International Journal of Management Science
Vol 4, No 2, November 1998

ASYMPTOTIC MAXIMUM PACKET SWITCH THROUGHPUT UNDER NONUNIFORM TRAFFIC

JEONG-HUN PARK
College of Business Administration, Chung-Ang University
Seoul 156-756, Korea

(Received March 1998; revision received July 1998)

ABSTRACT

Packet switch is a key component in high speed digital networks. This paper investigates congestion phenomena in the packet switching with input buffers. For large value of switch size N, mathematical models have been developed to analyze asymptotic maximum switch throughput under nonuniform traffic. Simulation study has also been done for small values of finite N. The rapid convergence of the switch performance with finite switch size to asymptotic solutions implies that asymptotic analytical solutions approximate very closely to maximum throughputs for reasonably large but finite N. Numerical examples show that non-uniformity in traffic pattern could result in serious degradation in packet switch performance, while the maximum switch throughput is 0.586 when the traffic load is uniform over the output trunks. Window scheduling policy seems to work only when the traffic is relatively uniformly distributed. As traffic non-uniformity increases, the effect of window size on throughput is getting mediocre.
(PACKET SWITCH; INPUT QUEUEING; ASYMPTOTIC THROUGHPUT; NONUNIFORM TRAFFIC)

1. INTRODUCTION

In the move toward high speed digital networks, fast packet switching is a key component for interconnecting a large number of transmission lines. Typically, the futuristic asynchronous transfer mode (ATM) optical networks will run at 100s Mbits/sec transmission rate. To support such backbone networks, high-speed and high efficiency are required in the packet switching operation that routes the incoming packets to appropriate output trunk lines [3]. In this paper, we investigate the congestion phenomena in the packet switching and develop math-

* This research was supported by the Chung-Ang University Research Grants in 1997.

ematical models to analyze the packet switch throughput under nonuniform traffic.

Packet switch is a box that routes the packet on its N inputs to the appropriate N outputs. Figure 1 illustrates simple crossbar switch fabrics with N^2 switch points. The routing information contained in the header of arriving packets (i.e., the first 5 bytes of each cell in ATM networks) is used to set up certain paths from inputs to outputs. We assume that the switch operates synchronously with fixed-size packets (i.e., 53 byte cells in ATM networks). Packet arrivals to the switch at each time slot are unscheduled as is the case in ATM networks. Thus, if there are more than one packet contending for the same output, packets may have to be buffered until they are connected to the appropriate output. The location and size of buffers depend on the switch architecture and the input traffic characteristics. Figure 1 illustrates a switch architecture with input buffers. Arriving packets join at the end of input buffers. The buffers are served on a FIFO discipline, so that at the beginning of each time slot only the packets at the heads of buffers contend for connection to outputs. If all packets are contending for all different outputs, there is no conflict and the switch allows each packet to pass through to an appropriate output. If many packets are addressed to the same output during a time slot, the switch allows only one packet to pass through and the other packets have to wait until the next time slot and contend again. This so called head-of-line (HOL) blocking is a key source of congestion in the packet switch with input queueing, i.e., a packet in a FIFO queue has no chance to access an available output port because the packet ahead of it in the buffer is blocked.

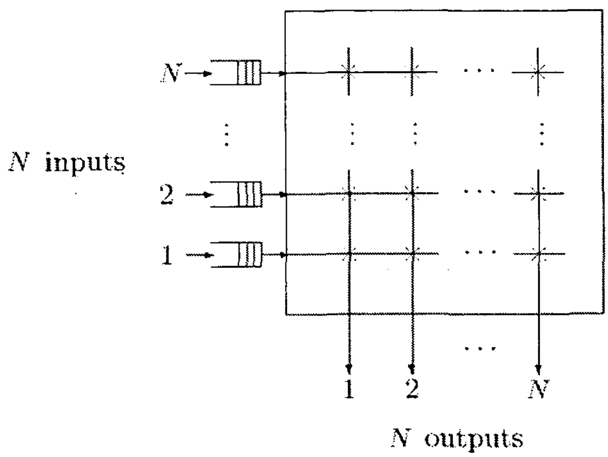

Figure 1. Crossbar Packet Switch Architecture with Input Queueing

Performance analyses for crossbar architecture packet switches can be found in Karol et al. [9], Hluchyj and Karol [7], Oie et al. [13], Fuhrmann [5], and Thomas [14]. All of their works are based on the uniform traffic assumption. It is well known in literature that for high value of N, the maximum possible switch throughput (the utilization of output trunk per input trunk) is 0.5858 when the traffic load is uniform over the output trunks [5, 7, 9, 13, 14]. However, in real ATM networks, a wide range of bandwidths need to be accommodated to serve integrated multimedia data and therefore, incoming packet traffic could be typically non-uniformly addressed to output trunks. That is, more packets may go to some of output ports than others. Hot spot traffic is an another example of nonuniform traffic instance, in which many sources try to communicate with a popular destination at the same time. For example, many telephone callers may contend to call a popular location; many nodes may report synchronously some information to the network control center for administrative purposes; many nodes may contend to connect to a popular web site for information fetching. Such non-uniformity in traffic pattern could result in serious degradation in packet switch performance.

For the past decade, numerous studies have been reported for the switch performance under nonuniform traffic situation. As a pioneer work, Wu [16] first presented an analysis of a single buffer banyan network and proposed an approximate analysis under nonuniform traffic conditions. Kim and Leon-Garcia [10] presented an alternative method of evaluating the performance of input buffered banyan networks. They modeled each input buffer of a switching element as a Markov chain and the relationship between switching element is described by average flow constraint. Including some blocked cells and link state information, Lee [12] and Atiquzzaman and Akhtar [1] provide more accurate results. For input ports using shared buffer slots, Valdimarsson [15] presented an exact model for very small switch elements of 2 and approximate model for large switch size. The nonuniform switch models with input buffers that have been proposed are for switches with limited numbers of input buffers only. Markov chain approaches adopted by those authors are computationally intensive to be used except for very small switch size. The switch performance under nonuniform traffic with shared concentration and output queueing buffers can be found in Chen and Mark [2]. Using fluid-flow approximation for incoming cells, Collier and Kim [4] analyzed the switch performance under non-homogeneous, correlated bursty traffic. All the work mentioned above has shown in general that nonuniform traffic has a detrimental effect on the performance of the network.

In this paper, we present a performance analysis of the input buffered switch with simple crossbar architecture shown in Figure 1 under the assumption that independent, statistically identical packet traffic arrives on each input buffers and the incoming packets may have different probability of being addressed to N output trunks. Our model is different from the previous nonuniform models reported in the literature in that we allow infinite numbers of input buffers, window scheduling policy, and we derive exact asymptotic and some closed-form solutions for the switch throughput under nonuniform traffic. To authors knowledge, such analytical results have not been reported yet in literature. Employing queueing analogy, we develop mathematical systems and derive an asymptotic solution of the maximum switch throughput for infinite switch size. Through simulation study, we then show very rapid convergence of the switch performance with finite switch size to asymptotic results. We also investigate the effect of nonuniform traffic and/or window scheduling policy on the packet switch throughput.

2. MATHEMATICAL MODEL FOR MAXIMUM THROUGHPUT ANALYSIS

To determine the maximum throughput of the packet switch, we assume that all input queues are saturated, i.e., packets are always waiting in every input FIFO queue. Whenever a packet at the head of the input queues is transmitted through the switch, a new packet behind moves ahead and replaces it. Following a similar approach as in Hluchyj and Karol [7], we define B_t^i as the number of packets at the heads of the input queues destined for output trunk i in the tth time slot, but not selected to pass through the switch. Recall that when many packets at the head of the input queues are addressed to the same output i during a time slot t, the switch allows only one packet to pass through and the other B_t^i packets have to wait until the next time slot and contend again. We also define A_t^i as the number of packets moving to the heads of the input queues during a time slot t and destined to output trunk i. Note that a packet can move to the heads of the input queues if and only if a packet in that queue is transmitted through the switch in the previous $(t-1)$st time slot. Therefore, it follows

$$B_{t+1}^i = \max[0,\ B_t^i + A_{t+1}^i - 1] \tag{1}$$

Note that (1) has the same form as the fundamental queueing relation for a single-server queueing system [6]. Although B_t^i does not represent any physical queue, it constitutes a virtual queue for packets blocked at the heads of input queues and waiting for next contention to access the packet switch.

For incoming packets, we assume that each packet at the heads of the input queues has the probability $p_i\,(\sum_{i=1}^{n} p_i = 1)$ of being addressed to the output i. For a special case when all $p_i = 1/N\,(i=1,2,\ldots,N)$, the traffic is assumed to be uniformly distributed. In typical traffic situations however, each p_i has different values and that represents nonuniform traffic in our formulation. Note that the number A_t^i follows a binomial distribution, that is

$$\Pr[A_{t+1}^i = k] = \binom{F_t}{k} p_i^k (1-p_i)^{F_t-k}, \qquad k=0,1,2,\ldots,F_t, \tag{2}$$

$$F_t = N - \sum_{i=1}^{N} B_t^i, \tag{3}$$

where F_t represents the number of packets transmitted through the switch during a time slot t. Let $\overline{F_t}$ and $\overline{B_t^i}$ be the mean of F_t and B_t^i, respectively. Then we have

$$\rho_t = \frac{\overline{F_t}}{N} = 1 - \frac{1}{N}\sum_{i=1}^{N} \overline{B_t^i}. \tag{4}$$

Note that ρ_t is the normalized switch throughput, or the switch utilization factor defined as the utilization of output trunk per input trunk.

Now we will consider a special case when $N \to \infty$. In equilibrium, the one-step transition probability matrix P for the virtual queue size B^i is given from (1) as

$$P = \begin{bmatrix} a_0 + a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

where $a_n = \Pr[A^i = n]$. Therefore, following the standard approach in queueing analysis (see Gross and Harris [6], for example), we obtain the probability generating function (PGF) of B^i as

$$B^i(z) = \lim_{N\to\infty} \frac{(1-Np_i\rho)(1-z)}{A^i(z) - z}. \tag{5}$$

The PGF of A^i is given from (2) as

$$A^i(z) = \lim_{N\to\infty} \sum_{k=0}^{N} z^k \Pr[A^i = k]$$

$$
\begin{aligned}
&= \lim_{N\to\infty} \sum_{k=0}^{N} z^k \sum_{l=k}^{N} \Pr[A^i = k/F = l] \Pr[F = l] \\
&= \lim_{N\to\infty} \sum_{l=0}^{N} \sum_{k=0}^{l} z^k \Pr[A^i = k/F = l] \Pr[F = l] \qquad (6) \\
&= \lim_{N\to\infty} \sum_{l=0}^{N} (1 - p_i + z p_i)^l \Pr[F = l].
\end{aligned}
$$

The mean of B^i can be derived from (5) and (6) as

$$\overline{B^i} = \frac{dB^i(z)}{dz}\bigg|_{z=1} = \lim_{N\to\infty} \frac{(Np_i\rho)^2}{2(1-Np_i\rho)}. \qquad (7)$$

The equation (7) can also be derived from the following argument. Since A^i follows a binomial distribution as is shown in (2), the number A^i of packets addressed to the output i follows a Poisson distribution with arrival intensity $\lambda_i = Np_i\rho$ as $N\to\infty$, i.e., the limiting process of the binomial produces the Poisson distribution [8, 9]. Thus, the number of packets at the heads of the input queues destined for output i can be derived from a $M/D/1$ queueing analogy where the arrival intensity is λ_i and the constant service time is 1. From theories in the queueing literature, we have (7) [6, 13].

From (4) and (7), and letting $N\to\infty$, we derive

$$\frac{\rho^2}{2} \lim_{N\to\infty} \sum_{i=1}^{N} \frac{Np_i^2}{(1-Np_i\rho)} = 1 - \rho. \qquad (8)$$

Therefore, for large values of N, the maximum throughput ρ will be given from (8) when values for p_i are specified. For a special case when $p_i = 1/N$ $(i = 1, 2, \ldots, N)$, i.e., the traffic is uniformly distributed, from (8) one can easily see that the switch throughput is $\rho = 2 - \sqrt{2} = .5858$, which is identical to Hluchyj and Karol [7]. For small values of N, Table 1 and Figure 2 show simulation results of maximum throughput for finite N. The simulation program is based on SIMLIB package in Law and Kelton [11]. Note the rapid convergence of the switch throughput with finite switch size to the asymptotic throughput of .5858.

Table 1. Maximum Throughput under Uniform Traffic

N	1	2	4	8	16	32	64	128	256	∞
ρ	1.0000	.7500	.6552	.6181	.6017	.5937	.5898	.5878	.5868	.5858

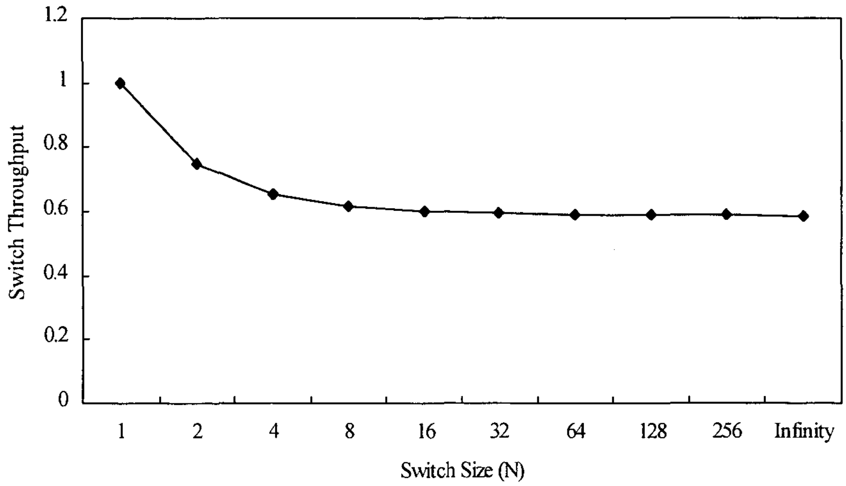

Figure 2. Maximum Throughput under Uniform Traffic

3. MAXIMUM THROUGHPUT UNDER NON-UNIFORM TRAFFIC

Nonuniform traffic could be any traffic patterns other than uniform, or $p_i = 1/N (i = 1, 2, \ldots, N)$. Theoretically speaking, for any nonuniform traffic pattern of p_i, the maximum switch throughput ρ could be derived from the analytic solution of the equation (8), if it exists. Since the equation (8) involves infinite sum of p_i terms, we need to somehow rearrange the equation for practical use. For this purpose, of the many possible nonuniform traffic patterns, we consider the following nonuniform traffic situation in which more packets may go to some portions of infinite output ports than others. Mathematically, we consider non-uniform traffic case where $100 a_j$ % $(0 \leq a_j \leq 1,\ \sum a_j = 1)$ of output lines are of type j $(j = 1, 2, \ldots, K)$ and

$$p_i = m_j p_0, \quad \text{if} \quad i \in \text{type } j \ (i = 1, 2, \ldots, \infty, \quad j = 1, 2, \ldots, K), \tag{9}$$

in which p_0 is the normalization factor, or

$$p_0 = [\sum_{i=1}^{\infty} \frac{p_i}{p_0}]^{-1} = \lim_{N \to \infty} [N \sum_{j=1}^{K} a_j m_j]^{-1}.$$

The nonuniform traffic pattern in (9) implies that each incoming packet has different traffic weight m_j than others of being addressed to each a_j portion of infinite output ports.

From (8), (9), and the sum-to-one condition for p_i, we have for sufficiently large N,

$$\frac{\rho^2}{2}\sum_{j=1}^{K}\frac{a_j(m_jc)^2}{(1-m_jc\rho)}=1-\rho, \tag{10}$$

where

$$c=[\sum_{j=1}^{K}a_jm_j]^{-1}.$$

Note that with the nonuniform traffic pattern of the form in (9), the infinite sum expression in the equation (8) is transformed into finite sum in (10). Therefore, once nonuniform traffic parameters, a_j and m_j $(j=1,2,\ldots K)$ are specified, the maximum switch throughput ρ can be easily derived from (10).

In a special case when $m_j \rightarrow \infty$ for a specific tagged j, i.e., all traffics are addressed to $100\,a_j\%$ output lines of type j, we have $c \rightarrow 0$ and $m_jc \rightarrow 1/a_j$. The equation (10) is further simplified in this case as

$$\frac{\rho^2}{2(a_j-\rho)}=1-\rho,$$

or

$$\rho=(1+a_j)-\sqrt{1+a_j^2}. \tag{11}$$

Note that the switch throughput ρ in this case has the minimum possible value of 0 when $a_j \rightarrow 0$, and the maximum possible value of .5858 when $a_j \rightarrow 1$, an equivalent of uniform traffic case.

Let us consider another simple non-uniform traffic case of two output types where each incoming packet traffic addressed to $100a\ \%$ $(1 \leq a \leq 1)$ output lines is m times as much as the traffic addressed to the remaining $100(1-a)\ \%$ output lines. Hot spot traffic is an example, in which many sources try to communicate with popular destinations at the same time. From (9) and (10), we have for sufficiently large N,

$$\frac{\rho^2}{2}[\frac{a(mc)^2}{(1-mc\rho)}+\frac{(1-a)c^2}{(1-c\rho)}] = 1-\rho, \tag{12}$$

where

$$c = \frac{1}{a(m-1)+1}.$$

Table 2. Maximum Throughput under Nonuniform Traffic

	a=.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
m=1	.5858	.5858	.5858	.5858	.5858	.5858	.5858	.5858	.5858	.5858
2	.4884	.4950	.5055	.5172	.5292	.5412	.5529	.5643	.5752	.5858
3	.3699	.4033	.4337	.4614	.4867	.5099	.5311	.5509	.5690	.5858
4	.3037	.3508	.3924	.4295	.4626	.4923	.5192	.5435	.5656	.5858
5	.2628	.3179	.3664	.4093	.4473	.4813	.5116	.5389	.5635	.5858
10	.1796	.2501	.3123	.3671	.4154	.4581	.4958	.5293	.5592	.5858
∞	.0950	.1802	.2560	.3230	.3820	.4338	.4793	.5194	.5546	.5858

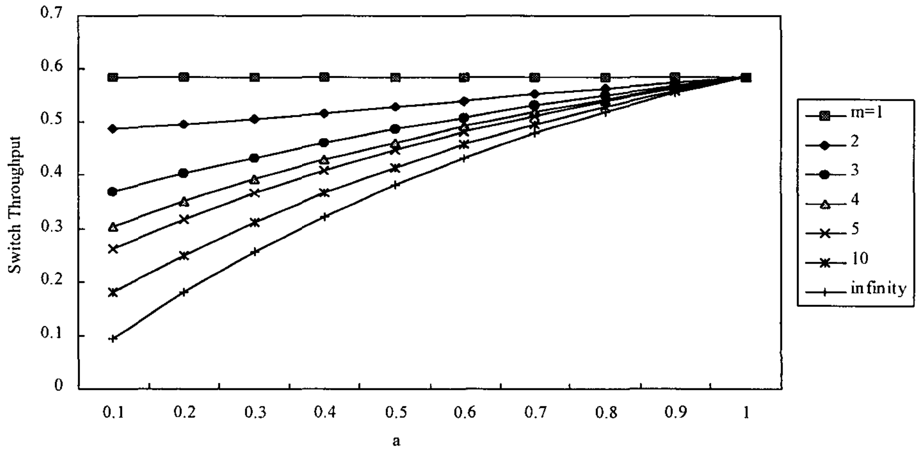

Figure 3. Maximum Throughput under Nonuniform Traffic

Table 2 and Figure 3 show the maximum throughput ρ with various combination of a and m. As is shown, the effect of traffic non-uniformity on the switch performance seems significant, which is often overlooked in packet switch performance evaluations. Since the maximum throughput presented in Table 2 are asymptotic analytical solutions, simulation has been done for small values of N to see the effect of the switch size on throughput. Table 3 and Figure 4 present simulation results for various switch size N when a=0.5. The rapid convergence of the switch performance with finite switch size to asymptotic solutions implies that for reasonably large N, asymptotic solutions approximate very closely to maximum throughputs for finite N.

Table 3. Maximum Throughput for Finite N (a =.5)

	N=1	2	4	8	16	32	64	128	∞
m=1	1.0000	.7500	.6552	.6181	.6017	.5937	.5898	.5878	.5858
2	1.0000	.7000	.6033	.5640	.5463	.5377	.5334	.5314	.5292
3	1.0000	.6498	.5577	.5199	.5031	.4947	.4907	.4888	.4867
4	1.0000	.6175	.5300	.4944	.4781	.4702	.4664	.4646	.4626
5	1.0000	.5959	.5126	.4783	.4622	.4548	.4510	.4493	.4473
10	1.0000	.5494	.4757	.4440	.4294	.4223	.4189	.4172	.4154
∞*	1.0000	.5000	.4375	.4082	.3947	.3883	.3853	.3836	.3820

* Set as m=100,000 for simulation purpose

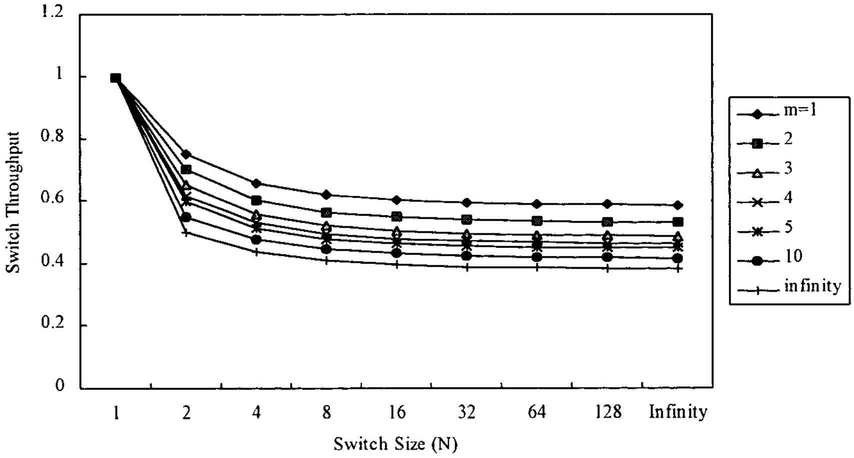

Figure 4. Maximum Throughput for Finite N (a =.5)

4. MAXIMUM THROUGHPUT WITH WINDOW SCHEDULING

In previous sections, we analysed switch throughput on the assumption that packets in the input buffers are served on a FIFO discipline, so that at the beginning of each time slot, only the packets at the heads of buffers contend for connection to outputs. In this section, we relax this strict FIFO discipline and consider a window scheduling policy in which first w packets in each input buffers can contend for connection to output lines. We allow that more than one packet can be retrieved from each input queue, i.e., all input buffers are fully shared. Hluchyj and Karol [7] considered a similar window scheduling policy under uniform traffic. From simulation study, they showed the maximum throughput can be increased by employing window scheduling under uniform traffic. Thomas [14] proposed bifurcated input queueing, in which each input line maintains w separate queues and obtained maximum throughput as $\rho=(w+1)-\sqrt{w^2+1}$ under uniform traffic. Bifurcated input queueing has a similar effect as window scheduling. In this section, we consider a window scheduling policy under non-uniform traffic and derive analytical results.

Let w be the window size. The equation (3) can be rewritten in this case as

$$F_t = wN - \sum_{i=1}^{N} B_t^i. \tag{13}$$

Assume that each packet inside the window (total ωN packets) of the input buffers has the probability p_i of being addressed to the output i. We define B_t^i as the number of packets inside the window of the input buffers destined for output trunk i in the tth time slot, but not selected to pass through the switch. Then, the number of packets inside the window w of the input buffers destined for output i can be derived from a $M/D/1$ queueing analogy where the arrival intensity is $\lambda_i = \rho c_i$ and a constant service time is 1. The mean of B^i is given same as the equation (7). Therefore, we derive from (7) and (13),

$$\frac{\rho^2}{2} \lim_{N\to\infty} \sum_{i=1}^{N} \frac{Np_i^2}{(1-Np_i\rho)} = w-\rho. \tag{14}$$

The equation (14) is a generalization of (8) with window scheduling policy. The maximum throughput ρ can be determined from (14) as a function of w and p_i.

First, let us consider a simple case of uniform traffic situation where all $p_i = 1/N$. In this case, we have

$$\frac{\rho^2}{2(1-\rho)} = w-\rho,$$

or

$$\rho = (w+1)-\sqrt{w^2+1}. \tag{15}$$

This result is identical to Thomas [14]. Table 4 shows maximum throughput under uniform traffic with various values of window size w. We can see that there is a big increase in maximum throughput by changing w from 1 to 2, 4, and 8, with diminishing improvements thereafter.

Table 4. Maximum Throughput with Window Scheduling under Uniform Traffic

w	1	2	4	8	16	∞
ρ	.5858	.7639	.8769	.9377	.9688	1

For the nonuniform traffic pattern of the form in (9), following the same approach in Section 3, the equation (10) is rewritten as

$$\frac{\rho^2}{2}\sum_{j=1}^{K}\frac{a_j(m_jc)^2}{(1-m_jc\rho)}=w-\rho, \tag{16}$$

where

$$c=[\sum_{j=1}^{K}a_jm_j]^{-1}.$$

Note that the equation (16) is a generalization of (10) with window scheduling policy. In a special case where $m_j\to\infty$ for a specific tagged j, i.e., all traffics are addressed to $100a_j\%$ output lines of type j, we have $c\to 0$ and $m_jc\to 1/a_j$. Therefore the equation (16) is further simplified in this case as

$$\frac{\rho^2}{2(a_j-\rho)}=w-\rho,$$

or

$$\rho=(w+a_j)-\sqrt{w^2+a_j^2}. \tag{17}$$

policy. As $w\to\infty$ and $m_j\to\infty$, we also have $\rho\to a_j$. Table 5 and Figure 5 show numerical results for maximum throughput possible with various values of w and a_j when $m_j\to\infty$. The results show that as traffic non-uniformity increases with smaller values of a_j, the effect of window size on throughput is getting mediocre. That strongly implies that window scheduling policy may be helpful only for reasonably uniform traffic.

Table 5. Maximum Throughput with Window Scheduling under Nonuniform Traffic ($m_j \rightarrow \infty$)

	a_j=.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
w=1	.0950	.1802	.2560	.3230	.3820	.4338	.4793	.5194	.5546	.5858
2	.0975	.1900	.2776	.3604	.4384	.5119	.5810	.6459	.7068	.7639
4	.0988	.1950	.2888	.3800	.4689	.5553	.6392	.7208	.8	.8769
8	.0994	.1975	.2944	.3900	.4844	.5775	.6694	.7601	.8495	.9377
16	.0997	.1988	.2972	.3950	.4922	.5888	.6847	.7800	.8747	.9688
∞	.1000	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9000	1.0000

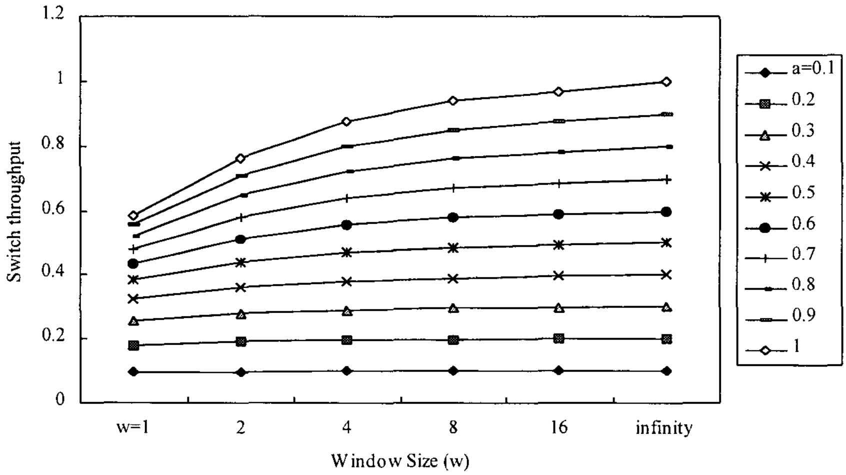

Figure 5. Maximum Throughput with Window Scheduling under Nonuniform Traffic ($m_j \rightarrow \infty$)

We will consider another simple non-uniform traffic case of two output types as is discussed in Section 3. With window scheduling policy, the equation (12) is rewritten as

$$\frac{\rho^2}{2}\left[\frac{a(mc)^2}{(1-mc\rho)}+\frac{(1-a)c^2}{(1-c\rho)}\right] = w-\rho. \qquad (18)$$

In a special case where $m \rightarrow \infty$, i.e., all traffics are addressed to $100\,a\%$ output lines of type j, the equation (18) is further simplified in this case as

$$\rho = (w+a)-\sqrt{w^2+a^2}.$$

As $w \rightarrow \infty$, we have

$$\rho = \frac{1}{mc} = \frac{a(m-1)+1}{m}.$$

As $w \to \infty$ and $m \to \infty$ simultaneously, we have $\rho \to a$. Table 6 and Figure 6 present numerical results for maximum throughput with various values of w and m when $a = 0.1$. Again, the results reveal that as traffic non-uniformity increases with larger values of m, the effect of window size on throughput is diminishing. That conforms to our previous observation that window scheduling policy may be helpful only for reasonably uniform traffic.

Table 6. Maximum Throughput with Window Scheduling under Nonuniform Traffic ($a = 0.1$)

	w=1	2	4	8	16	∞
m=1	.5858	.7639	.8768	.9377	.9688	1.0000
2	.4884	.5299	.5418	.5463	.5482	.5500
3	.3699	.3878	.3945	.3974	.3987	.4000
4	.3037	.3157	.3207	.3229	.3240	.3250
5	.2628	.2723	.2763	.2782	.2791	.2800
10	.1796	.1850	.1876	.1888	.1894	.1900
∞	.0959	.0975	.0988	.0994	.0997	.1000

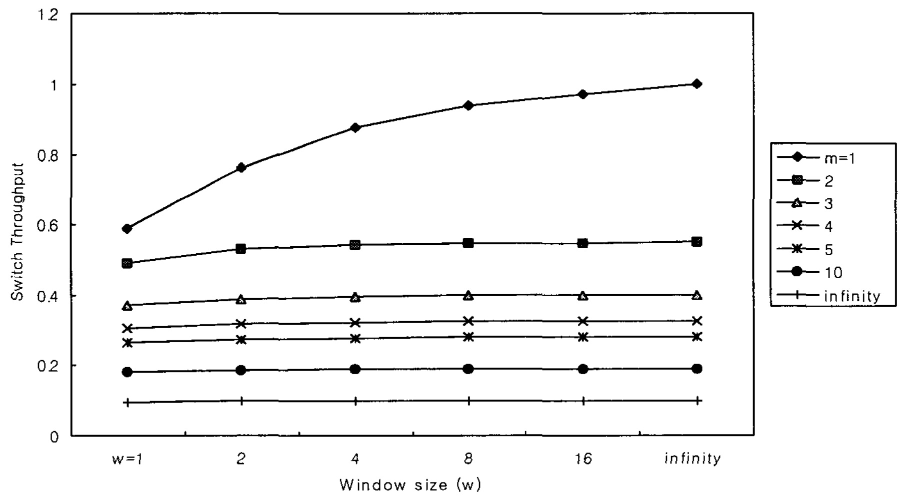

Figure 6. Maximum Throughput with Window Scheduling under Nonuniform Traffic ($a = 0.1$)

5. CONCLUSIONS

Mathematical models have been developed to evaluate maximum throughput of packet switch with input buffers under nonuniform traffic. While the maximum possible switch throughput is 0.5858 when the traffic load is uniform over the output trunks, our results reveal that when the incoming traffic is non-uniformly addressed to output trunks, such non-uniformity in traffic pattern could result in serious degradation in packet switch performance, which is often overlooked in packet switch performance evaluations. Since our analyses are based on asymptotic solutions for large value of N, we performed simulation study for small value of finite N. As is shown, the rapid convergence of the switch performance with finite switch size to asymptotic results implies that asymptotic solutions approximate very closely to maximum throughputs for reasonably large but finite N. Window scheduling policy seems to work only when the traffic is relatively uniformly distributed. As traffic non-uniformity increases, the effect of window size on throughput is getting mediocre. Therefore, window scheduling policy seems not promising to cope with nonuniform traffic environment. Oie et al. [13] studied packet switches with input and output buffers under uniform traffic, in which switch fabric runs s $(1 \leq s \leq N)$ times faster than the input queueing packet switches. We propose to investigate packet switches with input and output buffers under nonuniform traffic as future research agenda.

REFERENCES

[1] Atiquzzaman, M. and M. S. Akhtar, "Performance of Buffered Multistage Interconnection Networks in Nonuniform Traffic Environment," *7th International Parallel Processing Symposium*, Newport Beach, California, April (1993), pp.762-767.

[2] Chen, D. X. and J. W. Mark, "A Buffer Management Scheme for the SCOQ Switch under Nonuniform Traffic Loading," *IEEE Trans. Communication*, Vol.42, No.10 (1994), 2899-2907.

[3] Choudhury, G. L., D. M. Lucantoni, and W. Whitt, "Squeezing the Most Out of ATM," *IEEE Trans. Communication*, Vol.44, No.2 (1996), 203-217.

[4] Collier, B. R. and H. S. Kim, "Performance of Multistage ATM Switch Architectures under Nonuniform Bursty Traffic," *IEEE Infocom*, (1995), pp. 667-674.

[5] Fuhrmann, S. W., "Performance of a Packet Switch with Crossbar Architecture," *IEEE Trans. Communication*, Vol.41, No.3 (1993), 486-491.

[6] Gross, D. and C. M. Harris, *Fundamentals of Queueing Theory*, 2nd Ed., John Wiley & Sons, Inc., New York, (1985).

[7] Hluchyj, M. G. and M. J. Karol, "Queueing in High Performance Packet Switching," *IEEE J. Select. Areas Communication*, Vol.6, No.9 (1988), 1587-1579.

[8] Hogg, R. and A. T. Craig, *Introduction to Mathematical Statistics,* 4th Ed., Macmillan Publishing Co., Inc., New York, (1978).

[9] Karol, M. K., M. G. Hluchyj and S. P. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Trans. Communication*, Vol.35, No.12 (1987), 1347-1356.

[10] Kim, H. S. and A. Leon-Garcia, "Performance of Buffered Banyan Networks under Nonuniform Traffic Pattern," *IEEE Trans. Communication*, Vol.38, No.5 (1990), 648-658.

[11] Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis*, 2nd Ed., McGraw-Hill, (1991).

[12] Lee, T. H., "Analytic Models for Performance Evaluation of Single Buffered Banyan Networks under Nonuniform Traffic," *IEEE Proceedings, Part E, Computers and Digital Techniques*, Vol.138, No.1, (1991), pp.41-47.

[13] Oie, Y., M. Murata, K. Kubota, and H. Miyahara, "Performance Analysis of Nonblocking Packet Switch with Input and Output Buffers," *IEEE Trans. Communication*, Vol.40, No.8 (1992), 1294-1297.

[14] Thomas, G., "Bifurcated Queueing for Throughput Enhancement in Input-Queued Switches," *IEEE Communications Letters*, Vol.1, No.2 (1997), 56-57.

[15] Valdimarsson, E., "Queueing Analysis for Shared Buffer Switching Networks for Nonuniform Traffic," *IEEE Infocom*, (1995), pp.8-15.

[16] Wu, L. T., "Mixing Traffic in a Buffered Banyan Network," *Proceedings of 9th Data Communication Symposium*, Whistler Mountain, British Columbia, September, (1985).