Performance of Bus Allocation Policies for a Shared-Medium ATM Switch under Bursty Arrivals and Correlated Destinations

A. O. Zaghloul
H. G. Perros
I. Viniotis

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

TR-93/13
July 1993

Performance of Bus Allocation Policies for a Shared-Medium ATM Switch under Bursty Arrivals and Correlated Destinations

A. O. Zaghloul [1] , H. G. Perros [2] and I. Viniotis [3]

[1] IBM Corporation
Department C15
RTP, N.C. 27709
U S A

[2] Department of Computer Science, and
Center for Communications and Signal Processing
North Carolina State University
Raleigh, NC 27695-8206
U S A

[2] Department of Electrical and Computer Engineering, and
Center for Communications and Signal Processing
North Carolina State University
Raleigh, NC 27695-8206
U S A

Abstract

In this paper, we study, via simulation, the performance of different bus allocation policies for a generic shared-medium ATM switch with input and output queueing. The policies differ in the amount of information used by the bus arbitrator to decide which input queue to serve next. Classical polling systems (e.g. Round-Robin) consider only the state of the input queues. This may waste bus bandwidth on cells that are destined to output queues which are full. We consider a range of bus service policies that are based on the state of the input queues, or the state of both the input and the output queues. The performance of the switch is analyzed under realistic system characteristics such as bursty traffic, limited buffer size, asymmetric load conditions, and nonuniform destinations. Input traffic is modelled by an Interrupted Bernoulli Process (IBP). Furthermore, we assume that the destinations of the arriving cells within the same burst are correlated. Two types of destination correlations are introduced and their effect on the performance of the switch is analyzed. The performance measures considered are: cell loss probability, mean cell delay, and cell delay percentiles.

1.0 Introduction

The Asynchronous Transfer Mode (ATM) is the adopted transfer mode solution for broadband ISDN. ATM must be capable of efficiently multiplexing a large number of highly bursty sources, such as voice, video, and large file transfer. As a result, there is an increasing interest in the performance analysis of ATM based networks and in particular in ATM switches. There are several ATM switch architectures that have been proposed recently in the literature (see for instance Cidon et al [3], and Turner [11]). ATM switch architectures can be classified into three classes: ***shared-memory, shared-medium, and space-division*** . Tobagi [10] gives a detailed review of the different switch architectures and their implementation issues. In this study, we are interested in the analysis of the shared-medium architecture. In a shared-medium switch, all packets arriving on the input links are routed to the output links over a common high-speed medium such as a parallel bus. Each of the output links is capable of receiving all packets transmitted on the bus. The PARIS system described by Cidon et al [3] and the ATOM switch proposed by Suzuki et al [9], are two examples of such a switch architecture. One of the main advantages of the shared-medium architecture is the simplicity with which the multicast function is achieved since a transmitted packet over the shared-medium can be received by all output links simultaneously. Thus, the need to send multiple copies is eliminated. Also, priority among input links can be easily implemented within the bus arbitration policy. A shared-medium packet switch can provide packet queueing at the input links, the output links, or both the input and the output links (see Karol et al [6] and Pattavina [7]). A queueing model of a generic shared-medium packet switch with input and output queueing is shown in figure 1. The input queues are attached to the shared bus and contend for access when they have one or more packets to transmit. The order in which input queues are served is determined by the bus service policy. Cells arriving to a full input queue will be lost. Furthermore, if one of the output queues is full, the flow of customers destined to it may be momentarily stopped. This situation is commonly referred to as blocking. See Perros [8] for a comprehensive review of queueing networks with blocking.

Karlsson, Perros, and Viniotis [5] analyzed a high speed ATM bus under 4 different adaptive polling schemes that are based on status information sent by the input queues to the bus arbitrator. The objective of the study was to develop an adaptive polling scheme that equalizes the cell loss probability and the mean

waiting time at each input queue. However, they did not include the output queues in the analysis. Zaghloul and Perros [12] analyzed a generic shared-medium switch under three different bus service policies: TDM, Round-Robin, and Random. The analysis is based on the notion of decomposition whereby the switch is decomposed into smaller sub-systems. Iliadis [4] analyzed a single-stage NxN packet switch with input and output queueing. He studied the effect of the head of line arbitration policy on the average delay and the maximum throughput of the switch. Two policies, FCFS and LCFS, were considered in the analysis. Badran and Mouftah [1] analyzed an NxN switch under different bus selection criteria. They considered cell-based and input queue-based criteria. The cell-based criteria are the time of joining the input queues and the time of arrival to the top of the input queue. The queue-based criteria are the queue length and the queue age.

In this paper, we analyze the shared-medium switch shown in figure 1 under seven different bus service policies. The policies differ in the amount of information used in deciding the next input queue to be served. In order to model such a system accurately, we take into consideration the real system characteristics such as finite buffer capacities, bursty input traffic, nonsymmetric load conditions, and nonuniform destinations. Furthermore, we consider the case where the destinations of cells traversing the switch are correlated. Two types of destination correlation between successive cells within a single burst are introduced and analyzed.

The paper is organized as follows. In section 2, a description of the system under study is presented. Section 3 gives a description of the bus service policies considered in this paper. The results are given in section 4 and the conclusions in section 5 respectively.

2.0 The Queueing model under study

In this paper, we consider a single-stage N x N shared-medium switch with queue size (input and output) equal to m as shown in figure 1. Cells may be queued before switching at the inputs, if the output queue is full, as well as after switching at the outputs. A cell at the top of input queue i is destined for output queue j with probability d_{ij}, where $\sum_{j=1}^{N} d_{ij} = 1,\ i = 1,\ 2,\ldots,\ N$. Each input queue containing one or more packets arbitrates for the bus by activating the ***bus-request*** signal. A bus arbiter is used to select the input queue to

be served next according to the bus service policy. Once an input queue has been granted the bus, only one cell is forwarded to its destination output buffer. If the output buffer is full, then the cell will be blocked. The blocked cell remain at the top of its input queue until the queue is granted the bus again.

The bus bandwidth is N times the speed of a single input link, where N is the number of input links. The arrival process to each input queue is slotted, with a slot size equal to the transmission time of an ATM cell. The bus is also slotted, with a slot size equal to 1/N of the transmission time of an ATM cell. For instance, if N = 6, then there are 6 bus slots within the boundary of each arrival slot as shown in figure 2. We assume zero switch-over time which means that when the bus completes serving the current queue, it switches instantaneously and starts serving the next queue. This is normally accomplished by separating the control signals, which are used for bus arbitration, and data signals, which are used for data transmission. This scheme permits arbitration for the bus to be performed in parallel to data transmission; thus, justifying the zero switch-over assumption. Finally, the service time of each output link can be deterministic (equal to one arrival slot) or consists of n arrival slots, where n is geometrically distributed with parameter B_p.

2.1 The Arrival process

The arrival process to each of the input queues is assumed to be bursty and it is modelled by an Interrupted Bernoulli Process (IBP). That is, the incoming link into an input queue is slotted. Each slot is long enough to contain one cell. An arriving slot may or may not contain a cell. In an IBP, we have a geometrically distributed period during which no arrivals occur (idle state), followed by a geometrically distributed period during which arrivals occur in a Bernoulli fashion (active state). Given that the process is in the active state at slot t, it will remain in the active state during the next slot t + 1 with probability p, or it will change to the idle state with probability 1-p. If the process is in the idle state at slot t, it will remain in the idle state during the next slot t + 1 with probability q, or it will change to the active state with probability 1-q. During the active state, a slot contains a cell with probability α. The quantity α is also known as the peak bandwidth, i.e., the rate of arrivals during the active period. In [8], the average arrival rate, i.e., the probability that any slot contains a cell, is calculated as:

$$\rho = \frac{\alpha(1-q)}{2-p-q} \quad ,$$

and the squared coefficient of variation, C^2, of the time between successive arrivals is

$$C^2 = 1 + \alpha\left[\frac{(1-p)(p+q)}{(2-p-q)^2} - 1\right] .$$

Because of the finite buffer space at the input links, a cell arriving to a full input queue is lost. However, once a cell has been received, it will not be lost within the switch. In this paper, we assume that α is equal to 1 for each arrival process.

2.2 Correlated destinations

In this paper, we consider the case where the destinations of cells traversing the switch are correlated. We recognize the fact that in a wide area ATM network, as the cells from multiple sources are multiplexed, the destination correlation among successive cells should diminish as we move deeper into the network. On the other hand, as we move closer to the edges of the network, destination correlation becomes greater. This is due to the fact that packets are broken into smaller cells before transmission over the ATM network. As a result, we expect to see cells within the same burst coming from the same packet. The closer we are to the source or the destination, the higher the probability that successive cells have the same destination. Furthermore, in a Customer Premises Network (CPN) environment where the network comprises of few switches (see Lee Boudec [2]) , one can expect that there is less multiplexing of cells within the network. As a result, there is a high probability that most of the cells, within the same burst, are destined to the same output port.

In this paper we propose two types of destination correlation between successive cells within a single burst. We designate the two types of correlation by ***correlation type-1*** and ***correlation type-2*** respectively. The destinations of successive cells within a burst are correlated according to the matrix C. In correlation type-1,

the matrix element c_{ij} represents the probability that the current cell is destined to output port j given that the first cell of the present burst was destined to output port i. In correlation type-2, c_{ij} is the probability that the current cell is destined to output port j given that the previous cell was destined to output port i. Note that the destination of the first cell in the burst is randomly selected. The following matrix

$$C = \begin{vmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{vmatrix}$$

describes the cell destination correlation for a 3X3 switch as follows. In correlation type-1, for a given burst (i.e. the arrival process is in the active period), if the first cell in the current burst was destined to output port i, then the present cell is destined to output port i with probability 0.6 or destined to either one of the other two ports with probability 0.2. In correlation type-2, if the previous cell, within the current burst, was destined to output port i, then the present cell is destined to output port i with probability 0.6, or destined to either one of the other two ports with probability 0.2. Note that the uncorrelated destinations case is represented by the matrix C where c_{ij} has the same value for all i and j. We note that correlation type-1 is more appropriate to describe the traffic at the edges of the network, where correlation type-2 is more appropriate as we move closer to the center of the network.

3.0 The Bus service policies

Classical polling policies (e.g. Round-Robin) consider only the state of the input queues. However, these policies may waste bus bandwidth since some of the transmitted cells may be destined to output queues that are full. Bus allocation policies can be classified according to the information used in deciding which queue to be serve next as follows:

1. Independent of the state of the system

2. Input queues state dependent

3. Output queues state dependent

4. Both, input and output queues state dependent

The bus service policies considered in this paper differ in the information used to decide the next input queue to be served. In policy 1, TDM, no information is needed. Policies 2 and 3 are based only on the state of the input queues. Policies 4 and 5 are based on the state of the input queues plus limited knowledge about the state of the destination queues (full or non-full). Policies 6 and 7 require full knowledge of all input and output queue sizes. The policies are described below.

1.Time Division Multiplexing (TDM): TDM is the simplest bus allocation scheme. In TDM each bus slot i is preassigned to input queue i. Each input queue is only allowed to transmit during the bus slot assigned to it. If input queue i is empty, bus slot i is wasted since other busy input queues are not allowed to use it. The slot assignments follow a predetermined pattern that repeats itself periodically as shown in figure 2. Each such period is called a bus cycle. Once, a cell reaches the head of an input queue, it waits for its slot number and then it is forwarded to the output buffer, if the output buffer is not full. If the output buffer is full, then the cell will be blocked and will wait for its slot number in the next bus cycle. We note that this blocking mechanism is different from the first-blocked-first-unblocked mechanism that has been typically used in continuous-time queueing networks with blocking.

2.Round-Robin (RR): In this policy, the busy input queues are served in a cyclic fashion. The service policy in limited, i.e. only one cell is transmitted from the input queue being served before the server (the bus) switches to the next input queue. Once the present input queue has been served, the server proceeds in cyclic order until it finds an input queue which is non-empty. We assume zero switch-over time. If the destination output buffer is full, then the cell will be blocked. The blocked cell remains at the top of input queue i until the queue is granted the bus again.

3.Longest Input Queue (LIQ): In this policy the selection is based on the input queue lengths. The input queue with the longest queue length is selected first. If there is more than one input queue with the longest length, then one is selected randomly, i.e. the selection probability is uniformly distributed. If the destination output buffer is full, then the cell will be blocked and the bus slot is wasted.

4.Modified Round-Robin (MRR): This policy is an improvement over policy 2 (RR). Only the busy input queues with the top cell destined to nonfull output queues are considered. The selected input queues are served in a cyclic fashion. Therefore, bus slots are not wasted on cells destined to full output queues. Note that this selection process is repeated for every bus slot.

5.Modified Longest Input Queue (MLIQ): Again, this policy is an improvement over policy 3 (LIQ). We form a subset which consists of those input queues with the top cell destined to a nonfull output queue. The longest input queue in this subset is selected next. If there is more than one queue with the largest number of customers, then one is selected randomly.

6.Longest Input Queue - Shortest Output Queue (LIQ-SOQ): This policy is almost the same as policy 5 with the following exception. If there is more than one longest input queue, then we select the one whose top cell is destined to the shortest output queue. If there exists more than one pair of longest input queue/shortest output queue, then one pair is selected randomly. Note that policy 5 does not depend on the output queue length.

6.Shortest Input Queue - Shortest Output Queue (SIQ-SOQ): In this policy, we select the shortest input queue with the top cell destined to a nonfull output queue. If there is more than one shortest input queue, then the shortest input queue with the top cell destined to the shortest output queue is selected. Again, if there are more than one pair of shortest input/shortest output queues, one pair is selected randomly.

4.0 Results

The shared-medium switch under the bus service policies described in the previous sections was analyzed extensively through simulation. A representative sample of the results are given here for an 8x8 switch with queue size (input and output) equal to 16. The blocking probability at the output links B_p was set to zero, i.e. the output link service time is equal to one arrival slot. The 95% confidence intervals were obtained for all runs. Confidence intervals were small and therefore, are not shown. The performance measures of interest were: switch cell loss probability, mean cell delay, and cell delay percentiles.

Symmetric arrivals and uncorrelated uniform destinations: Figures 3 to 8 compare the performance of the different bus service policies under symmetric arrivals, uniform destination distribution (i.e. the routing probability d_{ij} was equal to 1/8 for all cells traversing the switch), and uncorrelated destinations (i.e destinations of successive cells within the same burst were uncorelatted).

Figures 5 to 8 give switch mean cell delay, cell loss probability, and 99th delay percentiles respectively, as a function of cell arrival rate assuming C^2 equals to 20. Figures 6 to 8 give the same performance measures under more bursty traffic where C^2 was set to 100. As shown in the graphs, all policies perform equally well up to an arrival rate of .6 (60% of switch capacity). As we increase the load beyond .6, we start to observe a difference , in terms of mean cell delay and cell loss probability, between the policies. Policies 1 (TDM) and 3 (LIQ) perform rather poorly in terms of both cell delay and cell loss probability measures. This is due to the fact that both policies waste more bus slots than the other policies. TDM policy wastes bus slots on empty input queues when they could have been used by other busy input queues. Policy 3 (LIQ) gives priority to the longest input queue independent of the state of the destination queue (full or nonfull). As a result, bus slots are wasted on cells that can not be forwarded. Note that under the LIQ policy, bus slots are wasted in groups of 8 since there are 8 bus slots within the boundaries of an arrival or a departure slot, i.e. the longest input queue remains the longest queue for 8 successive bus slots. It is interesting to note that policy 2 (RR) performs better than policy 3 (LIQ), in terms of both delay and cell loss probability, although policy 2 requires less information about the state of the input queues. This is due to the fact that RR policy changes the input queue being served every bus slot where LIQ policy continues on serving an input queue that is blocked. Therefore, more bus slots are wasted under LIQ than RR.

Next, we compare the policies that are based on the state of both input and output queues (policies 4, 5, 6, and 7). In terms of delay, policy 7 (SIQ-SOQ) gives the shortest delay; however, this policy has the highest cell loss probability because it gives the short input queue a higher priority at the expense of the long input queues thus resulting in high cell loss. Because of the high cell loss, policy 7 is not an appropriate policy to be considered and therefore, it will not be discussed further. Policies 4 (Modified RR), 5 (Modified LIQ), and 6 (LIQ-SOQ) waste the least bus slots among all policies considered in this study. Policy 4 gives lower switch delay where policy 6 gives the least cell loss. The higher cell delay under policy 6 is due to the fact that more cells are accepted which results in higher cell delay.

Figures 3 and 6 give the 99th cell delay percentiles for the switch as a function of the arrival rate ρ for C^2 = 20 and 100 respectively. Note that the percentiles are calculated for all cells traversing the switch. For example, under TDM, 99% of the cells accepted will traverse the switch in 300 bus slots or less (assuming ρ = 0.8 and C^2 = 20). Maximum cell delay through the switch can be approximated by the 99th cell delay percentiles. We note that this maximum delay is approximately the same for all policies under consideration; however, as shown in figure 9, there is a significant difference between TDM and LIQ on one hand, and MRR on the other hand in terms of cell delay percentiles. As we approach the maximum switch cell delay (99th percentile), the difference between the policies gets smaller. Figure 10 compares the cell delay percentiles (10% to 99%) for policies 4, 5, 6, and 7 assuming ρ = 0.8 and C^2 = 100. We note that the cell delay percentiles under the four policies are about the same. Note that the percentiles may differ if we consider only the queueing time at the input queues; however, in this paper we are concerned with the total switch delay which includes queueing times at both the input and output queues.

From the above data, we can see that policy 4 gives the lowest delay where policy 6 gives the lowest cell loss. Therefore, the following comparisons will be made only between these two policies.

Correlated destinations: Figure 11 shows the effect of destination correlations on switch delay under LIQ-SOQ policy. The correlation matrix C was set as shown below.

$$C = \begin{bmatrix} 0.3 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.3 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.3 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.3 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.3 \end{bmatrix}$$

As shown in the graph, both correlation types cause increase in switch cell delay. However, switch delay under correlation type-1 increases more rapidly than under correlation type-2 as we increase the arrival rate. For example, when $\rho = 0.7$ the mean switch delay is about 30 bus slots under type-2 correlation versus 75 bus slots under type-1 correlation. This is due to the fact that correlation type-1 is dependent on the burst size since it is based on the destination of the first cell in the burst. As we increase ρ (while C^2 is fixed to 100), the burst size is increased. For example, assuming $C^2 = 100$, the burst size is equal to 103 and 557 for ρ equal to .3 and .7 respectively. Consequently, both types of correlation are equal for small ρ . For large ρ , type-1 correlation gives much stronger correlation between successive cells in the same burst. This stronger correlation is reflected in the increase in switch cell delay.

Figures 12 to 15 compare policies 4 and 6 in terms of cell delay and cell loss under the two types of correlations. The arrival rate ρ was set to 0.7 assuming $C^2 = 20$. Under correlation type-1, policy 4 (Modified RR) gives lower cell delay than policy 6 (LIQ-SOQ) and the cell loss is approximately equal under both policies. However, under correlation type-2, the switch delay is equal; but, the cell loss probability is much less under policy 6 (LIQ-SOQ). This is due to the fact policy 6 tends to serve the same input queue for several successive bus slots which results in one output queue being full while the other output queues are not quit as full or perhaps empty. On the other hand, under policy 4, since the bus is granted to the busy input queues in cyclic fashion, the destination of cells transmitted across the bus will be more evenly distributed among the output queues. This leads to lower probability of blocking (i.e. Round-Robin policy minimizes head of line blocking).

<u>Nonuniform destinations:</u> Results in figure 16 were obtained under a nonuniform destination distribution (hot spot pattern). In particular, for each input queue 30% of the traffic was destined to output queue 1 and the remainder of the traffic was uniformly distributed among the rest of the output queues, i.e. output queues 2 to 8 get 10% each. Figure 16 gives the mean cell delay of output queue 1 (hot output), output queue 8 (a non-hot output), and the mean overall switch cell delay under policies 4 and 6. The arrival process to each input queue was an IBP with $C^2 = 100$ and $\rho = 0.4$. Note that when ρ is equal to 0.4, the average rate of arrivals destined for output 1 (hot output) is equal to 0.96 which is very close to the output link capacity. We note that the cell delay under policy 6 is about 37% more than the delay under policy 4. Also, the cell loss probability under policy 4 is 37% more than that under policy 6 (0.11 versus 0.08). The

99th cell delay percentiles under the same arrival pattern are shown in figure 17. Note that the 99th percentiles for policy 6 is much greater than that of policy 4 (more than double) where the difference in the mean switch delay was only about 25%.

Asymmetric arrivals: Figures 18 and 19 give the mean cell delay and cell loss probability under asymmetric arrival process ($\rho_1, \rho_2, \rho_3, \rho_4 = 0.7$, and $\rho_5, \rho_6, \rho_7, \rho_8 = 0.9$) for three values of C^2 : 20, 100, and 200. Uniform destinations are assumed. We note that the mean cell delay is approximately the same under both policies. However, the cell loss under policy 6 (LIQ-SOQ) is less than half of the cell loss under policy 4(RR). This is due to the fact that policy 6 (LIQ-SOQ) tends to equalize the queue sizes and as a result, the longer input queue (the input queue with high arrival rate) will be served more frequently than the input queues with lower arrival rates. On the other hand, policy 4 (RR) give equal priority to all busy input queues independent of the queue size. Thus, an input queue that is losing a lot of cells is served at the same rate as other input queues that are nearly empty which results in the high cell loss probability.

Asymmetric nonuniform arrivals with correlated destinations: Results in figure 20 were obtained under a nonuniform destination distribution (for each input queue 30% of the traffic was destined to output queue 1 and the remainder of the traffic was uniformly distributed among the rest of the output queues) and Asymmetric arrivals ($\rho_1, \rho_2, \rho_3, \rho_4 = 0.7$, and $\rho_5, \rho_6, \rho_7, \rho_8 = 0.9$) assuming $C^2 = 100$. Furthermore, the destination of successive cells were correlated according the matrix C used in the correlated destinations results previously. As shown in the graph, the two policies perform equally well in terms of cell delay and cell loss. Recall that policy 4 performed better under highly correlated destinations whereas policy 7 performed better under asymmetric and nonuniform arrivals. As a result, the mixing of asymmetric arrivals and correlated destinations offset each other.

Summary: In general, the results can be summarized as follows:

1. Considering policies that depends only on the state of the input queues, RR is superior to LIQ and TDM in terms of both cell delay and cell loss.

2. Considering policies that depend on the state of both input and output queues, policy 6 (LIQ-SOQ) is better than policies 4 (MRR) and 5 (MLIQ) under traffic with uncorrelated destinations (lowest cell loss

coupled with slight increase in cell delay). Under asymmetric arrivals , policy 6 is much superior to policy 4 (equal cell delay and much less cell loss).

3. Policy 4 (MRR) gives better performance measures under highly correlated cell destinations.

5.0 Conclusions

In this paper we analyzed the performance of different bus allocation policies for a generic shared-medium ATM switch with input and output queueing. The policies differed in the amount of information used by the bus arbitrator to decide which input queue to serve next. We considered a range of bus service policies that are based on the state of the input queues, or the state of both the input and the output queues. The performance of the switch is analyzed under realistic system characteristics such as bursty traffic, limited buffer size, asymmetric load conditions, and nonuniform destinations. Furthermore, we assume that the destinations of the arriving cells within the same burst are correlated. Two types of destination correlations were introduced and their effect on the performance of the switch was analyzed. Results show that policy 6 (Longest Input Queue-Shortest Output Queue) is superior for all cases except when there is a strong destination correlation between successive cells. For those environments where there is a strong destination correlation, policy 4 (Modified Round-Robin) performs better.

References

1. H. Badran and H. Mouftah, "Head of line arbitration in ATM switches with input-output buffering and backpressure control", IEEE GLOBCOM 91, Phoenix, Arizona, 1991.

2. J. Boudec, E. Port, and H. Truong, "Flight of the FALCON", IEEE Communications magazine, Vol. 31, NO. 2, February 1993.

3. I. Cidon, I. Gopal, G. Grover, and M. Sidi, "Real-Time Packet Switching: A Performance Analysis", IEEE JSAC Vol. 6 NO. 9, December 1988.

4. I. Iliadis, "Head of line arbitration of packet switches with input and output queueing", IFIP TC6 4th. Int. Conference on Data Communication Systems and Their Performance, Barcelona, Spain, June 1990.

5. J. Karlsson, H. Perros, I. Viniotis, "Adaptive polling schemes for an ATM bus with bursty arrivals", Computer Networks and ISDN Systems, 24 (1992), North-Holland.

6. M. Karol, M. Hluchyj, and S. Morgan, "Input versus output queueing on a space-division packet switch", IEEE Trans. Comm., VOL. 35, NO. 12, December 1987.

7. A. Pattavina, "Performance evaluation of ATM switches with input and output queueing", International Journal of Digital and Analog Communication Systems, VOL. 3, 1990.

8. H.G. Perros, "Approximation algorithms for open queueing networks with blocking", in Takagi (ed.), Stochastic Analysis of Computer and Communication Systems, North-Holland, 1990, 451-498.

9. H. Suzuki, H. Nagano, T. Takeuchi, and S. Iwasaki, "Output-buffer switch architecture for asynchronous transfer mode", Int. Conf. on Communications, Boston, MA, June 1989.

10. F. Tobagi, "Fast Packet Switch Architectures for Broadband Integrated Services Digital Networks", Proceedings of the IEEE, Vol. 78, NO. 1, January 1990.

11. J. Turner, "Design of a broadcast packet switching network", IEEE Trans. Comm., Vol. 36, NO. 6, June 1988.

12. A. Zaghloul and H. Perros, "Approximate analysis of a generic shared-medium ATM switch under bursty arrivals and nonuniform destinations", submitted

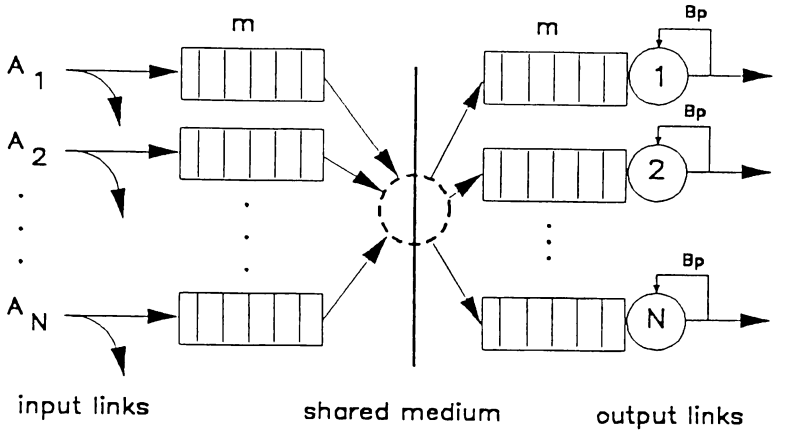

Figure 1. A Queueing model of a shared-medium switch

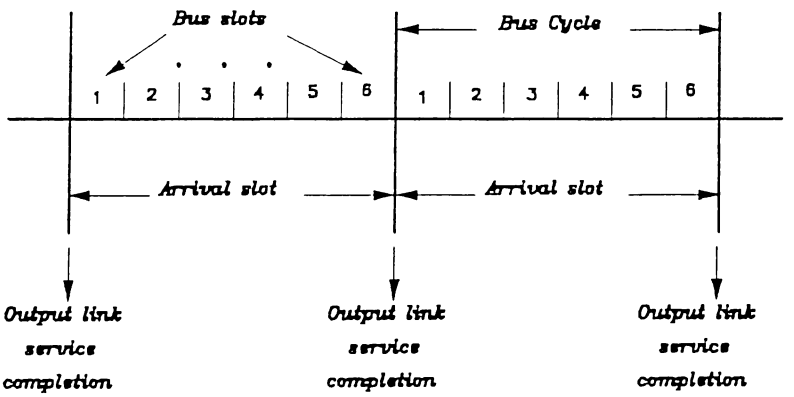

Figure 2. Bus slots and arrival slots

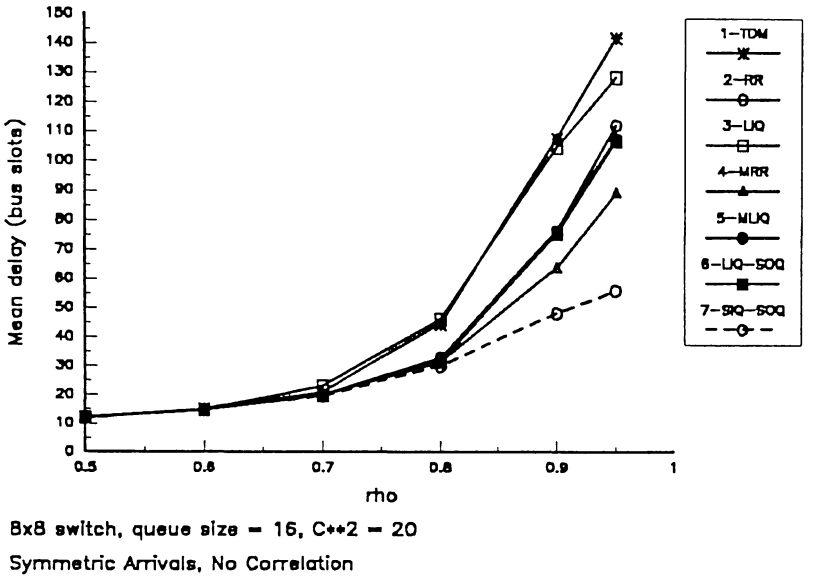

Figure 3. Switch cell delay comparison ($C^2 = 20$)

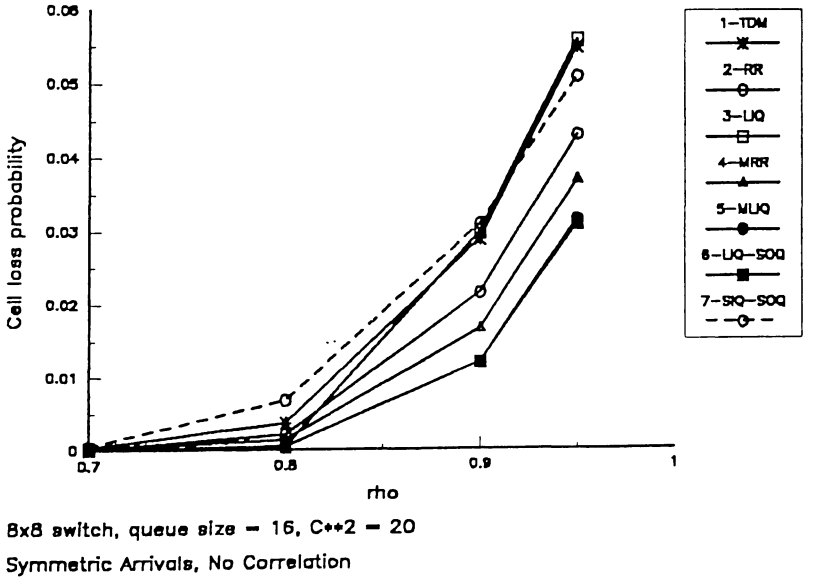

Figure 4. Switch cell loss probability comparison ($C^2 = 20$)

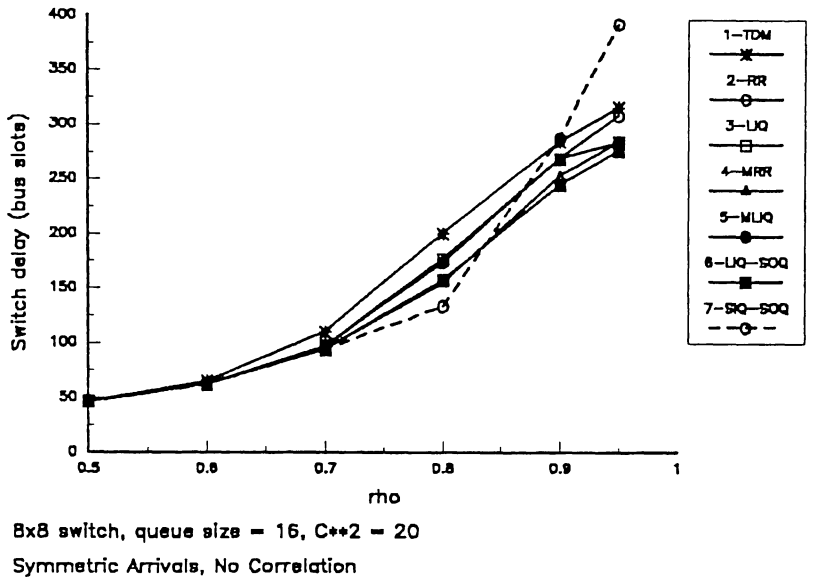

Figure 5. 99th switch cell delay percentile ($C^2 = 20$)

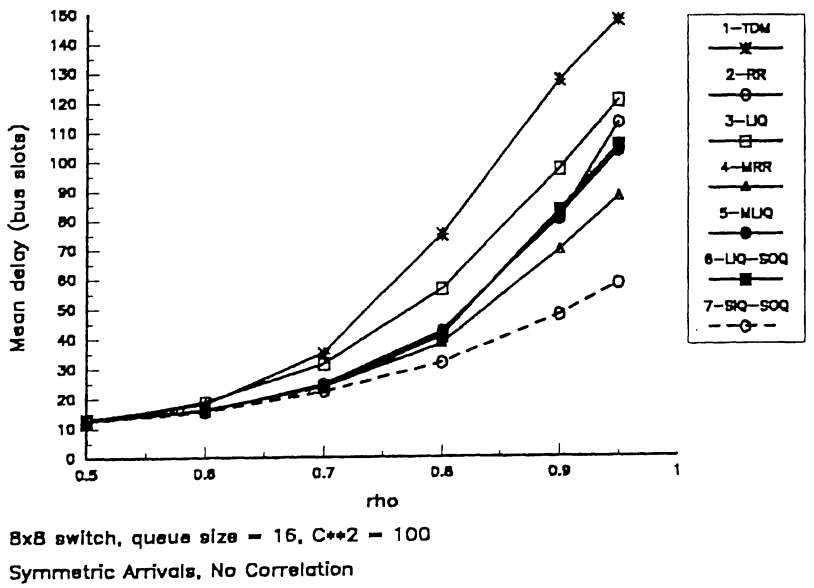

Figure 6. Switch cell delay comparison ($C^2 = 100$)

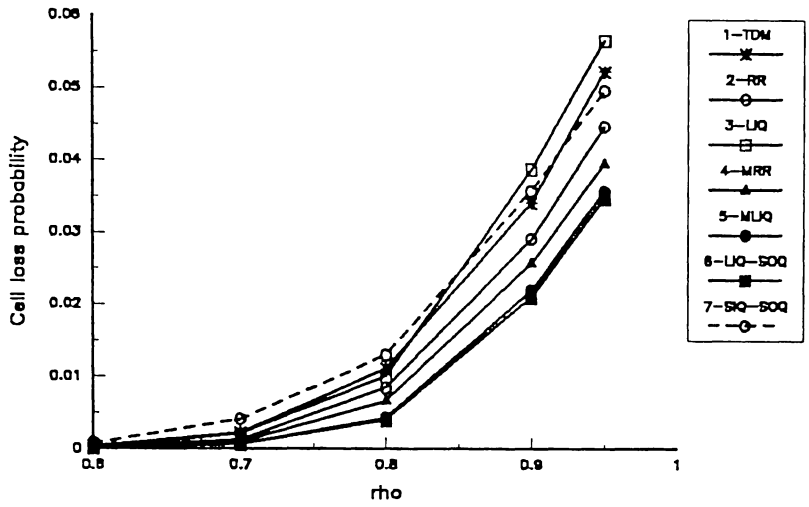

Figure 7. Switch cell loss probability comparison ($C^2 = 100$)

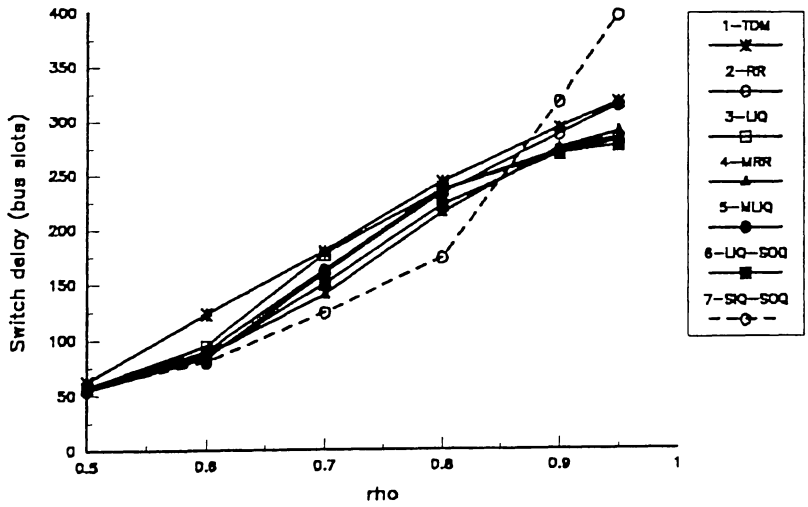

Figure 8. 99th switch cell delay percentile ($C^2 = 100$)

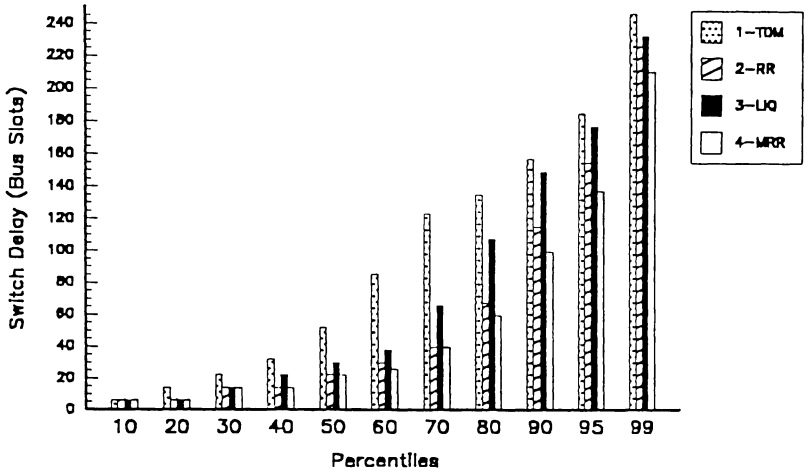

Figure 9. Cell delay percentiles (policies 1, 2, 3, 4)

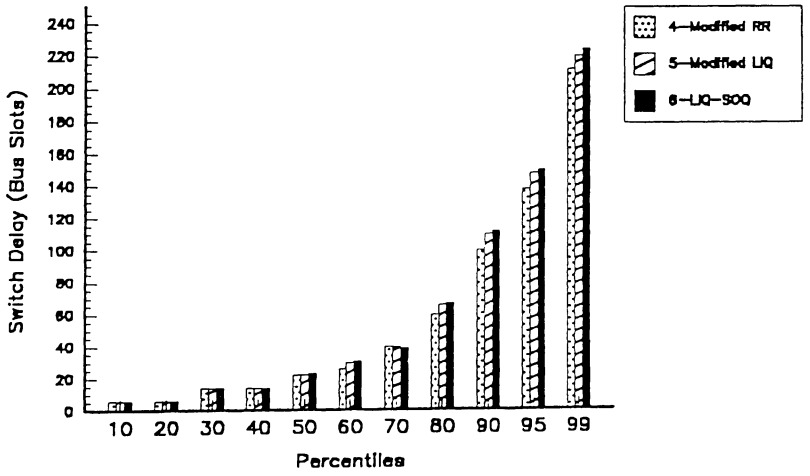

Figure 10. Cell delay percentiles (policies 4, 5, 6, 7)

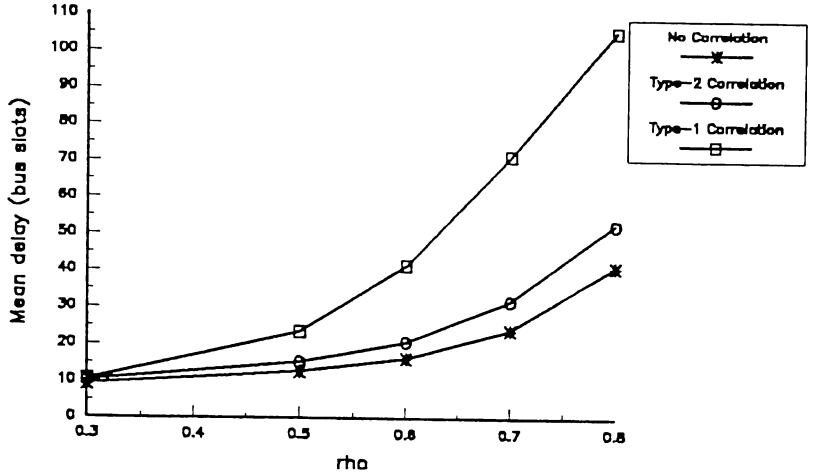

Figure 11. Effect of correlation on switch cell delay

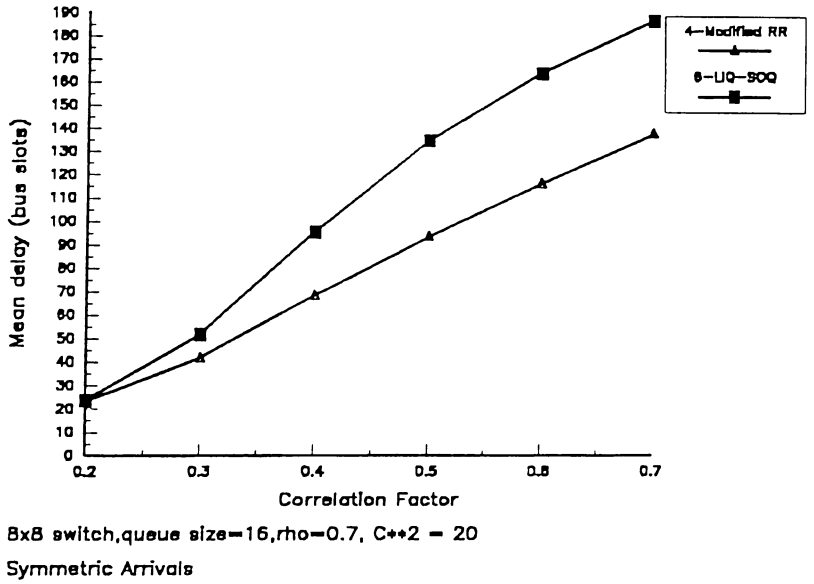

Figure 12. Switch cell delay under type-1 correlation

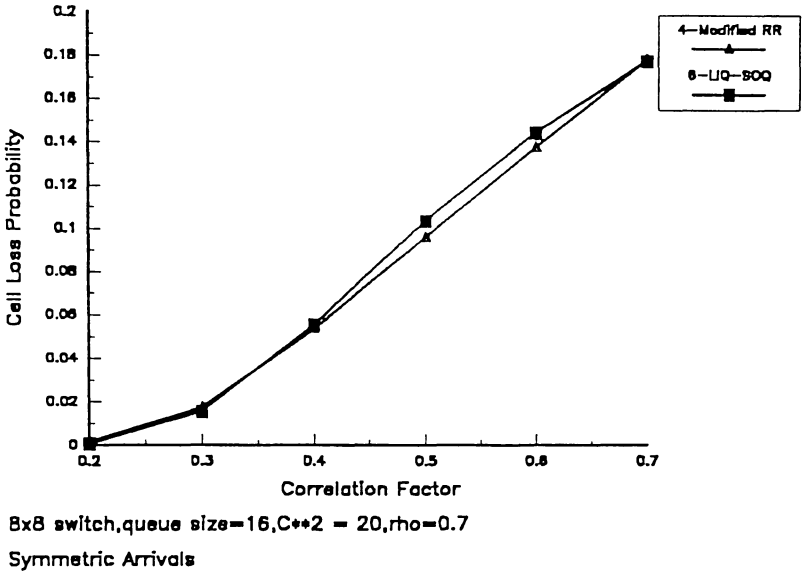

Figure 13. Switch cell loss probability under type-1 correlation

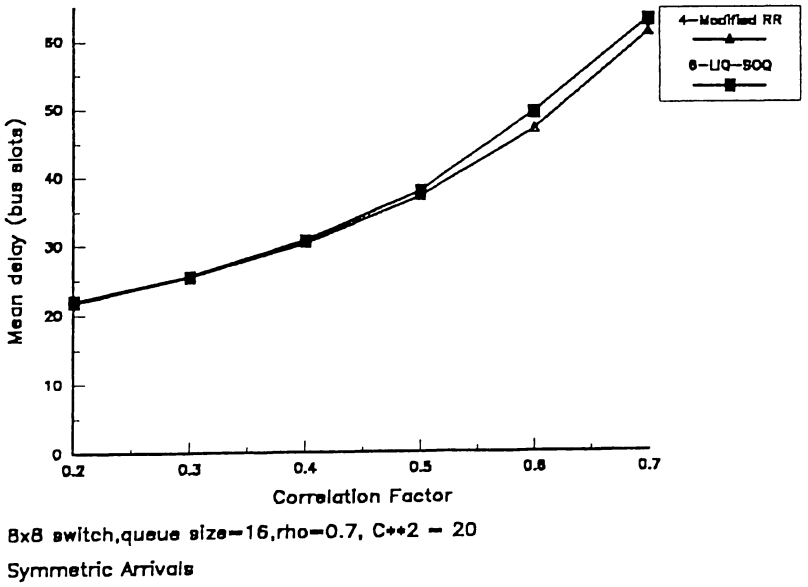

Figure 14. Switch cell delay under type-2 correlation

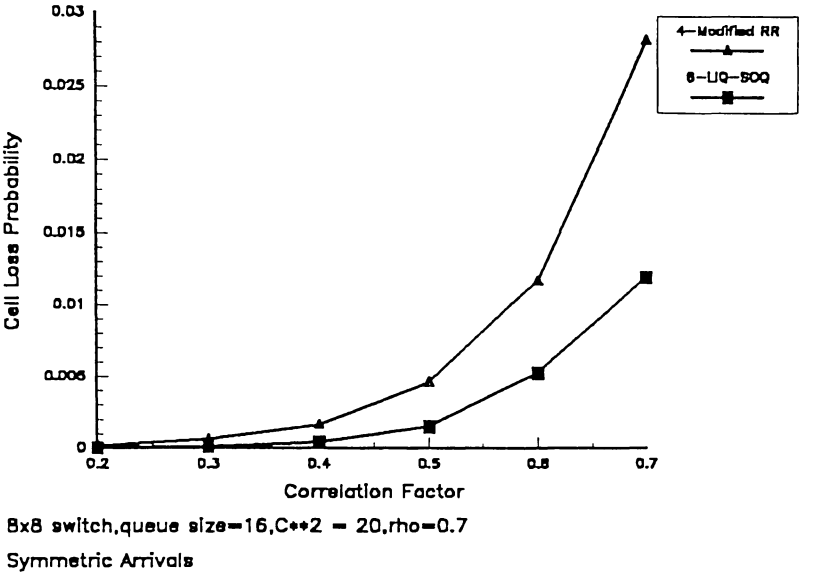

Figure 15. Switch cell loss probability under type-2 correlation

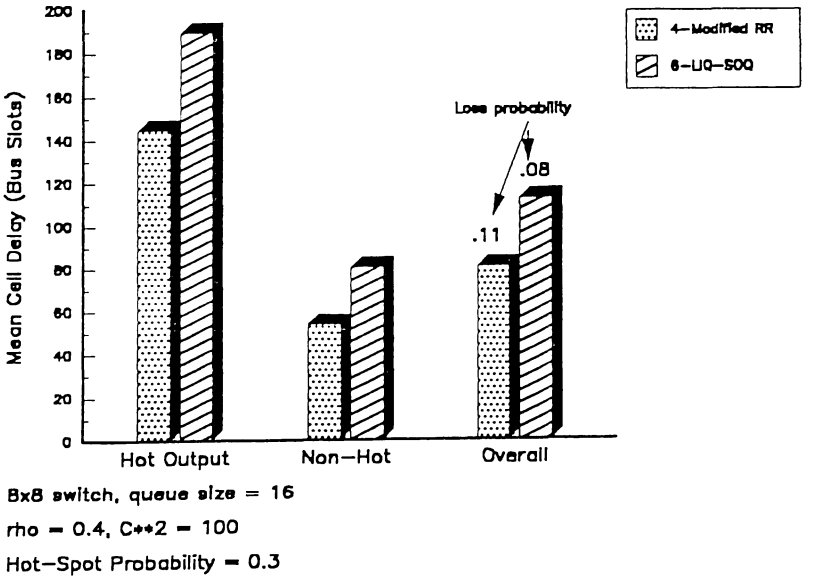

Figure 16. Switch cell delay under nonuniform destinations

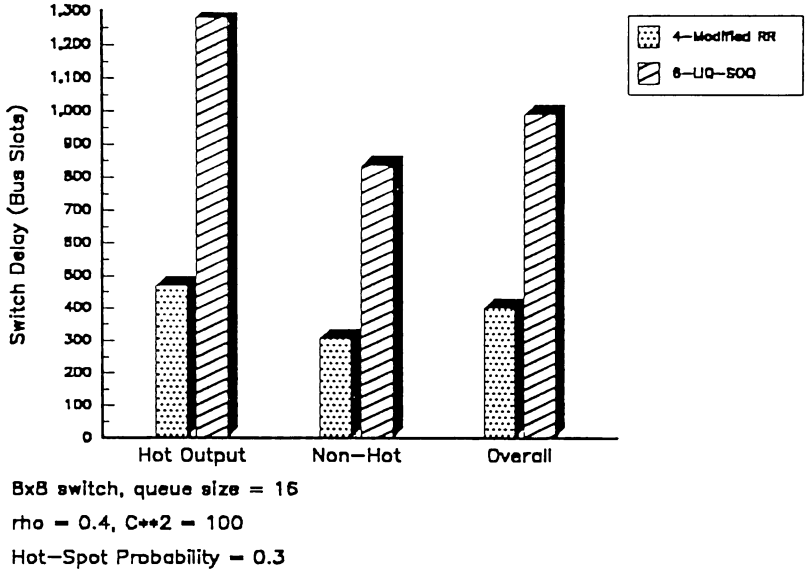

Figure 17. 99th cell delay percentile under nonuniform destinations

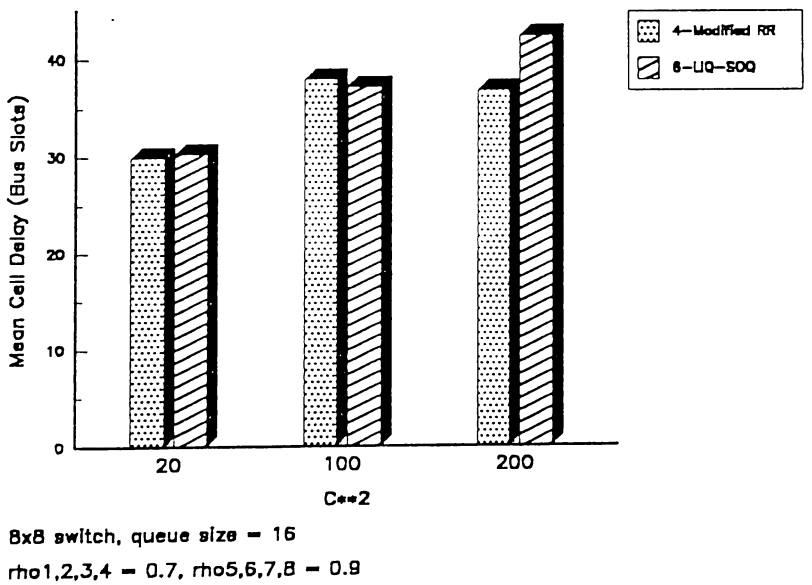

Figure 18. Switch cell delay under Asymmetric arrivals

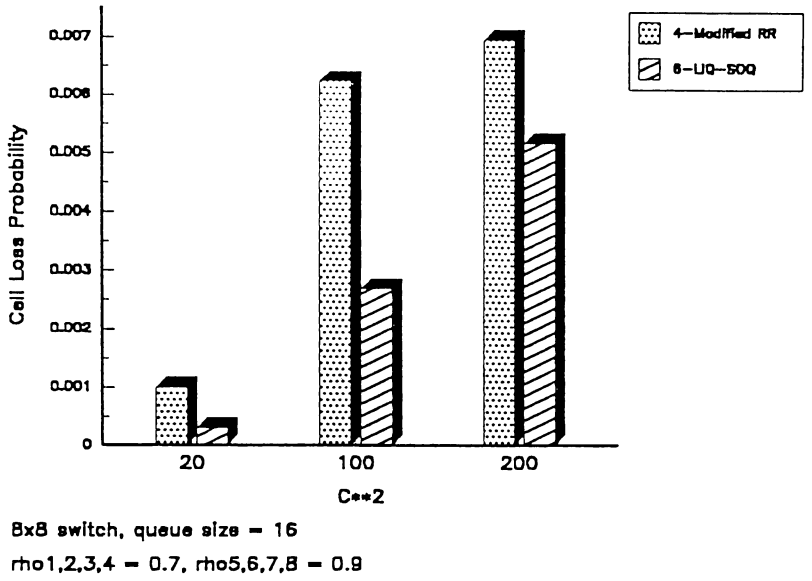

Figure 19. Switch cell loss probability under Asymmetric arrivals

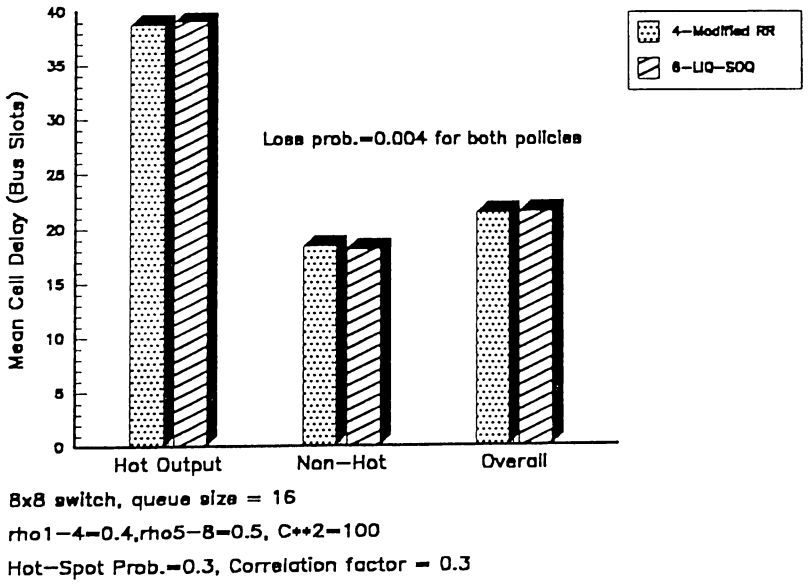

Figure 20. Cell delay and loss (asymm. arrivals, nonuniform dest., & corr. destinations)