Input Queued Packet Switches for Variable Length Packets : A Continious Time Analysis

D. Manjunath
Dept. of Electrical Engg.
Indian Institute of Technology
Powai, Mumbai 400 076 INDIA
Ph: +91-22-5767427
dmanju@ee.iitb.ernet.in

Biplab Sikdar
Dept. of ECSE
Rensselaer Polytechnic Institute
Troy, NY 12180 USA
Ph: +1-518-276-8289
bsikdar@networks.ecse.rpi.edu

Abstract— **We consider a general input queued $M \times N$ switch operating in continuous time. Analysis of such switches is important in the context of IP switches where the packet arrival process and packet lengths are drawn from continuous distributions. An approximate model to obtain the delay throughput characteristics of this switch is developed. The model is general enough to allow non uniform arrival rates, routing probabilities and output evacuation rates. After developing the general model, we consider $N \times N$ switches with uniform arrivals, speedup in the switch to allow multiple evacuations from input head-of-lines by an output and also the effect of hotspots in the output. All our analytical models are compared against simulation models. We also comment on some of the unexplained results in [3] where the first such system was considered.**

I. Introduction and Problem Statement

In this paper we consider space division packet switches switching variable length packets. Packet lengths and interarrival times are assumed to be random and drawn from continuous distributions. Such switches occur in processor memory interconnections in multiprocessor systems [1] and with the emergence of IP switching [6], such a continuous time analysis of space division packet switches becomes relevant and important. In [3] Fuhrman presents an analysis of such switches by considering a $M \times N$ crossbar switch with variable length packets as inputs. Using a saturation analysis with i.i.d. Poisson arrival processes at each node and uniform routing assumptions it is shown that the maximal throughput per port is $M/(M+N+1)$. A state dependent server model to obtain the service rate when there are i packets in the system and the use of this to model the switch as an M/M/1 queue is also given in [3]. For fixed packet lengths and slotted operation of the switch (with typically slot size equal to packet length), a discrete time analysis is possible. In [8] Patel analyses switches with no buffers under Bernoulli arrivals where output contention is resolved by dropping all but one of the contending packets. The maximum throughput under this model is $1-1/e \approx 0.63$. In [5], Karol et al show that the saturation throughput of an input queued switch is $2-\sqrt{2} \approx 0.586$. Output queued switches can provide a throughput of 1.0 but this requires that a $M \times N$ switch operate at M times the line rate. To reduce this increased complexity and obtain throughput close to that of an output queued switch, two alternatives are available - in each slot switch packets from head of lines of upto L inputs to an output [7], or operate the switch fabric at L times the line rate [2]. For a given L, the latter performs better than the former method.

In this paper, we consider an unslotted $M \times N$ input queued packet switch operating in continuous time. Packets arrive at input port i according to a Poisson process of rate λ_i and choose a destination j with probability p_{ij}. The rate at which a packet is evacuated from an input queue by output port j is μ_j. μ_j is the line rate on output port j and there is no queueing at the output. Our models can be easily generalized to relax this assumption but we will not pursue it in this paper. Service from the input queues is FCFS. When a packet moves to the head of its queue, if its destination is busy, the packet will wait at the head of the input queue till the destination output port is free and chooses to evacuate the packet. When an output port finishes service, of the packets that are waiting at the head of the queues of the inputs, the packet that was blocked first is served first. Service in random order, round robin or processor sharing disciplines are also possible but we do not investigate them. Finally, the packet switch is assumed to be internally non blocking. We will be drawing parallels with the results from a discrete time analysis of an input queued fast packet switch with Bernoulli arrivals of rate λ_i at port i and unit packet length.

The arrival rate to each output port, Λ_j and the load on output port j, ρ_j, are given by

$$\Lambda_j = \sum_{i=1}^{M} \lambda_i p_{ij} \qquad \rho_j = \frac{\Lambda_j}{\mu_j} \tag{1}$$

In Section II the analytical model will be developed using results from the analysis techniques used in queueing networks with finite buffer nodes. Section III extends this analysis to the case where the output ports are capable of simultaneous evacuation of multiple packets like the discrete time switch of [7]. Section IV investigates the effects of hotspots on the switch performance. In Section V we conclude by commenting on the following two observations made in [3]. Once a packet moves to the head of the input queue, it enters into a virtual queue corresponding to that of the destination output port. This queue could have a service discipline other than FIFO. The analytical model of

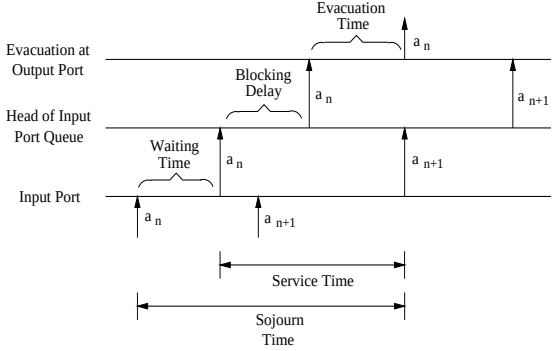


Fig. 1. Time diagram for the sojourn time in the switch. a_n represents the n^{th} arrival to the input queue and all times shown in this figure correspond to this packet.

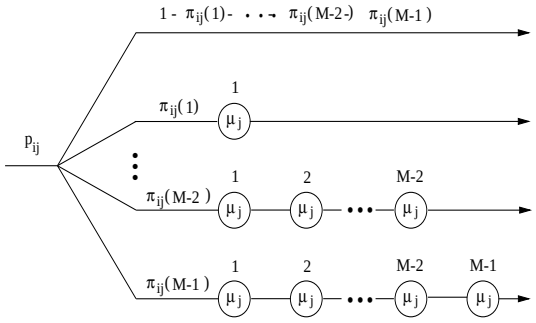


Fig. 2. Phase-type distribution for blocking delay of a packet at head-of-line of input port i and destined for output port j. A k-stage Erlangian is encountered with probability $\pi_{ij}(k)$, the probability that there are k packets in the "queue" of output port j.

[3] shows them to be equivalent but the simulation results clearly show that FIFO has a better delay performance. We explain this observation with our analytical model. Also, in [3] it is conjectured that the switch performance is symmetric in M and N. We comment on this conjecture too.

II. Analytical Model

The approximate analytical model that we develop here is based on the analysis techniques of feedforward open queueing networks with blocking [9]. Assume that the M input nodes have infinite buffer capacity. An arriving packet has to wait in two queues. First it has to wait in the input queue till it moves to the head of the line. Second, at the head of line (HOL), it has to wait in a virtual queue of the output port till its evacuation by the destination output port begins. It is easy to see that the total delay at the HOL is the service time of the input queue. Therefore we first model this waiting time.

Consider output port j. It has room for only the packet that is being evacuated. However, the HOL positions at the M input queues can contain a packet meant for output j which are waiting for the port to become free. These packets are served in FCFS manner. Since, at any given time, at most $M-1$ such packets may be waiting for a service completion at output port j, we can model the output port as a M/M/1/M queue. The approximation here is that the arrival process to the virtual queue is assumed to be Poisson which is not true when M is finite. The throughput of output port j is Λ_j. But the queue has finite buffers and the "arrival rate" corresponding to this throughput, let us call this the effective arrival rate Λ'_j, is obtained by solving for Λ'_j in the equation

$$\Lambda_j = \Lambda'_j \left[1 - \frac{1-\rho'_j}{1-(\rho'_j)^{M+1}}(\rho'_j)^M\right] = \frac{1-(\rho'_j)^M}{1-(\rho'_j)^{M+1}}\Lambda'_j \quad (2)$$

where $\rho'_j = \frac{\Lambda'_j}{\mu_j}$. The term in the square brackets in the first equality corresponds to the probability that an arriving packet into an M/M/1/M queue is not blocked.

Once a packet moves to the head of the input queue, the time it spends there has two components; the blocking delay, i.e., the time until the output starts evacuating it and the time taken for evacuation. The service time at the output is considered as part of the time spent at the head of the input queue because until the packet is evacuated by the output, subsequent packets do not move to the head of the queue. Thus the sum of the blocking delay and the evacuation time at the head of the input queue corresponds to the service time in that queue. (See Figure 1.) Thus the input queue is seen to be a M/G/1 queue with service time distribution given by the distribution of the sum of the blocking delay and the evacuation time. The waiting time for a packet in the input queue depends on this service distribution. If we obtain the first and second moments of this service distribution, we can use the Pollaczek-Khinchin formula to obtain the delay in the input queue. The total delay for a packet will be the sum of this delay, the blocking delay and the transmission delay.

Let us now obtain the blocking delay experienced by a packet whose destination is j and is at the head of input queue i. Let $\theta_j(k)$ be the probability that there are k packets that are either in service at output port j or waiting at the HOL of the input queues. Since the "queue" corresponding to the output port j has been approximated by a M/M/1/M queue,

$$\theta_j(k) = \frac{(1-\rho'_j)(\rho'_j)^k}{1-(\rho'_j)^{M+1}} \qquad \text{for } k = 0 \cdots M \quad (3)$$

$\pi_j(k)$, the probability that a packet moving to the head of an input queue and wanting to go to output j sees k packets ahead of it is

$$\pi_j(k) = \frac{\theta_j(k)}{1-\theta_j(M)} = \frac{(1-\rho'_j)(\rho'_j)^k}{1-(\rho'_j)^M} \quad \text{for } 0 \le k \le M-1 \quad (4)$$

$\pi_{ij}(k)$, the probability that a packet moving to the head of input queue i, and wanting to go to output j sees k packets ahead of it, is given by

$$\pi_{ij}(k) = p_{ij}\pi_j(k) \qquad \text{for } k = 0 \cdots M-1 \quad (5)$$

When a packet enters the "queue" at output port j and sees k packets ahead of it, it has to wait for the evacuation of these packets before it can begin its service and its waiting time is a k stage Erlangian distribution (sum of the k

independent, exponentially distributed evacuation times). Thus the blocking delay has a phase type distribution like that shown in Figure 2. The first and second moments of the blocking delay at input queue i of the packets meant for output port j, $\overline{B_{ij}}$ and $\overline{B_{ij}^2}$ respectively, are

$$\overline{B_{ij}} = \sum_{k=1}^{M-1} \pi_{ij}(k)\frac{k}{\mu_j} \qquad \overline{B_{ij}^2} = \sum_{k=1}^{M-1} \pi_{ij}(k)\frac{k(k+1)}{\mu_j^2} \quad (6)$$

The first and second moments of the blocking delay at input i are

$$\overline{B_i} = \sum_{j=1}^{N} \overline{B_{ij}} \qquad \overline{B_i^2} = \sum_{j=1}^{N} \overline{B_{ij}^2} \quad (7)$$

The time spent at the head of the input queue is the sum of two independent random variables - the blocking delay with a phase distribution described earlier and the evacuation time with an exponential distribution. Thus the first and second moments of the service time for a packet at the head of input queue i, $\overline{X_i}$ and $\overline{X_i^2}$ respectively, are

$$\begin{aligned} \overline{X_i} &= \overline{B_i} + \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} \\ \overline{X_i^2} &= \overline{B_i^2} + 2\overline{B_i}\sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} + 2\sum_{j=1}^{N} \frac{p_{ij}}{\mu_j^2} \end{aligned} \quad (8)$$

The sojourn time in the switch for an input packet at port i, D_i, is (from the Pollaczek-Khinchin formula)

$$D_i = \frac{\lambda_i \overline{X_i^2}}{2(1-\lambda_i\overline{X_i})} + \overline{B_i} + \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} \quad (9)$$

The maximum arrival rate that input port i can support is obtained by solving for λ_i in $\lambda_i\overline{X_i} = 1.0$.

Let us now consider the special case of an $N \times N$ switch with $p_{ij} = \frac{1}{N}$ for all i, j; $\lambda_i = \lambda$ for all i and $\mu_j = 1.0$ for all j. The throughput from each output port is λ. The effective arrival rate corresponding to this throughput, Λ', is obtained by solving

$$\lambda = \frac{1-(\Lambda')^N}{1-(\Lambda')^{N+1}}\Lambda'$$

After simplification, the following will be the expressions for $\pi_{ij}(k)$, $\overline{B_i}$, $\overline{B_i^2}$, $\overline{X_i}$, $\overline{X_i^2}$ and D_i

$$\pi_{ij}(k) \equiv \pi(k) = \frac{(1-\Lambda')(\Lambda')^k}{1-(\Lambda')^N}$$

$$\begin{aligned} \overline{B_i} \equiv \overline{B} &= \frac{\Lambda\left(1-N(\Lambda')^{N-1}+(N-1)(\Lambda')^N\right)}{(1-(\Lambda')^N)(1-\Lambda')} \\ \overline{B_i^2} \equiv \overline{B^2} &= \sum_{k=1}^{N-1} k(k+1)\left[\frac{1}{\lambda}\frac{(1-\Lambda')(\Lambda')^{k+1}}{1-(\Lambda')^{N+1}}\right] \end{aligned}$$

$$\overline{X_i} \equiv \overline{X} = \overline{B} + 1.0 \qquad \overline{X_i^2} \equiv \overline{X_i} = \overline{B^2} + 2(\overline{B}+1)$$

N	Continuous Time (Approx)	Continuous Time (Exact)	Discrete Time
1	1.000000	1.00	1.0000
2	0.666667	0.667	0.7500
3	0.576849	0.600	0.6825
4	0.540642	0.571	0.6553
5	0.521880	0.556	0.6399
6	0.512015	0.545	0.6309
7	0.506655	0.538	0.6234
8	0.503694	0.533	0.6184
∞	0.500000	0.500	0.5858

TABLE I

MAXIMUM ACHIEVABLE THROUGHPUT FOR VARIOUS VALUES OF N.

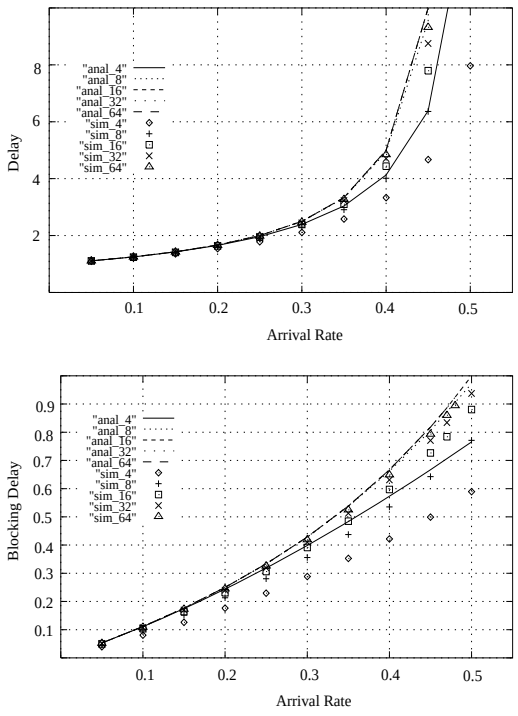


Fig. 3. Delay-throughput and blocking delay characteristics from analytical and simulation models for different values of N.

$$D_i \equiv D = \frac{\lambda\overline{X^2}}{2(1-\lambda\overline{X})} + \overline{B} + 1.0$$

Maximum throughput is obtained by solving for λ in $\lambda\overline{X} = 1.0$. Table I, shows the maximum throughput of the continuous time input queued switch as obtained from our analytical model and the formula of Fuhrman [3], $M/(2M-1)$. Throughput of the discrete time switch [5] is also shown . The maximum error between ours and Fuhrman's model is about 6%.

Figure 3 shows the total delay and the blocking delay for various values of N from analytical and simulation models as a function of λ. Note that the difference between the analytical and simulation models improves for both the total and the blocking delay as the switch size increases.

As $N \to \infty$, the M/M/1/N queue corresponding to each

output port becomes an M/M/1 queue with arrival rate λ and service rate 1.0. We now show that our delay model is exact for $N \to \infty$. The arrival process to the input queue is Poisson. Since each packet chooses a destination independent of the packets before it, its blocking delay at the head of the input queue is independent of the blocking delay of the others before it. Thus the "service times" are independent and the input queue is an M/G/1 queue. The arrival process at the server of this queue of packets destined for a tagged output port corresponds to one stream of arrivals into the tagged port. It is easy to see that the interarrival times are i.i.d. with a general distribution with arrival rate λ/N. The total arrival process into the tagged port is a the superposition of N such streams. As $N \to \infty$, this is Poisson with rate λ [4]. Thus, as $N \to \infty$ the blocking delay distribution is equal to the distribution of the total delay in an M/M/1 queue and hence this is the service time distribution for the input M/G/1 queue. Thus the input queue can be characterized exactly as an M/G/1 queue in which the service time depends on the arrival rate and its first and second moments are given by the first and second moments of the average time spent in an M/M/1 queue. Thus for the input queue to be stable, λ should be less than the reciprocal of the average time spent in an M/M/1 queue with arrival rate λ and service rate 1.0. This yields the condition, $\lambda \leq 1 - \lambda$ or $\lambda < 0.5$ for stable queues at the input.

III. Evacuating Multiple Packets to an Output

Throughput can be increased by simultaneously evacuating multiple packets by the output ports. Thus, for a speedup factor of m, each output port may be treated as an m server system. If there are more than m HOL packets at the inputs destined for the same output, m of them are served on a FCFS basis and others are blocked. Note that this speedup is similar to that of [7] rather than that of [2] in that multiple input queue HOLs are served simultaneously rather than the outputs evacuating faster.

Consider output port j of an $M \times N$ input queued switch with a speedup factor of m. It has room for m packets corresponding to the m servers. It easily seen that output port j can be modelled as a M/M/m/M queue. As in Section II, the approximation involved is that the arrival process is assumed to be Poisson, which is not true when M is finite. The effective arrival rate to this finite buffer queue, Λ'_j, is obtained by solving for Λ'_j in the equation

$$\Lambda_j = \Lambda'_j \left[1 - \frac{\left[\frac{(\rho'_j)^M m^m}{m!} \right]}{\left[\sum_{k=0}^{m-1} \frac{(m\rho'_j)^k}{k!} + \sum_{k=m}^{M} \frac{(\rho'_j)^k m^m}{m!} \right]} \right] \quad (10)$$

where $\rho'_j = \frac{\Lambda'_j}{m\mu_j}$. Proceeding as in Section II, $\theta_j(k)$ and $\pi_j(k)$ are now given by

$$\theta_j(k) = \begin{cases} P_0 \frac{(m\rho'_j)^k}{k!} & k \leq m \\ P_0 \frac{(\rho'_j)^k m^m}{m!} & m \leq k \leq M \end{cases}$$

N	Speedup Factor					
	2		3		4	
	Contin	Discr	Contin	Disc	Contin	Disc
4	0.8670	-	0.9795	-	1.0000	-
8	0.8304	-	0.9616	-	0.9934	-
16	0.8284	-	0.9611	-	0.9934	-
32	0.8284	-	0.9611	-	0.9934	-
∞	0.8284	0.8845	0.9611	0.9755	0.9934	0.9956

TABLE II

Maximum throughput for various speedup factors. Results for the discrete time switch are from [9]

$$\pi_j(k) = \begin{cases} \frac{P_0}{1-\theta_j(M)} \frac{(m\rho'_j)^k}{k!} & k \leq m-1 \\ \frac{P_0}{1-\theta_j(M)} \frac{(\rho'_j)^k m^m}{m!} & m-1 \leq k \leq M-1 \end{cases}$$

where P_0, the probability that the M/M/m/M queue is empty, is given by

$$P_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho'_j)^k}{k!} + \sum_{k=m}^{M} \frac{(\rho'_j)^k m^m}{m!} \right]^{-1} \quad (11)$$

Proceeding along the lines of the analysis in Section II, $\pi_{ij}(k)$ is obtained using the expression in Eqn 5, $\overline{B_{ij}}$ and $\overline{B_{ij}^2}$ are

$$\overline{B_{ij}} = \sum_{k=m}^{M-1} \pi_{ij}(k) \frac{k}{m\mu_j} \qquad \overline{B_{ij}^2} = \sum_{k=m}^{M-1} \pi_{ij}(k) \frac{k(k+1)}{(m\mu_j)^2}$$

and $\overline{B_i}$ and $\overline{B_i^2}$ are obtained from Eqn 7. As before, each input queue can be modeled as a M/G/1 queue whose service time is given by the sum of the blocking delay and the evacuation time by the output port. The first and second moments of the service time for a packet at the head of input port i, $\overline{X_i}$ and $\overline{X_i^2}$ are obtained using Eqn 8 and the average delay for an input packet is obtained using Eqn 9 and the maximum arrival rate that input port i can support is obtained by solving for λ_i in $\lambda_i \overline{X_i} = 1.0$.

In Table II we show the maximum throughput with various speedup factors for different switch sizes. For comparison, we have also shown the maximum throughput for a discrete time switch with $N = \infty$ as obtained in [7]. Just as in the case of the discrete time switch, a speedup of 4 is needed to achieve 99% throughput. It may be mentioned here that doubling the evacuation rate can double the throughput. In Figure 4 we plot the throughput delay characteristics for various values of N from analytical and simulation models for speedup factors of 2 and 4. It may be observed that for a speedup of 4, the delay curve resembles that of the M/M/1.

IV. Hotspot Analysis

For this analysis we consider the case where the incoming packets select a particular output with higher probability than the others. Let packet arrivals at input i form a

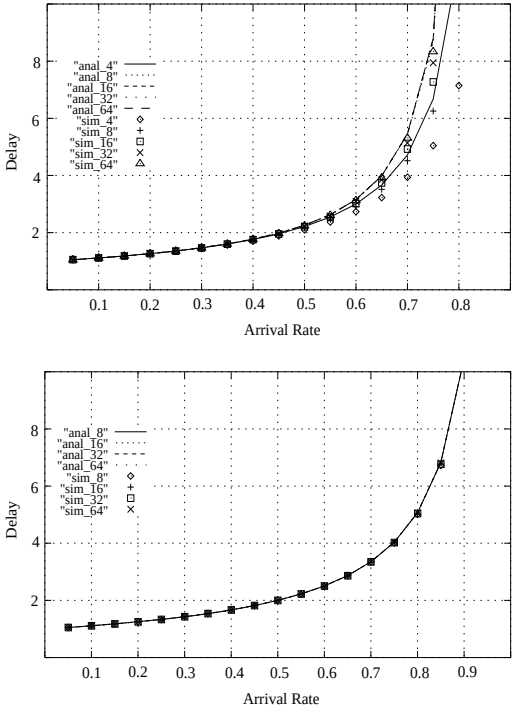
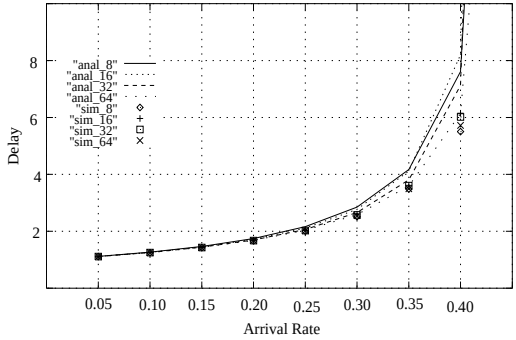


Fig. 4. Delay-throughput characteristics of input queued switches in continuous time with speedup. Results from analytical and simulation models are shown for different values of N with speedup factors of 2 and 4.



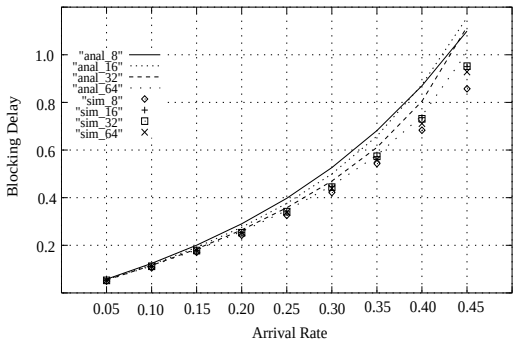


Fig. 5. Delay-throughput characteristics and the blocking delay of input queued switches in continuous time with hotspots. Results from analytical and simulation models are shown for $\gamma = 2$ for different values of N.

Poisson process with rate λ_i and choose destination j with probability p_{ij}. A hotspot at h is defined as follows

$$p_{ij} = \begin{cases} \beta & \text{for } j \neq h \\ \gamma\beta & \text{for } j = h, \gamma > 1 \end{cases} \quad \text{and} \quad \sum_{j=1}^{N} p_{ij} = 1 \quad \text{for all } i$$

As γ increases, contention for the hotspot output port h increases and hence the blocking delay for the HOL packets in the inputs increases. The increased blocking delay increases the "input service time" and hence the total delay of all the packets. The delay and the maximum throughput for this case can be obtained using the methods used in Sections II and will not be described here. In Figure 5 we show the total delay and the blocking delay for various values of N from analytical and simulation models for $\gamma = 2$ as a function of λ. Maximum throughput saturates at about 0.4 as compared to 0.5 without hotspots.

V. Conclusion

In this paper, we have presented a generalized analytical model for an input queued switch operating in continuous time. Although we have presented the analysis for switches with infinite input buffers our model can be easily extended to analyse finite buffer switches.

From our analysis we can explain why the FCFS service of the virtual queue of the output gives the least average total delay for an input packet. The total delay in the virtual queue is the service time of the input queue and FCFS has the minimum variance among all the service disciplines. Thus the variance of the "service time" of the input queue, which is an M/G/1 queue, is minimum for FCFS service of the virtual output queue. Since the mean delay in an M/G/1 queue depends on the mean and varaince of the service time, this gives the minimum mean total delay. Also, from our models it is clear that the conjecture that the performance of an $M \times N$ switch is symmetric in M and N is obviously true for the maximum achievable throughput but not for the delay analysis.

References

[1] D. P. Bhandarker, "Analysis of memory interference in multiprocessors" *IEEE Trans on Comput.,* vol C-24, pp 897-908, Sept 1975.

[2] A. S. Diwan, Performance Analysis of Speeded-up High Speed Packet Switches, *M.Sc. Thesis* Deptt of Elecl Communication Engg, Indian Inst of Science, Bangalore India, 1998.

[3] S. W. Fuhrman, "Performance of a Packet Switch with a Crossbar Architecture," *IEEE Trans on Commun.,* vol COM-41, pp 486–491, Mar 1993.

[4] J. F. Hui, "Switching and Traffic Theory for Integrated Broadband Networks," *Kluwer Academic Publishers,* Boston, 1990.

[5] M. J. Karol, M. G. Hluchyj and S. P. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Trans on Commun,* vol. COM-35, no. 12, pp. 1347-1356, Dec 1987.

[6] P. Newman, W. L. Edwards, R. Hinden, E. Hoffman, F. C. Liaw, T. Lyon and G. Minshall, "Ipsilon Flow Management Protocol Specification for IPv4," *IETF RFC 1953,* May 1996.

[7] Y. Oie, M. Murata, K. Kubota and H. Miyahara, "Effect of Speedup in Nonblocking Packet Switch," *Proceedings of IEEE ICC, 1989,* pp 410-414.

[8] J. H. Patel, "Performance of processor-memory interconnections for multiprocessors," *IEEE Trans on Comput.,* vol C-30, pp 771-780, Oct 1981.

[9] H. Perros, "Queueing Networks with Blocking," *Oxford University Press,* 1994.