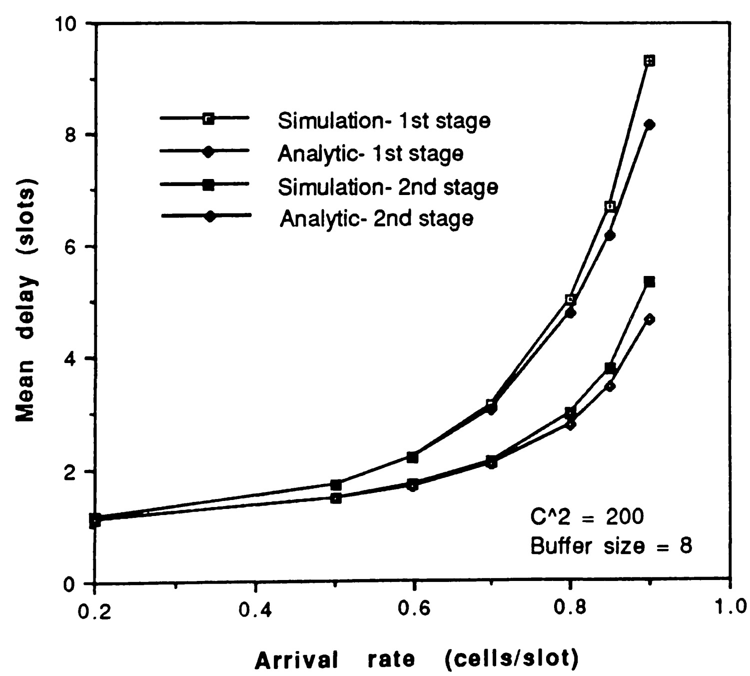


Figure 13: Mean delays incurred in the first and second stages of MSRN when $C^2 = 200$, queue capacity = 8.

References

[1] A. A. Nilsson, F.-Y. Lai, and H. G. Perros, "An approximate analysis of a bufferless N × N synchronous Clos ATM switch," *Proceedings, Canadian Conference on Electrical and Computer Engineering*, pp. 39.1.1–39.1.4, 1990.

[2] Y. Oie, T. Suda, M. Masayuki, and H. Miyahara, "Survey of the performance of non-blocking switches with FIFO input buffers," *IEEE International Conference on Communications*, vol. 2, pp. 737–741, 1990.

[3] M. G. Hluchyj and M. J. Karol, "Queueing in high-performance packet switching," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1587–1597, 1988.

[4] T. H. Theimer, E. P. Rathgeb, and M. N. Huber, "Performance analysis of buffered Banyan networks," *Transactions on Communications*, vol. 39, no. 2, pp. 269–277, 1991.

[5] H. Yoon K. Lee and M. Liu, "Performance analysis of multibuffered packet-switching networks in multiprocessor systems," *IEEE Transactions on Computers*, vol. 39, no. 3, pp. 319–327, 1990.

[6] K. Hajikano, K. Murakami, E. Iwabuchi, O. Isono, and T. Kobayashi, "Asynchronous transfer mode switching architecture for broadband ISDN - multistage self-routing switching," *IEEE International Conference on Communications*, vol. 2, pp. 911–915, 1988.

[7] Takagi, "Queueing analysis of polling models: an update," *Stochastic Analysis of Computer and Communication Systems*, pp. 267–318, H. Takagi (editor), Elsevier Science Publishers B.V. (North–Holland) Amsterdam, 1990.

[8] P. Tran-Gia and T. Raith, "Performance analysis of finite capacity polling systems with nonexhaustive service," *Performance Evaluation*, vol. 9, pp. 1–16, 1988.

[9] M. Lang and M. Bosch, "Performance analysis of finite capacity polling systems with limited-m service," *Proc. ITC-13*, vol. 14, pp. 731–735, 1991.

[10] P. J. Kuehn, "Multiqueue systems with nonexhaustive cyclic service," *B. S. T. J.*, vol. 58, no. 3, pp. 671–699, 1979.

[11] Fuyung Lai, *Performance evaluation of an ATM switch and error control schemes for high speed networks*, PhD thesis, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, February 1991.

[12] L. Kleinrock, *Queueing Systems, Vol. I: Theory.* New York, NY: John Wiley & Sons, Inc, 1975.

[13] J. J. Hunter, *Mathematical Techniques of Applied Probability, Vol. 2, Ch. 9.* New York, NY: Academic Press, Inc., 1983.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. I, 3rd ed.* New York, NY: John Wiley & Sons, Inc, 1968.