

Modeling and Performance Evaluation of Multistage Interconnection Networks with Nonuniform Traffic Pattern*

Youngsong Mun¹ and Hyunseung Choo²

¹ School of Computing, Soongsil University, Seoul, KOREA
mun@computing.ssu.ac.kr

² School of Electrical and Computer Engineering
Sungkyunkwan University, Suwon 440-746, KOREA
choo@ece.skku.ac.kr

Abstract Even though there have been a number of studies about modeling MINs, almost all of them are for studying the MINs under uniform traffic which cannot reflect the realistic traffic pattern. In this paper, we propose an analytical method to evaluate the performance of ATM switch based on MINs under non-uniform traffic. Simulation results show that the proposed model is effective for predicting the performance of ATM switch under realistic nonuniform traffic. Also it shows that the detrimental effect of hot spot traffic on the network performance turns out to get more significant as the switch size increases.

1 Introduction

Since ATM has been adopted as a standard for broadband ISDN, many research efforts have been focused on the design of the next generation switching systems for ATM. The three main approaches employed for the design of an ATM switch are shared medium, shared memory, and space-division architecture [1]. In all these designs, the limitation on the switching size is the primary constraint in the implementation. To make a larger size ATM switch, thus, more than one system is interconnected in a multistage configuration [2].

Multistage interconnection networks (MINs) [3] constructed by connecting simple switching elements (SEs) in several stages have been recognized as an efficient interconnection structure for parallel computer systems and communication systems. There have been a number of studies investigating the performance of MINs in the literature [4-8]. However, almost all of these previous works are for studying the MINs under the uniform traffic pattern. Nonuniform traffic reflects the realistic traffic pattern of currently deployed integrated service network where a wide range of bandwidths needs to be accommodated. Therefore, the performance of the MINs under nonuniform traffic must be studied for obtaining efficient switch-based system. Even though

* This work was supported by Brain Korea 21 Project.

there have been some models considering nonuniform traffic patterns [5,7], they are not precise enough since the performance of the models has not been verified.

In this paper, we propose an analytical method to evaluate the performance of ATM switch based on MINs under nonuniform traffic. It is mainly achieved by properly reflecting the nonuniform dispatch probability in modeling the operation of each switch element. To evaluate the accuracy of the proposed model, comprehensive computer simulation is performed for two performance measures – throughput and delay. MINs of 6 and 10 stages with buffer modules holding single or multiple cells are considered for evaluation. As nonuniform traffic pattern, hot spot traffic of 3.5% and 7% are investigated. Comparison of the simulation data with the data obtained from the analytical model shows that the proposed model is effective for predicting the performance of ATM switch under realistic nonuniform traffic. The detrimental effect of hot spot traffic on the network performance turns out to get more significant as the switch size increases. For example, the throughput is about 0.3 for 6-stage switch with 3.5% hot spot traffic, while it becomes only about 0.03 for 10-stage switch.

2 The Proposed Model

2.1 Assumptions, Buffer States, and Definitions

In our models, 2×2 switching elements with the buffer modules of size m are used, and a network cycle consists of two phases. The sending buffer modules check the buffer space availability of the receiving buffer modules in the first phase. Based on the availability (and routing information) propagated backward from the last stage to the first stage, each buffer module sends a packet to its destination or enters into the blocked state in the second phase.

In each network cycle packets at the head of each buffer module (head packets) in an SE contend with each other if the destinations of them are same. Based on the status of the head packet, the state of a buffer module can be defined as follows. Figure 1 shows the state transition diagram of a buffer module in SEs.

- *State-0* : a buffer module is empty.
- *State- n_k* : a buffer module has k packets and the head packet moved into the current position in the previous network cycle.
- *State- b_k* : a buffer module has k packets and the head packet could not move forward due to the empty space of its destined buffer module in the previous network cycle.

The following variables are defined to develop our analytical model. Here $Q(ij)$ denotes the j -th buffer module in Stage- i . And its conjugate buffer module is represented as $Q(ij^c)$. Also t_b represents the time instance when a network cycle begins, while t_d represents the duration of a network cycle.

- m : the number of buffers in a buffer module.
- n : the number of switching stages. There are $n = \log_2 N$ stages for $N \times N$ MINs.

- $r_{bn}^u(ij, t) / r_{bn}^l(ij, t)$: the probability that a blocked packet at the head of $Q(ij)$ is able to move forward to the upper/lower output port during t_d . Here it is assumed that $Q(ij^c)$ is empty or in the *State-n*.
- $r_{bb}^u(ij, t) / r_{bb}^l(ij, t)$: the probability that a blocked packet at the head of $Q(ij)$ is able to move forward to the upper/lower output port during t_d . Here it is assumed that $Q(ij^c)$ also has a blocked packet.
- $P^{na}(ij, t) / P^{ba}(ij, t) / P^{bba}(ij, t)$: the probability that a buffer space in $Q(ij)$ is available (ready to accept packets) during t_d , given that no blocked packet/only one blocked packet/two blocked packets in the previous stage is destined to that buffer.
- $X_n^u(ij, t) / X_n^l(ij, t)$: the probability that a normal packet destined to the upper/lower output port is blocked during t_d .
- $X_b^u(ij, t) / X_b^l(ij, t)$: the probability that a blocked packet destined to the upper/lower output port is blocked during t_d .
- $T(ij, t)$: the probability that an input port of $Q(ij)$ receives a packet.

2.2 Calculations of Required Measures

2.2.1 Obtaining $r_n(ij, t)$

A normal packet in an SE is always able to get to the desired output port when the other buffer module is empty or destined to a different port from it. When two normal packets compete, each packet has the equal probability to win the contention. The probability that a normal packet in $Q(ij)$ does not compete with a blocked packet in the other buffer module is $r(ij)\{1 - r_x(ij^c, t)\} + \{1 - r(ij)\}r_x(ij^c, t)$. Therefore, the probabilities $r_{mn}^u(ij, t)$ is as follows and $r_{mn}^l(ij, t)$ is obtained similarly.

$$r_{mn}^u(ij, t) = r(ij)P_0(ij^c, t) + [0.5r(ij)r(ij^c) + r(ij)\{1 - r(ij^c)\}]SP_n(ij^c, t) + r(ij)\{1 - r_x(ij^c, t)\}SP_b(ij^c, t) \tag{1}$$

$r_{nb}^u(ij, t)$ and $r_{nb}^l(ij, t)$ are the probabilities that a normal packet has the same destination as the blocked one in the other buffer module and wins the contention. Thus they are as follows:

$$r_{nb}^u(ij, t) = 0.5r(ij)r_x(ij^c, t)SP_b(ij^c, t), \tag{2}$$

The probability that a buffer module is not full ($\overline{P(ij, t)}$) is simply

$$\overline{P(ij, t)} = 1 - P_{n_m}(ij, t) - P_{b_m}(ij, t). \tag{3}$$

If the originating buffer module of a packet is in $State - b_i$ ($1 \leq i \leq m$), then the destined buffer module must be in either $State - n_j$ ($1 \leq j \leq m$) or $State - b_k$ ($2 \leq k \leq m$). If it has received a packet in the previous network cycle, it can be in $State - n_j$ ($1 \leq j \leq m$) or $State - b_k$ ($2 \leq k \leq m$). If it does not have received a packet, it must be in $State - b_m$. Thus

$$P^{ba}(ij, t) = T(ij, t - 1) \times A + \{1 - T(ij, t - 1)\} \frac{P_{b_m}(ij, t)r_b(ij, t)}{P_{b_m}(ij, t)} \tag{4}$$

Here $A = \frac{\sum_{k=1}^{m-1} P_{n_k}(ij, t) + \sum_{k=2}^{m-1} P_{b_k}(ij, t) + P_{n_m}(ij, t)r_n(ij, t) + P_{b_m}(ij, t) \times r_b(ij, t)}{1 - P_0(ij, t) - P_{b_1}(ij, t)}$.

The probabilities $r_n(ij, t)$ and $r_b(ij, t)$ will be discussed later in this section. $P^{na}(ij, t)$ is obtained similarly. If the destined buffer module has not received a packet, it must be in any state except $State - n_m$. Then

$$P^{na}(ij, t) = T(ij, t - 1) \times A + \{1 - T(ij, t - 1)\} \times B \tag{5}$$

Here $B = \frac{P_0(ij, t) + \sum_{k=1}^{m-1} P_{n_k}(ij, t) + \sum_{k=1}^{m-1} P_{b_k}(ij, t) + P_{b_m}(ij, t) \times r_b(ij, t)}{1 - P_{n_m}(ij, t)}$.

For a packet to move to the succeeding stage, it should be able to get to the desired output port and the destined buffer module should be available. Thus $r_n^u(ij, t)$ is as follows and $r_n^l(ij, t)$ is obtained similarly.

$$r_n^u(ij, t) = r_{nm}^u(ij, t)P^{na}((i + 1), t) + r_{nb}^u(ij, t)P^{ba}((i + 1), t) \tag{6}$$

So $r_n(ij, t)$ is

$$r_n(ij, t) = r_n^u(ij, t) + r_n^l(ij, t) \tag{7}$$

We can calculate $r_b(ij, t)$ using the similar method.

2.2.2 Obtaining $X_n^u(ij, t)$, $X_b^u(ij, t)$, $X_n^l(ij, t)$, $X_b^l(ij, t)$, and $r_x(ij, t)$

$X_n^u(ij, t)$ is the probability that a normal packet destined to the upper output port is blocked.

$$X_n^u(ij, t) = r_{nm}^u(ij, t)\{1 - P^{na}((i + 1), t)\} + r_{nb}^u(ij, t)\{1 - P^{ba}((i + 1), t)\} + 0.5r(ij)r(ij^c)SP_n(ij^c, t) + 0.5r(ij)r_x(ij^c, t)SP_b(ij^c, t) \tag{8}$$

The first two terms in the equation above are the probabilities that the destination has no available space. The last two terms are for the case of lost contention. $X_b^u(ij, t)$ is the probability that a blocked packet destined to the upper output port is blocked again. We can calculate this probability easily by the approach employed in $X_n^u(ij, t)$.

$$X_b^u(ij, t) = r_{bn}^u(ij, t)\{1 - P^{ba}((i+1), t)\} + r_{bb}^u(ij, t)\{1 - P^{bba}((i+1), t)\} + 0.5r_x(ij)r(ij^c)SP_n(ij^c, t) + 0.5r_x(ij)r_x(ij^c, t)SP_b(ij^c, t) \quad (9)$$

$X_n^l(ij, t)$ and $X_b^l(ij, t)$ are obtained similarly.

Also $r_x(ij, t)$, which is the probability that a blocked head packet is destined to the upper output port, is calculated as follows.

$$r_x(ij, t) = \frac{P_b^h(ij, t-1)}{P_b^h(ij, t-1) + P_b^l(ij, t-1)} \quad (P_b^u(ij, t-1) + P_b^l(ij, t-1) \neq 0) \quad (10)$$

Here $P_b^u(ij, t)$ and $P_b^l(ij, t)$ are calculated as follows.

$$P_b^u(ij, t) = X_n^u(ij, t)SP_n(ij, t) + X_b^u(ij, t)SP_b(ij, t), \quad (11)$$

$$P_b^l(ij, t) = X_n^l(ij, t)P_n(ij, t) + X_b^l(ij, t)P_b(ij, t). \quad (12)$$

2.2.3 Obtaining $T(ij, t)$ and $q(ij, t)$

Due to its inherent connection property of MINs, the two buffer modules in an SE are connected to either upper or lower output ports of the SE of the previous stage. On the contrary, the buffer modules below it are connected to the lower output ports. We denote $T(ij, t)$ for the buffer modules connected to upper output ports as

$$T(ij, t) = SP_n((i-1)g, t)r_n^u((i-1)g, t) + SP_n((i-1)g^c, t)r_n^u((i-1)g^c, t) + SP_b((i-1)g, t)r_b^u((i-1)g, t) + P_b((i-1)g^c, t)r_b^u((i-1)g^c, t) \quad (13)$$

The buffer modules which are connected to lower output ports of the previous stage are obtained similarly. $T(ij, t)$ ($1 \leq i \leq n$) also has the following relation with $T(ij, t)$.

$$T(ij, t) = q(ij, t)[\overline{P(ij, t)} + P_{n_m}(ij, t)r_n(ij, t) + P_{b_m}(ij, t)r_b(ij, t)] \quad (14)$$

Finally, $q(ij, t)$ ($2 \leq i \leq n$) is obtained.

$$q(ij, t) = \frac{T(ij, t)}{\overline{P(ij, t)} + P_{n_m}(ij, t)r_n(ij, t) + P_{b_m}(ij, t)r_b(ij, t)} \quad (15)$$

2.2.4 Calculating $r(ij, t)$

$r(ij)$ is calculated by using the transformation method proposed in [7]. It is a mapping scheme that transforms the given reference pattern into a set of $r(ij)$'s which reflect the steady state traffic flow in the network. For a steady state reference pattern, we represent it in terms of destination accessing probabilities A_j , the probability that a new packet generated by an inlet chooses the output port j as its destination. Then $r(ij)$ can be represented as the conditional probability that the sum of A_j 's which are connected to the upper output port of $Q(ij)$ given the sum of A_j 's of all possible destined output ports which are connected to the upper or lower output port of $Q(ij)$. For example, $r(ij)$'s in three stage MIN are described as follows.

For the last stage:

$$r(31) = r(32) = \frac{A_1}{A_1 + A_2}, \quad r(33) = r(34) = \frac{A_3}{A_3 + A_4},$$

$$r(35) = r(36) = \frac{A_5}{A_5 + A_6}, \quad r(37) = r(38) = \frac{A_7}{A_7 + A_8}.$$

For the second stage:

$$r(21) = r(22) = r(25) = r(26) = \frac{A_1 + A_2}{A_1 + A_2 + A_3 + A_4},$$

$$r(23) = r(24) = r(26) = r(27) = \frac{A_5 + A_6}{A_5 + A_6 + A_7 + A_8}.$$

For the first stage, all $r(1j)$ ($1 \leq j \leq N$) are same:

$$r(11) = r(12) = \dots = r(18) = \frac{A_1 + A_2 + A_3 + A_4}{A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8} = \frac{A_1 + A_2 + A_3 + A_4}{A_1 + A_2 + A_3 + A_4}$$

2.3 Throughput and Delay

Normalized throughput of a MIN is defined to be the throughput of an output port of the last stage. If $Port-j$ is the upper output port of an SE, the normalized throughput in this port is as follows.

$$TNET(j, t) = SP_n(nj, t)r_n^u(nj, t) + SP_n(nj^c, t)r_n^u(nj^c, t) \tag{16}$$

$$+ SP_b(nj, t)r_b^u(nj, t) + SP_b(nj^c, t)r_b^u(nj^c, t)$$

The delay occurred for a packet at the buffer module $Q(ij)$ in the steady state is calculated by using Little's formula.

$$D(ij) = \lim_{t \rightarrow \infty} \frac{\sum_{k=1}^m k \{P_{n_k}(ij, t) + P_{b_k}(ij, t)\}}{T(ij, t)} \tag{17}$$

As delay at each output port are different, the weight of it should be considered for obtaining the mean delay. Hence the mean delay is

$$D = \sum_{j=1}^N w_j D(j) \quad (18)$$

Here w_j – the weight of *Port-j* for the mean delay – is obtained by the rate of the normalized throughput of that port as follows.

$$w_j = \lim_{t \rightarrow \infty} \frac{TNET(j)}{\sum_{k=1}^N TNET(k,t)} \quad (19)$$

3 Verification of the Proposed Model

Correctness of our model in terms of network throughput and delay is verified by comparing them with the data obtained from computer simulation for various buffer sizes and traffic conditions. For the simulation, 95% confidence interval is used and the following approaches are employed for the computer simulation.

- Each inlet generates requests at the rate of the offered input traffic load.
- The destination of each packet follows the given hot spot nonuniform traffic pattern. Here each inlet makes a fraction h of their requests to a hot spot port, while the remaining $(1 - hP^{nd}(ij,t))$ of their requests are distributed uniformly over all output ports including the hot spot port.
- If there is a contention between the packets in an SE, it is resolved randomly.
- The buffer operation is based on the FCFS principle.

Figure 2 shows the mean throughput and delay comparison of a 6-stage single buffered MIN with 7% the hot spot traffic. The offered traffic load varies from 0.1 to 1, and simulation data are obtained by averaging 10 runs. In each run, 1,000,000 iterations are taken to collect reliable data. The variations in the last 100,000 iterations are less than 0.1%. Figure 3 shows the comparison of the throughput of the hot spot port and other ports between the analytical model and computer simulation in this case. It reveals that the throughput of the hot spot port is more than two times higher than that of other ports since the access probability to the hot spot port is higher than others. Also Figures 4 and 5 show the comparison results of multiple buffer MINs. In case of uniform traffic, more buffer entries can increase the performance of MINs 10% to 20%. As identified here, in case of the nonuniform traffic, the increase in the throughput is as small as about 2% even though more buffers are added since blocking among the packets is more likely due to the nonuniform traffic. Similar result are shown in case of the 3.5% hot spot traffic.

The figures show that our models are effective for predicting the performance of MINs with realistic traffic. In case of the large sized MIN (1024×1024), the throughput of the hot spot port is always close to 1 since there always exists a packet to that

port coming from a large number of input ports. However, those of other ports are as low as less than 0.03 since blocking is so severe.

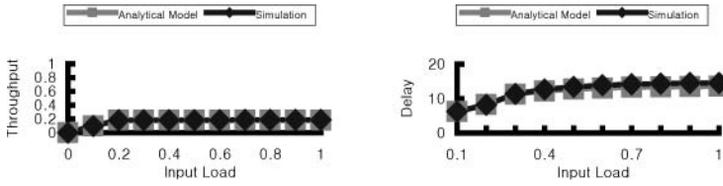
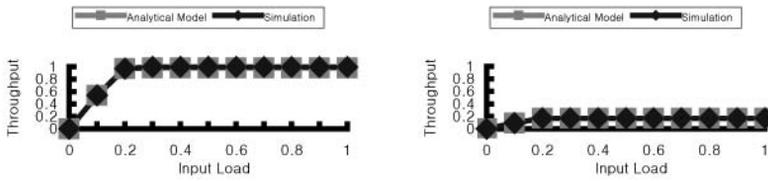


Figure 2. Comparison of the throughput and mean delay for 6-stage, single-buffered MIN delay with 7% hot spot traffic.



a) Hot spot port b) Other ports

Figure 3. Comparison of the throughput of hot spot port and other ports with 7% hot spot traffic.

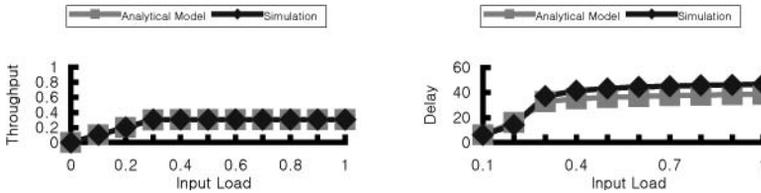
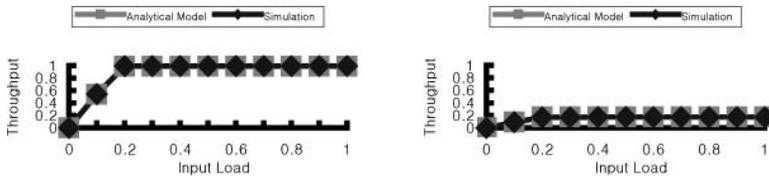


Figure 4. Comparison of the throughput and mean delay for 6-stage 4-buffered MIN with 7% hot spot traffic.



a) Hot spot port b) Other ports

Figure 5. Comparison of the throughput of hot spot port and other ports with 7% hot spot traffic.

4 Conclusion

This paper has proposed an analytical modeling method for the performance evaluation of MINs under nonuniform traffic. The effectiveness of the proposed model was verified by computer simulation for various practical MINs; 6×6 and 10×10 switches, single and 4-buffered MIN with 3.5% and 7% hot spot traffic. According to the results, the proposed model is accurate in terms of throughput and delay. The detrimental effect of hot spot traffic on the network performance turns out to get more significant as the switch size increases. For example, the throughput is about 0.3 for 6-stage switch with 3.5% hot spot traffic, while it becomes only about 0.03 for 10-stage switch. Therefore hot spot traffic needs to be avoided as much as possible for especially relatively large size switches. Performance analysis of other structures such as gigabit ethernet switches and terabit routers, or MINs for optical switching networks under nonuniform traffic are underway.

References

1. Hyoung-IL Lee, Seung-Woo Seo and Hyuk-jae Jang, "A High performance ATM Switch Based on the Augmented Composite Banyan Network," IEEE International Conference on Communications, Vol.1, pp.309-313, June 1998.
2. Muh-rong Yang and GnoKou Ma, "BATMAN : A New Architectural Design of a Very Large Next Generation Gigabit Switch," IEEE International Conference on Communications, Vol.2/3, pp.740-744, May 1997.
3. K. Hwang, *Advanced Computer Architecture: Parallelism, Scalability, Programmability*. New York: McGraw-Hill, 1993.
4. Y. C. Jenq, "Performance analysis of a packet switch based on single buffered Banyan network," IEEE J. Select. Areas Commun, vol. SAC-3, pp. 1014-1021, Dec. 1983.
5. H. Kim and A. Leon-Garcia, "Performance of Buffered Banyan Networks Under Nonuniform Traffic Patterns," IEEE Transaction on Communication, Vol. 38, No. 5, May 1990.
6. Y. Mun and H.Y. Youn, "Performance Analysis of Finite Buffered Multistage Interconnection Networks," IEEE Transaction on Computers, pp.153-162, Feb. 1994
7. T. Lin and L. Kleinrock, "Performance Analysis of Finite-Buffered Multistage Interconnection Networks with a General Traffic Pattern", ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems, San Diego, CA, pp. 68-78, May 21-24, 1991.
8. H.Y. Youn and H. Choo, "Performance Enhancement of Multistage Interconnection Networks with Unit Step Buffering," IEEE Trans. on Commun. Vol. 47, No. 4, April 1999.