

# An Improved Model for the Performance Analysis of Multistage Switches

C.-K. Chen

Department of Computer Science  
& Computer Engineering  
La Trobe University  
Melbourne 3083, Australia.  
chenc@LATCS1.lat.oz.au

M. Atiquzzaman

Department of Electrical & Computer  
Systems Engineering,  
Monash University,  
Clayton, Melbourne 3168, Australia.  
atiq@LATCS1.lat.oz.au

## Abstract

*Previous models for performance evaluation of multistage switches have been neither accurate enough nor based on realistic assumptions regarding the correlation of the blocked cells in successive stages of the switch. In this paper, a new analytical model which permits an accurate analysis of multistage switches is proposed. The new model reflects the behavior of blocked cells, and takes into account the fact that a blocked cell always hunts for the same output link in successive cycles.*

## 1 Introduction

Multistage switches (MS) have been used to connect processors to memories in massively parallel multiprocessor systems and also in ATM switching fabrics in Broadband ISDN networks. An MS consists of several stages of small crossbar switching elements (SE) connected by a permutation function. Routing conflicts inside the MS prevent simultaneous arbitrary input output connection requests. Buffers are used to store the cells which lose the routing conflicts.

Previously, the approximate performance modeling of MSs under uniform reference [1, 2] and non-uniform and hot-spot traffic patterns [3, 4, 5] have been presented. Mun [1] proposed an improved model over Hsiao's model [2] by introducing a correlation between the blocked state and the empty state. In the analysis, Mun implicitly assumes that a blocked cell has equal probabilities of requesting the output links of the SE during subsequent cycles. The impact of this implicit assumption is reflected in the increased errors with an increase in the blocking in the switch and the number of stages.

In this paper we propose an accurate model which takes into account the correlation between the states of the buffer of adjacent stages and the history of cells blocked during routing conflicts. The novelty of the model lies in removing the assumptions of previous models regarding blocked cells. The higher accuracy of the proposed model is due to its ability to memorize

the history of a blocked cell. It is shown that the accuracy of the proposed model is higher than the most recently [1] published model.

Section 2 presents the MS and the assumptions used in the model. The proposed model is developed in Section 3. Results obtained from the proposed model are compared with results obtained from simulation and Mun's model [1] in Section 4, followed by some concluding remarks and summary of our findings in Section 5.

## 2 Network and Modeling Assumptions

An MS consists of  $n = \log_a N$  stages of  $N/a$  SEs per stage, each SE having  $a$  input and  $a$  output lines. A perfect shuffle interconnection is used to connect the different stages as shown in Figure 1 for a switch of size  $N = 8$  using  $2 \times 2$  SEs.

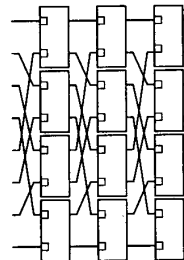


Figure 1: A Multistage ATM switch.

A switch cycle is assumed to be composed of two phases [4]. In phase 1, cells are forwarded and cells are received during phase 2. The following assumptions are used in the proposed model.

1. Each input receives a cell during a cycle with probability  $r$ . The cells are synchronized at the beginning of a cycle and are uniformly distributed over all the outputs. The cell received at an input is independent of the cells received by other inputs.

2. Each input of an SE has a single buffer.
3. The queuing and dequeuing of cells in a buffer take place during the same cycle.
4. Every cell has equal probability of winning the contention and the blocked cells are resubmitted to the original destination.
5. The arrival of a cell during a cycle is independent of whether the cell received at the previous cycle was accepted or rejected.

We represent the states of a buffer in an SE by:

**State 0** : Buffer is empty.

**State n** : Buffer has a new cell which has just arrived in the previous cycle.

**State u** : Buffer has a cell which is blocked for the upper output of the SE.

**State l** : Buffer has a cell which is blocked for the lower output of the SE.

Figure 2 shows the state transition diagram of the proposed model.

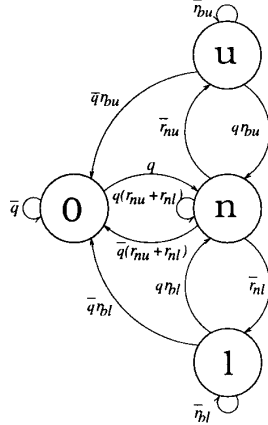


Figure 2: Markov chain for the proposed model.

### 3 Proposed Model

Without loss of generality, we assume a  $2 \times 2$  SE to illustrate our model. The two outputs of an SE will be called the *upper* and *lower* outputs. The following notations will be used in developing the model. An SE at stage  $k$  will be denoted by  $SE(k)$ .

$q(k, t)$ : Probability that a cell is ready to come to  $SE(k)$  during cycle  $t$ . The offered load to the input of the switch is therefore  $q(1, t)$ .

$\rho(k, t)$ : Probability that a buffer at  $SE(k)$  receives a cell during cycle  $t$ .

$\pi_0(k, t)$ : Probability that a buffer of  $SE(k)$  is empty at the beginning of cycle  $t$ .

$\pi_n(k, t), \pi_u(k, t), \pi_l(k, t)$ : Probability that a buffer of  $SE(k)$  has a new cell, a cell blocked for the upper output, or a cell blocked for the lower output, respectively, at the beginning of cycle  $t$ .

$r_{nu}(k, t)$ : Probability that a new cell in  $SE(k)$  moves to the upper output during cycle  $t$ .

$r_{nu}^{0n}(k, t)$ : Probability that a new cell in  $SE(k)$  is able to get to the upper output during cycle  $t$  and the state of the other buffer in the same SE is either 0,  $n$  or  $l$ .

$r_{nu}^u(k, t)$ : Probability that a new cell in  $SE(k)$  is able to get to the upper output during cycle  $t$  and the state of the other buffer in the same SE is  $u$ .

$r_{bu}(k, t)$ : Probability that a cell which is blocked for the upper output in  $SE(k)$  moves during cycle  $t$ .

$r_{bu}^{0n}(k, t)$ : Probability that a cell which is blocked for the upper output in  $SE(k)$  is able to get to the upper output during cycle  $t$  and the state of the other buffer in the same SE is 0 or  $n$ .

$r_{bu}^{ul}(k, t)$ : Probability that a cell which is blocked for the upper output in  $SE(k)$  is able to get to the upper output during cycle  $t$  and the state of the other buffer in the same SE is  $u$  or  $l$ .

$r_{nl}(k, t)$ : Probability that a new cell in  $SE(k)$  moves to the lower output during cycle  $t$ .

$r_{nl}^{0n}(k, t)$ : Probability that a new cell in  $SE(k)$  is able to get to the lower output during cycle  $t$  and the state of the other buffer in the same SE is either 0,  $n$  or  $u$ .

$r_{nl}^l(k, t)$ : Probability that a new cell in  $SE(k)$  is able to get to the lower output during cycle  $t$  and the state of the other buffer in the same SE is  $l$ .

$r_{bl}(k, t)$ : Probability that a cell which is blocked for the lower output in  $SE(k)$  moves during cycle  $t$ .

$r_{bl}^{0n}(k, t)$ : Probability that a cell which is blocked for the lower output in  $SE(k)$  is able to get to the lower output during cycle  $t$  and the state of the other buffer in the same SE is 0 or  $n$ .

$r_{bl}^{ul}(k, t)$ : Probability that a cell which is blocked for the lower output in  $SE(k)$  is able to get to the lower output during cycle  $t$  and the state of the other buffer in the same SE is  $u$  or  $l$ .

$\alpha(k, t)$ : Probability that at least one buffer space in  $SE(k)$  is available during cycle  $t$ , given that there is no blocked cell in  $SE(k-1)$  which is destined to this buffer.

$\alpha^b(k, t)$ : Probability that at least one buffer space in SE( $k$ ) is available during cycle  $t$ , given that only one blocked cell in SE( $k-1$ ) is destined to this buffer.

$\alpha^{bb}(k, t)$ : Probability that at least one buffer space in SE( $k$ ) is available during cycle  $t$ , given that SE( $k-1$ ) contains two blocked cells which are destined to this buffer.

Next we derive the different state and routing probabilities based on the state diagram of Figure 2.

### 3.1 State Probabilities

The state probabilities are given by:

$$\pi_0(k, t) = 1 - \pi_n(k, t) - \pi_u(k, t) - \pi_l(k, t) \quad (1)$$

$$\pi_n(k, t) = q(k, t-1)[\pi_0(k, t-1) + \pi_n(k, t-1) + [r_{nu}(k, t-1) + r_{nl}(k, t-1)] + \pi_u(k, t-1) + r_{bu}(k, t-1) + \pi_l(k, t-1)r_{bl}(k, t-1)] \quad (2)$$

$$\pi_u(k, t) = \pi_n(k, t-1)\overline{r_{nu}(k, t-1)} + \pi_u(k, t-1)\overline{r_{bu}(k, t-1)} \quad (3)$$

$$\pi_l(k, t) = \pi_n(k, t-1)\overline{r_{nl}(k, t-1)} + \pi_l(k, t-1)\overline{r_{bl}(k, t-1)} \quad (4)$$

where  $\overline{x(k, t)} = 1 - x(k, t)$  for any variable  $x(k, t)$ .

### 3.2 Routing and Acceptance Probabilities

The relationship between the routing and blocking probabilities for new and blocked cells are expressed as follows.

$$r_{nu}(k, t) + r_{nl}(k, t) + \overline{r_{nu}(k, t)} + \overline{r_{nl}(k, t)} = 1 \quad (5)$$

$$r_{bu}(k, t) + \overline{r_{bu}(k, t)} = 1 \quad (6)$$

$$r_{bl}(k, t) + \overline{r_{bl}(k, t)} = 1 \quad (7)$$

where  $r$  and  $\overline{r}$  are the routing and blocking probabilities respectively. For uniform incoming traffic which is uniformly distributed over the outputs,  $r_{nu}(k, t) = r_{nl}(k, t)$  and  $\overline{r_{nu}(k, t)} = \overline{r_{nl}(k, t)}$ . Therefore,

$$\overline{r_{nu}(k, t)} = 0.5 - r_{nu}(k, t) \quad (8)$$

$$\overline{r_{nl}(k, t)} = 0.5 - r_{nl}(k, t) \quad (9)$$

When considering  $r_{nu}^{0nl}(k, t)$ , the new cell under consideration is able to get to its desired output when there is no conflict with any cell in the other buffer of the same SE, or can win in the case of a conflict. Therefore,

$$\begin{aligned} r_{nu}^{0nl}(k, t) &= 0.5\pi_0(k, t) + 0.5 \times 0.5 \times 0.5\pi_n(k, t) + \\ & 0.5 \times 0.5\pi_n(k, t) + 0.5\pi_l(k, t) \\ &= 0.5\pi_0(k, t) + 0.375\pi_n(k, t) + 0.5\pi_l(k, t) \end{aligned} \quad (10)$$

$$r_{nu}^u(k, t) = 0.5 \times 0.5\pi_u(k, t) = 0.25\pi_u(k, t) \quad (11)$$

Similarly  $r_{nl}^{0nu}(k, t)$  can be obtained.

When the two buffers of SE( $k-1$ ) contain a new and a blocked cell during cycle  $t$  and both of them are destined to the same output link of the SE, the probability that the buffer at SE( $k$ ) is able to accept one of these cells is

$$\alpha^b(k, t) = 0.5\rho(k, t-1)[r_{nu}(k, t) + r_{nl}(k, t)] + \left[ \frac{1 - 0.5\rho(k, t-1)}{2} \right] [r_{bu}(k, t) + r_{bl}(k, t)] \quad (12)$$

where  $0.5\rho(k, t-1)$  is the probability that the buffer at SE( $k$ ) has received a cell during cycle  $t-1$  from the buffer in SE( $k-1$ ). Since the buffer at SE( $k$ ) can not be in an empty state, for uniform traffic at the input of the switch, the probability that the buffer is in state  $u$  or  $l$  is equal and is given by  $[1 - 0.5\rho(k, t-1)]/2$ . Now,  $\alpha(k, t)$  is given by

$$\alpha(k, t) = \pi_0(k, t) + \pi_n(k, t)[r_{nu}(k, t) + r_{nl}(k, t)] + \pi_u(k, t)r_{bu}(k, t) + \pi_l(k, t)r_{bl}(k, t) \quad (13)$$

Finally, for a cell to move to the succeeding stage, it should be able to get to the desired output port and the destined buffer should be available. Thus, for  $1 \leq k \leq n-1$ ,

$$r_{nu}(k, t) = r_{nu}^{0nl}(k, t)\alpha(k+1, t) + r_{nu}^u(k, t)\alpha^b(k+1, t) \quad (14)$$

$$r_{nl}(k, t) = r_{nl}^{0nu}(k, t)\alpha(k+1, t) + r_{nl}^l(k, t)\alpha^b(k+1, t) \quad (15)$$

When a cell is in state  $u$ ,  $r_{bu}^{0n}(k, t)$  and  $r_{bu}^{ul}(k, t)$  for  $1 \leq k \leq n-1$  are given by

$$\begin{aligned} r_{bu}^{0n}(k, t) &= \pi_0(k, t) + 0.5\pi_n(k, t) + 0.5 \times 0.5\pi_n(k, t) \\ &= \pi_0(k, t) + 0.75\pi_n(k, t) \end{aligned} \quad (16)$$

$$r_{bu}^{ul}(k, t) = \pi_l(k, t) + 0.5\pi_u(k, t) \quad (17)$$

If both buffers at SE( $k-1$ ) are blocked, the acceptance probability of the corresponding destination buffer at stage  $k$ ,  $2 \leq k \leq n$ , which must be in a blocked state, is given by

$$\alpha^{bb}(k, t) = \frac{1 - 0.5\rho(k, t-1)}{2} (r_{bu}(k, t) + r_{bl}(k, t)) \quad (18)$$

Therefore, for  $1 \leq k \leq n-1$ ,

$$r_{bu}(k, t) = r_{bu}^{0n}(k, t)\alpha^b(k+1, t) + r_{bu}^{ul}(k, t)\alpha^{bb}(k+1, t) \quad (19)$$

Similar logic applies for the routing of a cell blocked for the lower output port.

### 3.3 Throughput and Delay

The traffic rate at any link connecting the SEs between SE( $k-1$ ) and SE( $k$ ) can be obtained by finding the probability that a cell is received at SE( $k$ ) from SE( $k-1$ ). For a buffer at the first stage, i.e., for  $k=1$ ,

$$\rho(k, t) = q(k, t)[\pi_0(k, t) + \pi_n(k, t)[r_{nu}(k, t) + r_{nl}(k, t)] + \pi_u(k, t)r_{bu}(k, t) + \pi_l(k, t)r_{bl}(k, t)] \quad (20)$$

For  $SE(k)$ ,  $2 \leq k \leq n$ ,

$$\rho(k, t) = \pi_n(k-1, t)[r_{nu}(k-1, t) + r_{nl}(k-1, t)] + \pi_u(k-1, t)r_{bu}(k-1, t) + \pi_l(k-1, t)r_{bl}(k-1, t) \quad (21)$$

The probability that a cell is offered to a buffer at  $SE(k)$  is the ratio of the probability that a cell is received by the buffer to the probability that the buffer can accept the cell. For  $SE(k)$ ,  $2 \leq k \leq n$ ,

$$q(k, t) = \frac{\rho(k, t)}{\pi_0(k, t) + \pi_n(k, t)[r_{nu}(k, t) + r_{nl}(k, t)] + \pi_u(k, t)r_{bu}(k, t) + \pi_l(k, t)r_{bl}(k, t)} \quad (22)$$

where  $\rho(k, t)$  is obtained from Equation (21).

For the first stage,  $q(1, t)$  is specified as an input to the model, and since the last stage can always accept cells  $\alpha(n, t) = \alpha^b(n, t) = \alpha^{bb}(n, t) = 1$ . The routing probabilities at the last stage are therefore, obtained by setting the acceptance probabilities of the next stage equal to unity in Equations (14), (15) and (19).

*Normalized throughput* ( $\lambda$ ) of a switch is defined to be the probability of a cell leaving an output port of the switch during a cycle. Thus the normalized throughput at time  $t$  is given by

$$\lambda(n, t) = \pi_n(n, t)[r_{nu}(n, t) + r_{nl}(n, t)] + \pi_u(n, t)r_{bu}(n, t) + \pi_l(n, t)r_{bl}(n, t) \quad (23)$$

The *mean delay* ( $\delta$ ) is defined to be the number of clock cycles taken by a cell to traverse the switch and is given by  $\delta = \sum_{k=1}^n \delta(k)$  where  $\delta(k)$  is the delay for a cell at stage  $k$  in the steady state. Thus, for  $1 \leq k \leq n$

$$\delta(k) = \frac{\pi_n(k) + \pi_u(k) + \pi_l(k)}{\rho(k)} \quad (24)$$

where  $x(k)$  represents the time independent steady state value of the variable  $x(k, t)$ .

## 4 Results

Figures 3 to 6 show the normalized throughput and the mean delay as a function of the traffic rates obtained from the proposed model and Mun's model for  $512 \times 512$  and  $4096 \times 4096$  switches. The proposed model is found to be *significantly more accurate* than Mun's model. The reason for the improved results from the proposed model is the fact that the model memorizes the history of the output link for which a cell was blocked, whereas Mun's model does not.

Figure 7 and 8 show the normalized throughput and the mean delay as a function of the number of stages obtained from the proposed model, Mun's model and simulation. Figure 7 reveals that the proposed model is much more accurate than Mun's model when the switch size is large. This is because as the switch size increases, the number of stages also increase and results in increased blocking in the switch. By taking

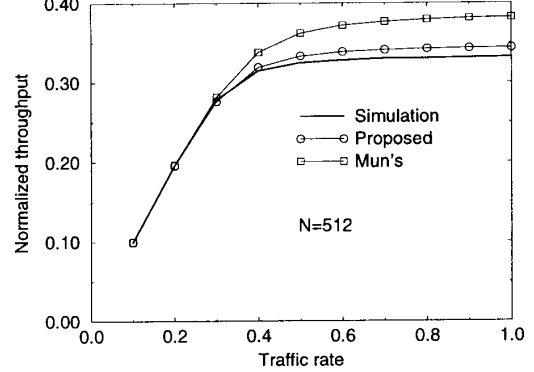


Figure 3: Throughput of a  $512 \times 512$  switch.

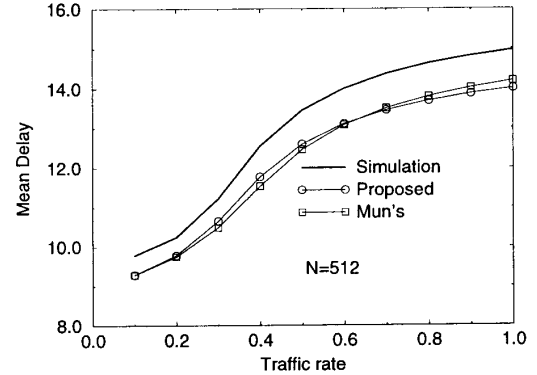


Figure 4: Delay of a  $512 \times 512$  switch.

a rigorous account of blocking in the proposed model, increased accuracy is obtained.

The throughput from the proposed model is very close to the simulation results and better than Mun's model. In Mun's model, the analysis of the acceptance probabilities and the transition rate implied that the destination of a blocked cell is randomly selected in succeeding cycles rather than pre-determined. This exemplifies the dependencies between buffer operations in consecutive cycles. Therefore, the delay of Mun's model is always higher than the proposed model and incidentally happens to be closer to the simulation results. The delay figures from Mun's model happen to be closer to simulation because two errors (due to assumptions in the model) fortunately cancel each other.

## 5 Conclusions

A new analytical model for buffered banyan type switches have been presented in this paper. Performance measurements obtained from the proposed model were compared against Mun's [1] single buffer model. The proposed model has been found to have higher accuracy. Comparison with simulation results

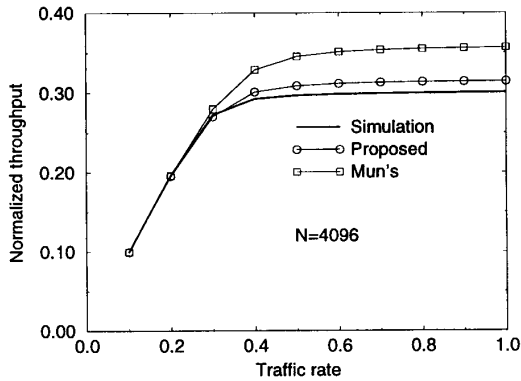


Figure 5: Throughput of a  $4096 \times 4096$  switch.

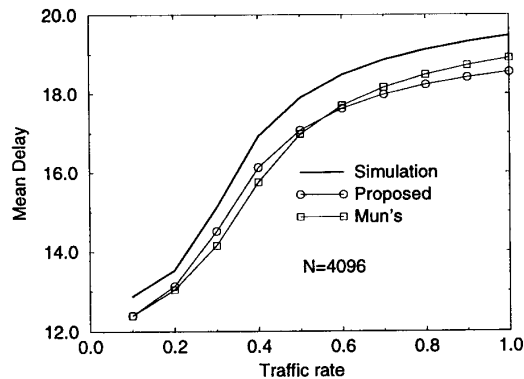


Figure 6: Delay of a  $4096 \times 4096$  switch.

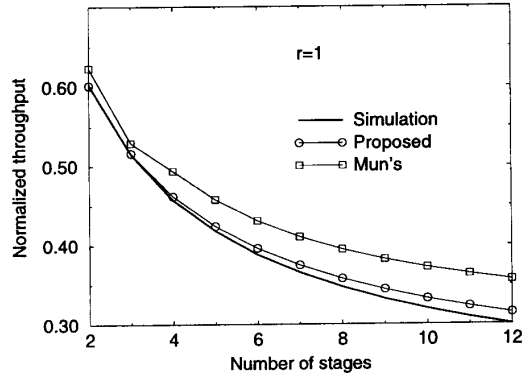


Figure 7: Throughput of switches of different sizes.

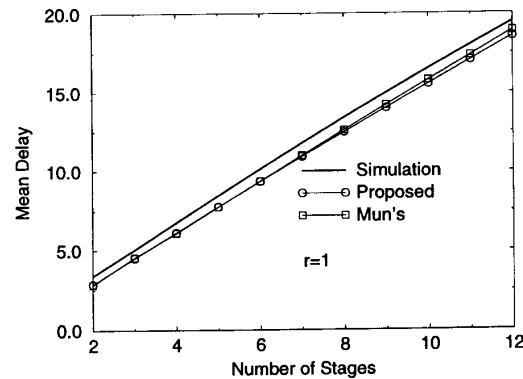


Figure 8: Delay of switches of different sizes.

revealed that the proposed model is very accurate when the switch size is large. The reason for the improved results from the proposed model is the fact that the model memorizes the history of a blocked cell, whereas other models do not.

Extension of the proposed model to multistage switches using multibuffered SEs is straight forward [6]. In [6] we show that our *multibuffered model also produces results which are more accurate* than those from [1]. The models can be extended to other types of non-uniform traffic patterns, structures, and operating conditions of switches.

## References

- [1] Y. Mun and H.Y. Youn, "Performance analysis of finite buffered multistage interconnection networks," *IEEE Transactions on Computers*, vol. 43, no. 2, pp. 153–162, February 1994.
- [2] S.H. Hsiao and R.Y. Chen, "Performance analysis of single-buffered multistage interconnection networks," *Third IEEE Symposium on Parallel and Distributed Processing*, pp. 864–867, December 1-5, 1991.
- [3] B. Zhou and M. Atiquzzaman, "Improved performance model of multibuffered multistage interconnection network under general traffic patterns," *IEEE INFOCOM '94: Conference on Computer Communications*, Toronto, Canada, pp. 1448–1455, June 12-16, 1994.
- [4] M. Atiquzzaman and M.S. Akhtar, "Performance of buffered multistage interconnection networks in non uniform traffic environment," *Seventh International Parallel Processing Symposium*, California, pp. 762–767, April 13-16, 1993.
- [5] M. Atiquzzaman and M.S. Akhtar, "Effect of non-uniform traffic on the performance of unbuffered multistage interconnection networks," *IEE Proceedings - Computer and Digital Techniques*, vol. 141, no. 3, pp. 169–176, May 1994.
- [6] C.K. Chen and M. Atiquzzaman, "An improved Markov chain model for the performance analysis of finite input buffered multistage switches," Tech. Rep. 13/94, La Trobe University, Department of Computer Science, Melbourne, July 1994.