Variable Length Packet Switches: Delay Analysis of Crossbar Switches under Poisson and Self Similar Traffic

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 crossbar switches for switching variable length packets. Analysis of such switches is important in the context of IP switches where the packet interarrival times and packet lengths are drawn from continuous distributions. Assuming a single stage $M \times N$ switch we obtain a very general throughput delay model for Poisson packet arrivals and exponential service times. We then analyze an $M \times N$ switch for self similar packet arrivals and exponential packet lengths. An MMPP-based self similar arrival process model corresponding to the arrival rate, the autocorrelation, the Hurst parameter and the time scales over which burstiness exists in the input process is first obtained using results from [1]. We then use queing theory available for MMPP/G/1 queues to model the switch performance for self similar packet arrivals. The results from the analytical model are compared against those from a simulation model that is driven by traces that are statistically similar to the Bellcore traces. We also analyse the effect of link multiplicities (speedup) to the output and asymmetries in the input traffic.**

Keywords—**Variable Length Packet Switches, IP Switching, Self Similar Traffic.**

I. Introduction

We analyse input-queued, space-division, variable-length-packet switches. The packet lengths and interarrival times are assumed to be random and drawn from continuous distributions. With the emergence of IP switching technologies [15], such a continuous time analysis of space division packet switches becomes relevant and important. For fixed packet lengths with the switch operation synchronized into slots (typically of the size of the packet lengths), discrete time analyses under various assumptions are available. Patel [18] analyzes such switches with no input or output buffers under Bernoulli arrivals. In this model, an output contention is resolved by dropping all but one of the contending packets. The maximum throughput under this model is $1 - 1/e \approx 0.63$. Karol et al [8] show that the saturation throughput of an input queued switch with infinite input buffers is $2 - \sqrt{2} \approx 0.586$. Li [11] shows that when the arrivals are correlated, the maximum throughput of the in these switches comes down to 0.5. Fuhrman [7] presents a continuous time analysis of an input queued packet switch (a crossbar) by considering a $M \times N$ crossbar switch with variable length packets as inputs. Assuming *iid* Poisson arrival processes at each node and uniform routing probabilities he shows that $M/(M + N + 1)$ is the saturation throughput per port of such a switch. He also presents a delay analysis by first developing a state dependent server model to obtain the service rate when there are i packets in the system and then using these rates in an M/M/1 queue model for the switch. The first part of our paper can be considered to be a generalization of the results of [7] where we present a throughput delay analysis for an $M \times N$ input queued switch with arbitrary Poisson arrivals at each input, exponential packet lengths, arbitrary output line rates and arbitrary routing probabilities.

The very little literature that there is on the analysis of variable length packet switches are for Poisson packet arrival processes. Recent measurement studies over a wide range of packet networks have established the self-similar nature of packet traffic and the failure of the traditional Poisson models to capture the *long range dependence* (LRD) and the burstiness of such packet arrival processes. The long range dependence in the arrival process is marked by the presence of correlations and burstiness over many time scales which are known to have a considerable impact on the queuing performance. We now know that queuing behavior with LRD arrival processes has a marked variation from those with Poisson arrivals. Extreme burstiness of packet traffic spanning over a number of time scales give rise to extended periods of large queue build ups and also to sustained periods low activity. Thus if the arrival process feeding each port of an input queued switch is from a LRD process, their interaction with the HOL blocking in an input queued switch can lead to a very bad queuing behavior. In view of the extreme queuing behavior expected, a deeper understanding of the switch behavior becomes necessary because the switch is the critical component in providing various quality of service guarantees in the multiservice Internet of the future. In this paper we extend the delay models for Poisson traffic arrivals to LRD input processes and present some results from our investigations into the queuing behavior of input queued, variable length packet switches under such input.

The rest of the paper is organised as follows. Section II introduces the delay throughput analysis for a $M \times N$ switch with Poisson arrivals and exponentially distributed packet lengths. In section III we present the analysis technique for a $M \times N$ switch with the arrival stream at each input port characterised by a self-similar process. We also analyse the switch under link multiplicities and asymetries in the traffic conditions. Finally, Section IV presents a discussion on the results and concluding remarks.

II. $M \times N$ Switch with Poisson Arrivals, Exponential Packet Lengths

We first consider a single stage unslotted, internally nonblocking $M \times N$ input queued packet switch. Packet arrivals to input port i form a Poisson process of rate λ_i and choose a destination j with probability p_{ij}. The line rate on output port j is μ_j and there are no buffers at the output. Input packets are served according to FIFO. 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 can also be analyzed using the method developed here but we do not investigate them. From above, the arrival rate to output port j, Λ_j, and its utilization, η_j, are

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

The sojourn time of an input packet has two components - waiting time in the input queue till it moves to the head of the line (HOL) and the time spent at the HOL of the input queue till the HOL packets from other input queues that were blocked earlier finish their service and the packet is evacuated. The time spent at the HOL of the input queue corresponds to the "service time" in the input queue. This service time, once again, has two components - a blocking delay, the time until the output starts evacuating it, and the actual service time, the time taken to evacuate the packet by the destination port. Figure 1 shows these times in detail. Since the arrivals to the input queue are Poisson, each input queue can be seen to be a M/G/1 queue with service time distribution given by the time spent by a packet at its HOL. To analyse the queuing behavior the distribution of the time spent at the HOL of the queue needs to be obtained and this is derived below. In this derivation, we use techniques similar to the analysis of queueing networks with blocking [20].

Consider output port j. It has room for only the packet that is being evacuated (served). 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 form a virtual queue for output j and are served FCFS. Thus the virtual queue of any output has at most M buffers. The time taken by the output port to evacuate a packet from the HOL of the inputs is exponentially distributed with mean $1/\mu_j$. If we approximate the arrival process to the virtual queue by a Poisson process of throughput Λ_j, then output queue j can be modeled as a M/M/1/M queue. We can easily show that as $M \to \infty$, the arrival process to the output queue is indeed Poisson under certain conditions. Since the queue has finite buffers, the throughput is not equal to the arrival rate. The throughput of output port j should be Λ_j. Therefore the "arrival rate" corresponding to this throughput, let us call this the effective arrival rate Λ'_j, will be obtained by solving for Λ'_j in the equation

$$\Lambda_j = \Lambda'_j \left[1 - \frac{1-{\eta'}_j}{1-{\eta'}_j^{M+1}} {\eta'}_j^{M} \right] = \Lambda'_j \frac{1-{\eta'}_j^{M}}{1-{\eta'}_j^{M+1}} \tag{2}$$

where $\eta'_j = \Lambda'_j/\mu_j$. 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.

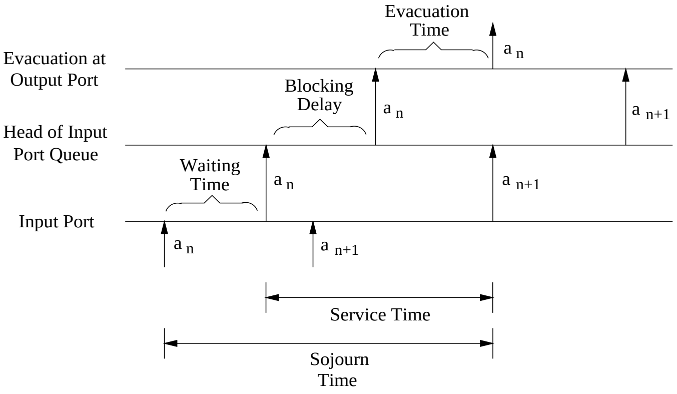

Fig. 1. Time diagram for the sojourn time in the switch. a_n represents the n^{th} arrival to the input queue and all times shown in this figure correspond to this packet.

The probability that there are k packets in the virtual queue of output port j, $\theta_j(k)$, is given by

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

Packet arrivals to the head of an input queue are approximated to form a Poisson process. Thus the probability that it will see k packets ahead of it in the virtual queue of the output will be $\theta_j(k)$. However a packet moving to the head of an input queue can see only $0, 1, \cdots M-1$ and will never see M packets ahead of it. Therefore the probability that a packet arriving to the head of an input queue wanting to go to output j sees k packets ahead of it, $\pi_j(k)$, will be

$$\pi_j(k) = \left[\frac{\theta_j(k)}{1-\theta_j(M)} \right] = \left[\frac{(1-\eta'_j)(\eta'_j)^k}{1-(\eta'_j)^{M}} \right] \quad \text{for } k = 0, 1, \cdots M-1 \tag{4}$$

In the virtual queue of output port j if there are k packets ahead of it, the packet 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). In addition to the blocking delay there is the evacuation time that has an exponential distribution of mean $1/\mu_j$. Thus the conditional (conditioned on the packet wanting to go to output port j) sojourn time of a packet at the HOL of the input queue has a phase type distribution like that shown in Figure 2. The Laplace-Stieltjes Transform (LST) of the unconditional distribution of the sojourn time at the head of input i, $\mathcal{X}_i(s)$, can be seen to be

$$\mathcal{X}_i(s) = \sum_{j=1}^{N} p_{ij} \left[\sum_{k=1}^{M-1} \pi_i(k) \left(\frac{\mu_j}{\mu_j + s} \right)^k \right] \left[\frac{\mu_j}{\mu_j + s} \right] \tag{5}$$

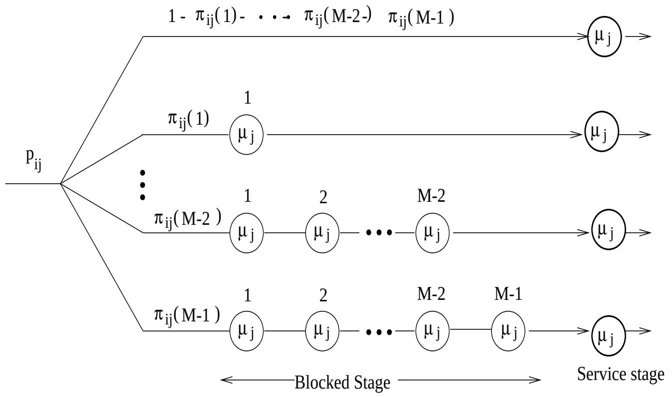

Fig. 2. Phase-type distribution for sojourn time in virtual queue of output j when a packet reaches the HOL of input i. The blocking delay is a k-stage Erlangian with probability $\pi_{ij}(k)$, the probability that there are k packets in the virtual queue of output port j ahead of this packet. There is an additional service stage corresponding to the evacuation of the packet from the input queue by the output port.

Here the term in the first square brackets corresponds to the blocking delay and that in the second corresponds to the evacuation time given that the packet wants to go to output j. The first three moments of the blocking delay at input queue i, $\overline{B_i}$, $\overline{B_i^2}$ and $\overline{B_i^3}$ respectively, are

$$\begin{aligned}\overline{B_i} &= \sum_{j=1}^{N} p_{ij} \sum_{k=1}^{M-1} \pi_j(k) \frac{k}{\mu_j} \\ \overline{B_i^2} &= \sum_{j=1}^{N} p_{ij} \sum_{k=1}^{M-1} \pi_j(k) \frac{k(k+1)}{\mu_j^2} \\ \overline{B_i^3} &= \sum_{j=1}^{N} p_{ij} \sum_{k=1}^{M-1} \pi_j(k) \frac{k(k+1)(k+2)}{\mu_j^3} \qquad (6)\end{aligned}$$

Likewise, the first three moments of the service time for the input queue, $\overline{X_i}$, $\overline{X_i^2}$, $\overline{X_i^3}$ respectively, are

$$\begin{aligned}\overline{X_i} &= \overline{B_i} + \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} \\ \overline{X_i^2} &= \overline{B_i^2} + 2\overline{B_i} \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} + 2 \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j^2} \\ \overline{X_i^3} &= \overline{B_i^3} + 3\overline{B_i^2} \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} + 6\overline{B_i} \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j^2} + 6 \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j^3} \qquad (7)\end{aligned}$$

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

$$D_i = \frac{\lambda_i \overline{X_i^2}}{2(1-\lambda_i \overline{X_i})} + \overline{B_i} + \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j} \qquad (8)$$

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

Consider the special case of an $N \times N$ switch with $p_{ij} = 1/N$ for all i, j; $\lambda_i = \lambda$ for all i and $\mu_j = 1.0$ for all j. 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. It is easily seen that our delay model is exact for $N \to \infty$. As $N \to \infty$, the virtual M/M/1/N queue of the outputs becomes an M/M/1 queue with arrival rate λ and service rate 1.0. As $N \to \infty$, the arrival process to the input queue is Poisson with rate λ and it in turn is an M/G/1 queue with service time equal to the sojourn time in an M/M/1 queue with arrival rate λ and service rate 1.0. Thus for the input queue to be stable, λ should be less than the reciprocal of the sojourn time of 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.

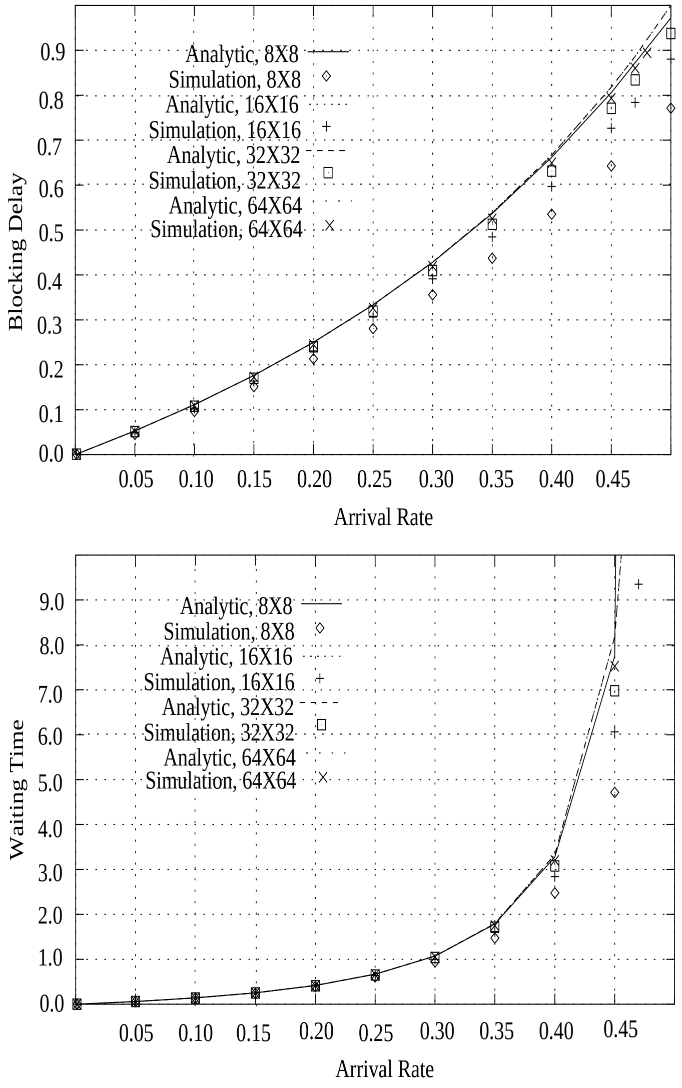

Fig. 3. Mean blocking total delays *vs* throughput. From analytical and simulation models for $N = 4, 8, 16, 32$ and 64.

III. $M \times N$ Switch with Self Similar Input and Exponential Packet Lengths

Having modeled switch behavior under the somewhat idealized model of Poisson inputs we will now examine the behavior under a more realistic model of self similar inputs. Before presenting the delay analyses for self similar arrival processes we give a brief overview of the various equivalent definitions of self similarity and the packet arrival models that can be used with each of these. Finally, we will select the self similar packet arrival model that has a well developed queueing theory.

Packet arrival instants are modeled as point processes. Divide the time axis into nonoverlaping intervals of unit length and let $\mathcal{X} = \{X_t : t = 0, 1, 2, \cdots\}$ be the number of points (packet arrivals) in the t^{th} interval. Measurements and analysis of such packet arrival processes in real networks has indicated that $\mathcal{X}$ is a self similar process. This means that although analysis of packet switches for the Poisson packet arrival model gives us a "first-order-feel" for their performance, to understand their performance in real networks, it is necessary to study their performance for self similar packet arrivals.

Mathematically, self similarity in the process $\mathcal{X}$ can be expressed in many ways. Let $\mathcal{X}$ be covariance stationary with mean λ, variance σ^2 and autocorrelation function $r(k), k \geq 0$. For each $m = 1, 2, \cdots$, let $X^{(m)} = (X_k^{(m)} : k = 1, 2, \cdots)$ be the new covariance stationary time series (with corresponding autocorrelation function $r^{(m)}$) obtained by averaging the original series X over non-overlapping blocks of size m, i.e., for each $m = 1, 2, \cdots$, $X^{(m)} = (X_{km-m+1} + \cdots + X_{km})/m$, $k \geq 1$. Then self-similarity of $\mathcal{X}$ means any of the following

$\mathcal{X}$ has a slowly decaying variance: The variance of the sample mean decreases more slowly than the reciprocal of the sample size. $var(X^{(m)}) \sim am^{-\beta}$ as $m \to \infty$ and $0 < \beta < 1$ (a is a finite positive constant)

$\mathcal{X}$ is long range dependent (LRD): The autocorrelations decay hyperbolically rather than exponentially fast, implying a non-summable autocorrelation function. $\sum_k r(k) = \infty$ and

$\mathcal{X}$ is $1/f$-noise: The spectral density $f(\cdot)$ obeys a power-law near the origin, i.e., $f(\lambda) \sim b\lambda^{-\gamma}$, as $\lambda \to 0$ with $0 < \gamma < 1$ and $\gamma = 1 - \beta$ (b is a finite positive constant)

Each of the above descriptions of a self similar process can lead to a class of models for the packet arrival process. From the point of understanding queuing behavior of systems, we consider those that are derived to match the LRD statistics of the packet arrival process. In [10] Leland *et al* show that Gaussian noise or nonlinear transformations on Gaussian noise such as fractional ARIMA can be used to characterise a LRD $\mathcal{X}$. In [19], Paxson and Floyd show that superposition of on/off sources that have a fixed rate in the on period and have a heavy-tailed distribution for the on and off period lengths can be used to model LRD $\mathcal{X}$. Erramilli, Singh and Pruthi use deterministic nonlinear chaotic maps to define a LRD $\mathcal{X}$ [5]. Andersen and Nielsen propose a Markovian approach in which an LRD $\mathcal{X}$ is obtained by superposing a number of two state Markov Modulated Poisson Processes (MMPPs) [1] with the resultant arrival process being an MMPP. The advantage of this last method is that in addition to allowing the modeling of burstiness over a number of time scales with the desired covariance structure, since the packet arrivals are MMPP, a well developed queuing theory is available for analysis. Further, it can be shown that this model converges to fractional Brownian motion in the sense of finite dimensional distributions as the number of MMPPs increases. Therefore, we will use this in the analysis of the variable length packet switches with input queuing.

We first summarise the technique outlined in [1] to fit an MMPP process to an LRD arrival process. Let the packet arrival process to input port i be a second order self similar process with with mean λ_i, correlation at lag 1 ρ_i, Hurst parameter H_i, and the number of time scales over which the burstiness is to be modeled, n_i. This will be modeled as the superposition of a number of two state Interrupted Poisson Processes (IPPs), typically four, and a Poisson process. The covariance function of this superposed process is fitted to that of the self-similar process that we are modeling over several time scales. For input port i we will superpose d_i IPPs and the jth two-state IPP is parameterized by its generator matrix Q_i^j and rate matrix R_i^j as follows

$$Q_i^j = \begin{bmatrix} -c_i^{1j} & c_i^{1j} \\ c_i^{2j} & -c_i^{2j} \end{bmatrix} \qquad R_i^j = \begin{bmatrix} r_i^j & 0 \\ 0 & 0 \end{bmatrix} \tag{9}$$

The superposed process will be

$$\mathbf{Q}_i = \bigoplus_{j=1}^{d_i} Q_i^j \qquad \mathbf{R}_i = \bigoplus_{j=1}^{d_i} R_i^j \tag{10}$$

where $\bigoplus$ denotes the Kronecker sum. Note that the individual Poisson process in the fitting procedure may also be represented as a special case of a MMPP and added in the Kronecker sum of Eqn 10 to obtain the complete MAP model of the arrival process. The steady state probability vector of the Markov chain, $\mathbf{\Phi}_i$, can be obtained by simultaneously solving the following equations,

$$\mathbf{\Phi}_i \mathbf{Q}_i = 0 \qquad \mathbf{\Phi}_i \mathbf{e}_i = 1 \tag{11}$$

where $\mathbf{e}_i = [1, 1, \cdots, 1]^T$ is a unit column vector of length 2^{d_i}. Let $\mathbf{r}_i = [r_i^1, r_i^2, \cdots, r_i^{d_i}]$. Then the average arrival rate to input i, $\lambda_i = \mathbf{\Phi}_i \mathbf{r}_i^T$. The procedure to fit c_i^{1j}, c_i^{2j} and r_i^j to λ_i, ρ_i, H_i and n_i are described in [1].

As in the previous section we assume that each packet at input i chooses output j independent of other packets with probability p_{ij} and the the rate at which a packet is evacuated from an input queue by output port j is μ_j which is the line rate at output port j. Packets lengths are exponentially distributed with unit mean. There are infinite buffers at the input and none at the output. The output ports evacuate packets from the HOL of the input queues according to "first blocked first served" discipline. The "service time" of the input queue, time spent at the HOL by packet, is obtained exactly as before by making the approximation that the virtual queue to each output is an M/M/1/M queue. The sojourn time in this M/M/1/M queue is thus the service time for the input queue which we can now model as an MMPP/G/1 queue. Since the service time for the input queue is like before, the maximum throughput per port will be 0.5 and is derived exactly as before. Thus the moments of the service times are obtained exactly like in the previous section using Eqns 1-7. The first and second moments of the packet delays in the input queue can now be obtained using well known techniques for MMPP/G/1 queues [6]. The procedure is summarised in the appendix.

Numerical results are obtained as follows. We use the Bellcore traces [10] and derive their statistical properties in terms of the Hurst parameter, the correlation at lag 1 and the time scales over which the burstiness occurs. These parameters and the arrival rate λ are used to fit the parameters c_i^{1j}, c_i^{2j} and r_i for $j = 1, \cdots, 4$ of the MMPP model described in [1]. The analytical results are obtained for the MMPP/G/1 queue as described earlier. To validate the analytical results we also develop a simulation model in which the arrivals are MMPP with parameters derived above. The arrival process generator is validated by simulating a single server queue and comparing with the results given in [4]. The magnitudes of our delays and the knee region of the delay-throughput graph match that given in Figure 2 of [4]. In the simulation model a separate and independent MMPP arrival process generator is used for each of the input ports with the traces generated by each of the sources having identical statistical properties. Thus, statistically identical self-similar traces but with different sample paths are used as the input processes to the simulation model. In this paper we primarily use the Bellcore traces `pAug.TL` ($H = 0.82$ and $\rho = 0.582$) and `pOct.TL` ($H = 0.92$ and $\rho = 0.356$). We model burstiness over 4 time scales.

We mention here that we considered feeding the traces to obtain the simulation results. Since the number of inputs was large, the size of the traces was insufficient. The same trace cannot be fed to all the inputs because in that case the arrivals at each input will have a correlation of one, an obviously wrong choice for an arrival process. Also, we did not use shuffled versions of a single trace because shuffling of the time series of the traces would lead to a loss of the correlation structure and consequently the long range dependence.

In Figures 4–7 we show the first and second moments of total and blocking delays in the switch. It can be seen that the simulation and analytical results are in extremely good agreement except at loads close to the capacity of the switch. We see a marked difference in the shape of the delay characteristics for the `pOct.TL` trace at low loads which can be attributed to its comparatively low correlation value at lag one. At low loads, the low correlation suggests a lower probability of successive intervals having packet arrivals, which in turn leads to low delays. Further investigation of the effect of the correlation structure is done in Section III-B. As discussed earlier, the throughput delay curves in Figure 4 show that the switch saturates at a load of 0.5. Also, note that the first and second moments of the blocking delay shown in Figures 6 and 7 are identical for both the traces for a given switch size. This is because the virtual queue at each output port is modeled as an M/M/1/M queue whose delay characteristics depend only on the average arrival rate of the input processes and not on any of their other statistical properties.

From Figure 4 we see that the mean delay increases exponentially as the arrival rate. The delay performance can be divided into three regions - low $(0.0 - 0.10)$, medium $(0.10 - 0.40)$ and high $(0.40 - 0.50)$ loads. Note that in the medium load load region the mean delay is of the order of the order of 10^3. In all these regions the mean delay increases exponentially with increasing arrival rate. For comparison, we have shown the delays that would have been experienced in a single server queue without HOL blocking. This would be the delay experienced in an output queued switch in which the arrival rate to an output port would be described by the corresponding MMPP process. This shows that for a given arrival rate mean delay in the input queued switch could be at least double and nearly 10 times higher even at medium load.

The moments of the blocking delay for the case of Poisson arrivals and that of the MMPP arrivals is identical in the analytical models. Comparisons with the simulation model suggests that the analytical models are a good approximation. Hence we note that the effect of increase in the second moment in the case of self similar arrivals is significantly larger.

We have performed extensive analysis and simulations to understand the switch behavior under self similar arrivals and we have observed that when the burstiness extends over 3 time scales, the delays are of the order of 10^2.

From the above results we note that the analytical results match the simulations reasonably well. Therefore, in the following we do not present any simulation results.

A. Evacuating Multiple Packets in Parallel to an Output

To increase the throughput and reduce the delay through the switch, we could introduce parallelism by increasing the link multiplicity to an output port similar to the discrete time switch described by Oie et al in [17]. Note that this will require queuing at the output too.

It is easy to see that in this case if there are more than m HOL packets at the inputs destined for a particular output port, m of them are served simultaneously while the others are blocked. Here too we assume the input process to the queue to be Poisson which is an approximation when M is finite. Thus the virtual queue of each output port will be modeled as an M/M/m/M queue and the effective arrival rate to output port j corresponding to a throughput of Λ_j is obtained by solving for Λ'_j in

$$\Lambda_j = \Lambda'_j \left[1 - \frac{\left[\frac{(\eta'_j)^M m^m}{m!} \right]}{\left[\sum_{k=0}^{m-1} \frac{(m\eta'_j)^k}{k!} + \sum_{k=m}^{M} \frac{(\eta'_j)^k m^m}{m!} \right]} \right] \tag{12}$$

where $\eta'_j = \frac{\Lambda'_j}{m\mu_j}$. $\theta_j(k)$, $\pi_j(k)$ and $\mathcal{X}_i(s)$ are obtained like before by considering an M/M/m/M queue rather than an M/M/1/M queue at the outputs. Similarly the blocking and total delay moments are also obtained like before and are given by,

$$\overline{B_i} = \sum_{j=1}^{N} p_{ij} \sum_{k=m}^{M-1} \pi_j(k) \frac{k - m + 1}{\mu_j}$$

$$\overline{B_i^2} = \sum_{j=1}^{N} p_{ij} \sum_{k=m}^{M-1} \pi_j(k) \frac{(k - m + 1)(k - m + 2)}{\mu_j^2}$$

$$\overline{B_i^3} = \sum_{j=1}^{N} p_{ij} \sum_{k=m}^{M-1} \pi_j(k) \frac{(k - m + 1)(k - m + 2)(k - m + 3)}{\mu_j^3} \tag{13}$$

$$\overline{X_i} = \overline{B_i} + \sum_{j=1}^{N} \frac{p_{ij}}{\mu_j}$$

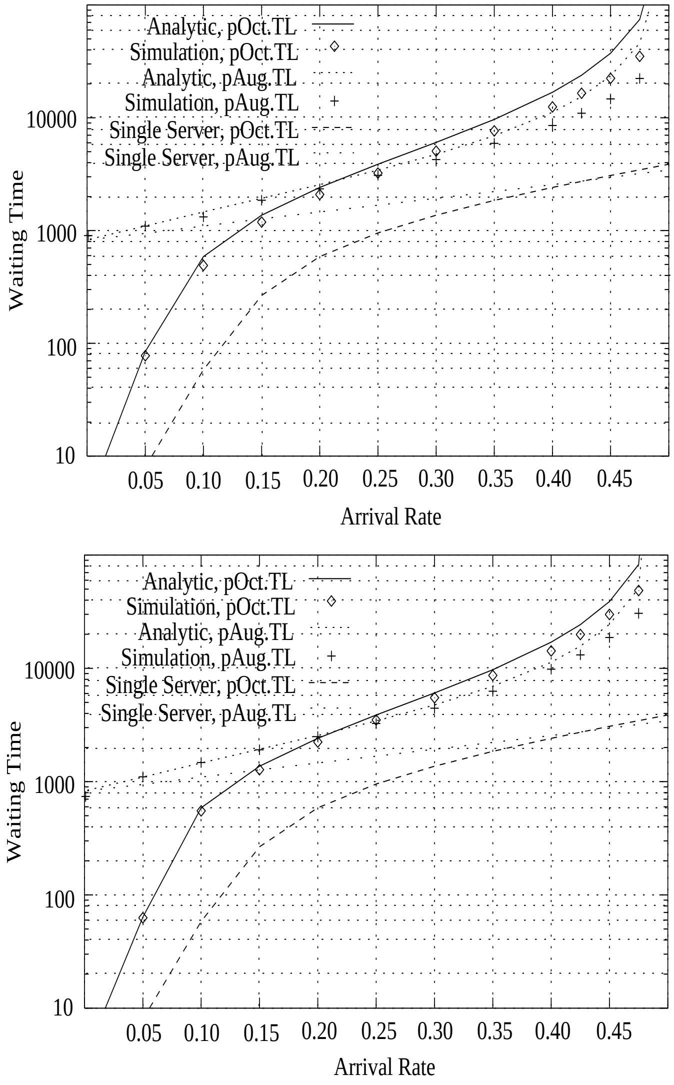

Fig. 4. First moment of total delay *vs* throughput. Results are from analytical and simulation models for Bellcore traces `pAug.TL` and `pOct.TL`. Top graph shows results for an 8×8 switch and bottom graph for a 16×16 switch.

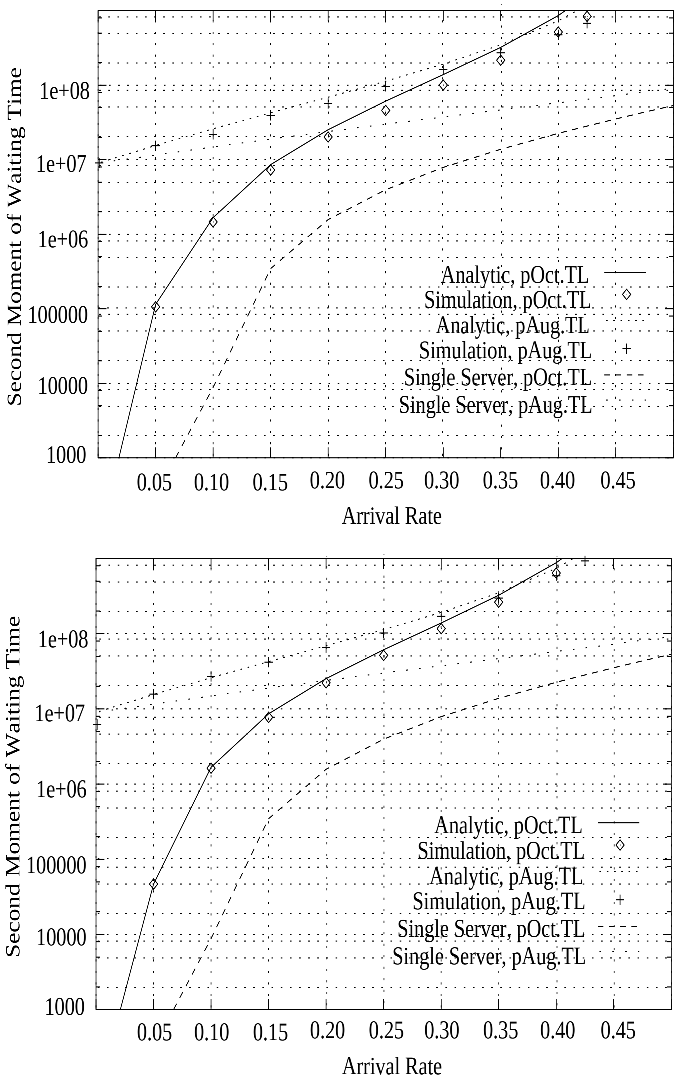

Fig. 5. Second moment of total delay *vs* throughput. From analytical and simulation models for Bellcore traces `pAug.TL` and `pOct.TL`. Top graph shows results for an 8×8 switch and bottom graph for a 16×16 switch.

$$\overline{X_i^2} = \overline{B_i^2} + 2\overline{B_i}\sum_{j=1}^{N}\frac{p_{ij}}{\mu_j} + 2\sum_{j=1}^{N}\frac{p_{ij}}{\mu_j^2}$$

$$\overline{X_i^3} = \overline{B_i^3} + 3\overline{B_i^2}\sum_{j=1}^{N}\frac{p_{ij}}{\mu_j} + 6\overline{B_i}\sum_{j=1}^{N}\frac{p_{ij}}{\mu_j^2} + 6\sum_{j=1}^{N}\frac{p_{ij}}{\mu_j^3} \qquad (14)$$

Note that the summations over the index k for the blocking delay is from $k = m$ to $k = M - 1$ because only when there are m or more packets waiting in the virtual queue will the packet at the HOL of an input queue have to wait. We can now use the expressions for the average delay and its second moment as given in Eqns 16 to obtain the latency for the arriving packets.

Figure 8 shows the analytical results for the delay throughput characteristics for $N \times N$ switches with $N = 8, 16, 32$ and 64 for speedup factors of 2 and 4. We assume identical loads on all the inputs and uniform routing probabilities p_{ij}. We see that effect of the switch size on the delay characteristics becomes negligible as the switch size increases. Also, the medium load region can be extended till the arrival rate of 0.75 for a speedup factor of 2 and upto 0.85 for a speedup factor of 4. Further, the mean delay is considerably lower with speedup than without. Also, the steep rise in the mean delay in the low load region does not manifest in the speeded up switch.

The maximum throughputs for a given speedup factor is obtained by solving for λ in $\lambda\overline{X} = 1.0$, where $\overline{X}$ is obtained from Eqn 14. Table I shows the maximum achievable throughputs for switches of various sizes and for speedup factors of 2, 3 and 4. Note that a switch with a speedup factor of 4 can support loads in excess of 99%.

B. *Effect of Asymmetries in Traffic*

Recall that the parameters in characterizing the input process are H the Hurst parameter, ρ the correlation a lag 1 and n the number of time scales over which burstiness occurs. In addition there are the routing probabilities and p_{ij} that can generate hotspots on some outputs. In this section we examine the ef-

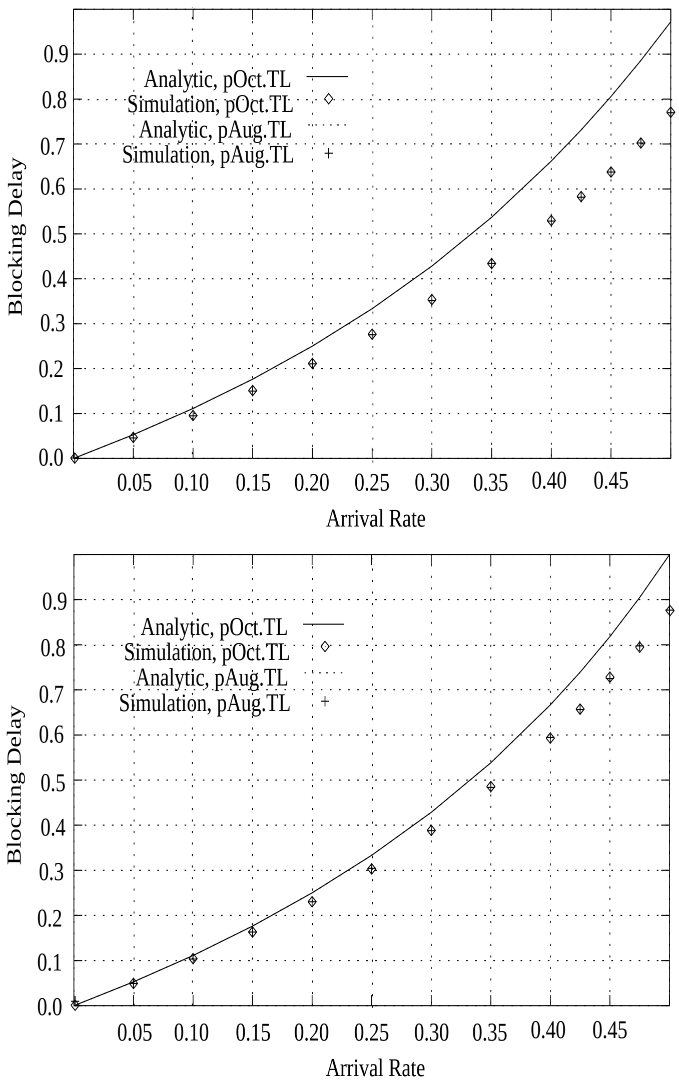

Fig. 6. First moment of blocking delay *vs* throughput. Results from analytical and simulation models for Bellcore traces `pAug.TL` and `pOct.TL`. Top graph shows results for an 8×8 switch and bottom graph for a 16×16 switch.

	Speedup Factor		
N	2	3	4
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	0.9611	0.9934

TABLE I

MAXIMUM THROUGHPUT FOR VARIOUS SPEEDUP FACTORS.

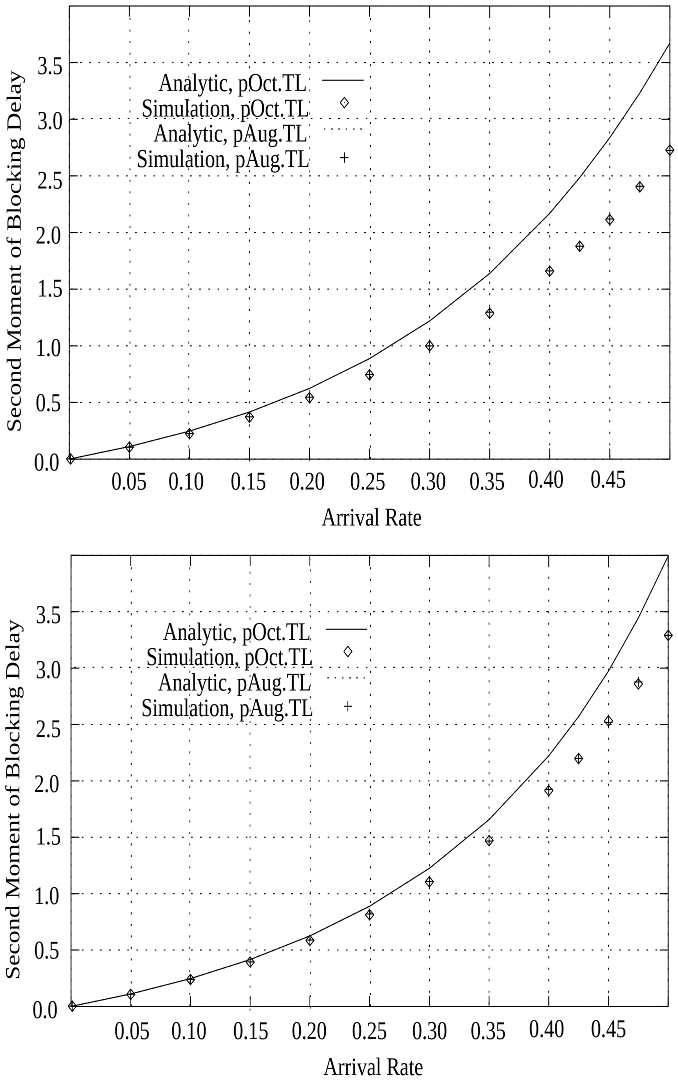

Fig. 7. Second moment of blocking delay *vs* throughput. From analytical and simulation models for Bellcore traces `pAug.TL` and `pOct.TL`. Top graph shows results for an 8×8 switch and the bottom graph for a 16×16 switch.

fect of asymmetries in these parameters across the inputs on the throughput delay characteristics for the input queued switch.

First, consider the effect of a hotspot on output port $h, 1 \leq h \leq N$ with

$$\begin{aligned} p_{ij} &= \begin{cases} \beta & \text{for } j \neq h \\ \gamma\beta & \text{for } j = h, \gamma > 1 \end{cases} \\ \sum_{j=1}^{N} p_{ij} &= 1 \qquad \text{for all } i \end{aligned} \qquad (15)$$

As γ increases, the contention for the hotspot output port h increases and hence the blocking delay for these packets at the head of their input queues increases. The increased blocking delay increases the "input service time" and hence the total delay of all the packets. In Figure 9 we show the effect of this hotspot for $\gamma = 2$. As is evident from the figures, there is a marked rise in the average delays in the presence of hotspots and a considerable reduction in the maximum achievable throughput.

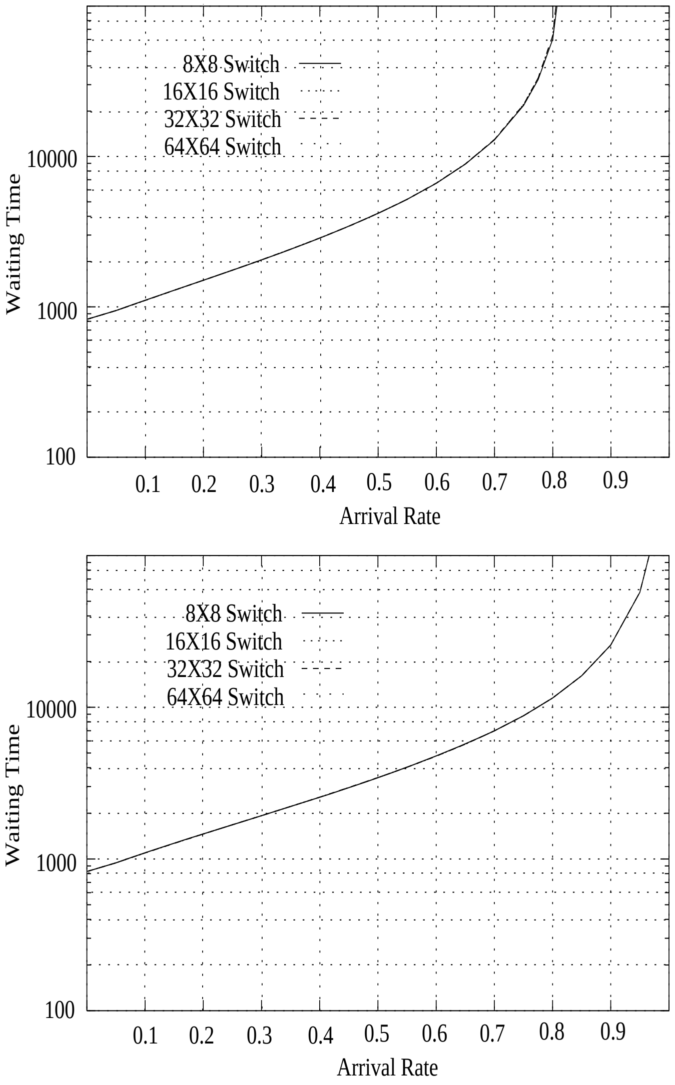

Fig. 8. First moment of total delay *vs* throughput. From analytical model for Bellcore trace `pAug.TL` for speedup factors of 2 and 4. Results are shown for $N \times N$ switches with $N = 8, 16, 32$ and 64.

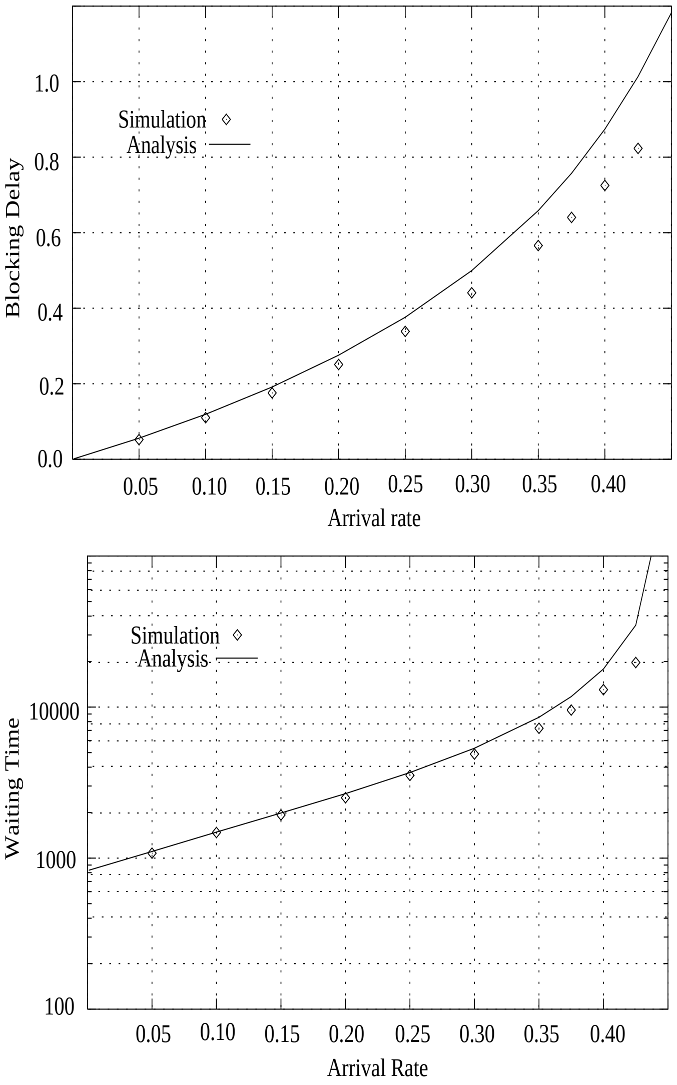

Fig. 9. Effect of an output hotspot on mean blocking and total delay for a 16×16 switch. The input process has the same characteristics as that of the Bellcore trace `pAug.TL`.

In our analytical model, asymmetry in the correlation or the Hurst parameter of the traffic at the input ports does not affect the delay performance of the other ports as long as the arrival rate remains constant. This is because the "service time" for a port depends on the blocking delay and the only factor affecting the blocking delay at the ports are the arrival rates into the virtual queues of the outputs. Thus the "service times" at all the ports in the presence of parameter asymetries is the same. Hence, if the arrival rates are the same, differences in H, ρ and n do not have any effect on the "service times" of the other ports. However, the total delay at the ports will depend on the traffic characteristics at that input port.

C. Effect of ρ and H on Total Delay

Now let us consider the effect of the correlation structure of the arrival process at each input on the delay throughput characteristics. Figure 10 shows the effect of variation of the correlation on delay characteristics. The three curves correspond to the case when the input processes have the same Hurst parameter ($H = 0.82$) and arrival rate but correlations at lag one of 0.532, 0.582 and 0.632. Each input port of the switch is fed with traces having the same parameters. Observe that the delay decreases substantially with lower correlations. This is due to the reduced probability of successive time units having packet arrivals and thus reducing the queuing at the inputs.

Finally we study the effect of variation in the Hurst Parameter. As in the previous case, we vary the Hurst parameter of the input streams keeping all other parameters constant. The delay throughput characteristics for the cases when the input steams at each port have Hurst parameters of 0.77, 0.82 and 0.87 for a correlation at lag one of 0.582 are shown in Figure 11. As before, each input port is fed with traces having the same statistical properties. Note that the delays decrease significantly with even slight reduction in the Hurst parameter. This can be explained by considering the fact that a lower H reduces the long range dependence and the burstiness thereby reducing the queue

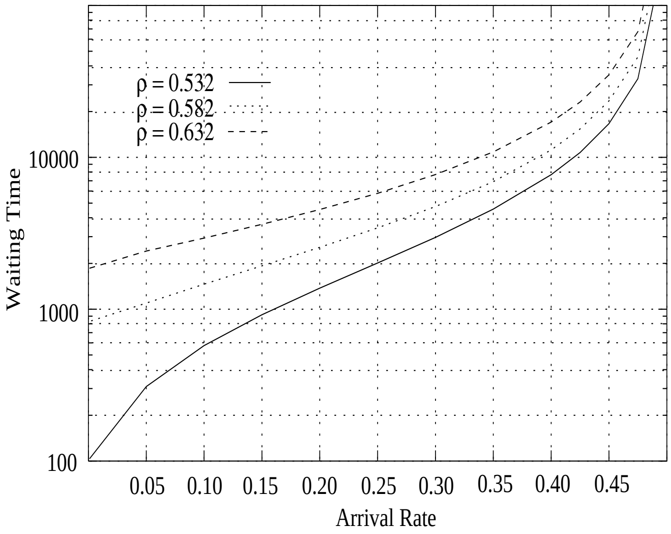

Fig. 10. The effect of variation in the correlation structure of the input traffic. We use $H = 0.82$ in the above results and the switch size is 16×16.

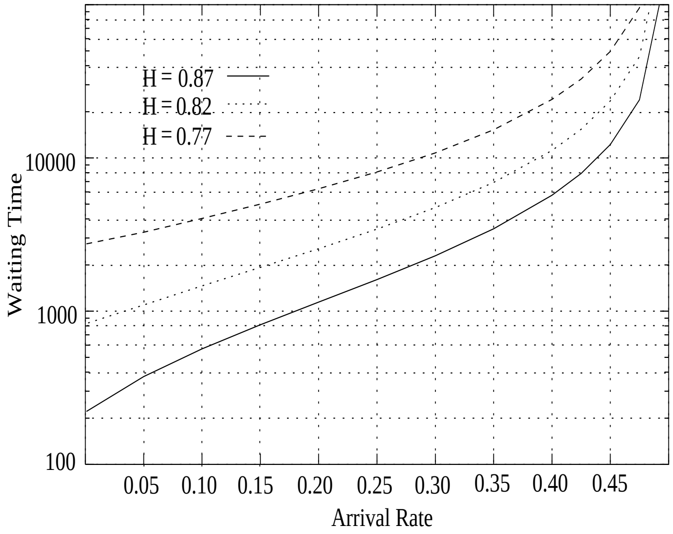

Fig. 11. The effect of variation in the Hurst parameter of the input traffic on different ports. We use $\rho = 0.582$ in the above results and the switch size is 16×16.

buildups at the inputs.

IV. Conclusion

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

In [7] it was conjectured that FCFS service in the virtual output queue gives the least average delay. Our analysis easily confirms this because FCFS service has the least variance and this is the variance of the "service time" of the input queue which is an M/G/1 queue. It is well known that for an M/G/1 queue the variance of the service time, in addition to the mean, contributes to the average delay. Also, from our models it is clear that the conjecture in [7] that the performance of an $M \times N$ switch is symmetric in M and N is not true.

From the throughput-delay characteristics of Figures 4, 10 and 11 we see that capturing all the statistical properties of the arrival processes is essential to characterizing the switch performance. Another important result to note is that operation in continuous time limits the maximum achievable throughput to 0.5, though, with a speedup factor of 4, the achievable throughput can be increased to more than 99%. Severe performance degradation takes place in the presence of hotspots, which can reduce the maximum throughput by 15% in a 16 × 16 switch. Also, Figures 10 and 11 highlight the large variations in the delay characteristics with changes in the correlation structure and the Hurst parameter. Lower Hurst parameters and correlation values reduce the burstiness of the arrival streams and reduces the queuing effects at the inputs and can give significantly lower delays at low loads. Thus the correlation structure and the Hurst parameter of the arrival processes are of extreme importance in determining the overall switch performance.

The model for variable length packet queues that we have developed here can easily extended to consider priorities in the input queue. Also extending it to analyse finite input buffer queues is rather straightforward and we do not present it due to lack of space. This can be done by considering the input queue as an M/G/1/K queue with the service time described by Eqn 5 for the case of Poisson arrivals using results from [3], [9], [16] and as an MMPP/G/1/K queue for the case of self similar arrivals modeled as an MMPP process using results from [2].

Finally, we add that our model does not address many architectures for variable length packet switches that are being considered today, specifically the virtual output queued (VOQ) switches. The buffer complexity of a VOQ switch is the same as that of a crossbar switch with crosspoint buffers – N buffers per input giving us N^2 buffers for a $N \times N$ switch. Furthermore, VOQ switches require complex scheduling algorithms to ensure fairness and a starvation-free operation. The scheduling algorithms proposed in [12], [13] are too impractical to be implemented in hardware [14]. Also, the "practical algorithm" of [14] still as a complexity of $\mathcal{O}(N^{2.5})$. Thus, although VOQ reduces the effect of head of line (HOL) blocking, it is complex and does not scale well enough to offset the throughput disadvantage of input queued switches for large N. Thus, alternate architectures like combined input output queued (CIOQ) switches with speedup are interesting and of practical value. They provide comparable throughputs for constant scale up of the switch. For CIOQ switches, our model gives the delay at the input buffer. The delay at the output buffer can be modeled separately.

Appendix

I. Delay Moments in an MMPP/G/1 Queue

The mean and second moment of the packet delay at input i, $\overline{D_i}$ and $\overline{D_i^2}$ respectively, are given by [6]

$$D_i = \frac{1}{\overline{X_i}\lambda_i}\left(w_v - \frac{1}{2}\overline{X_i^2}\lambda_i\right)$$

$$D_i^2 = \frac{1}{\overline{X_i}\lambda_i}\left(w_v^{(2)} - \frac{\overline{X_i^3}\lambda_i}{3} - D_i\overline{X_i^2}\lambda_i\right) \qquad (16)$$

where

$$w_v = \frac{1}{2(1-\overline{X_i}\lambda_i)}\Big[2\overline{X_i}\lambda_i + \overline{X_i^2}\lambda_i - 2\overline{X_i}((1-\overline{X_i}\lambda_i)\mathbf{g}_i$$
$$+\overline{X_i}\mathbf{\Phi}_i\mathbf{R}_i)(\mathbf{Q}_i + \mathbf{e}_i\mathbf{\Phi}_i)^{-1}\mathbf{r}_i\Big]$$

$$w_v^{(2)} = \frac{1}{3(1-\overline{X_i}\lambda_i)}\Big[3\overline{X_i}(2\mathbf{W}_i'(0)(\overline{X_i}\mathbf{R}_i - \mathbf{I}) - \overline{X_i^2}\mathbf{\Phi}_i$$
$$\mathbf{R}_i)\overline{X_i^2}\mathbf{\Phi}_i\mathbf{R}_i)(\mathbf{Q}_i + \mathbf{e}_i\mathbf{\Phi}_i)^{-1}\mathbf{r}_i - 3\overline{X_i^2}\mathbf{W}_i'(0)h+$$
$$\overline{X_i^3}\lambda_i\Big]$$

with

$$\mathbf{W}_i'(0) = (\overline{X_i}\mathbf{\Phi}_i\mathbf{R}_i + (1-\overline{X_i}\lambda_i)\mathbf{g}_i)(\mathbf{Q}_i + \mathbf{e}_i\mathbf{\Phi}_i)^{-1}$$
$$-\mathbf{\Phi}_i(1+w_v)$$

and g_i representing the steady state probability vector of the matrix G_i, the transition rate matrix of the embedded Markov chain at departure epochs with k packets in the queue and the MMPP arrival process in state j. We now present the procedure for calculating the matrix G and a general algorithm to calculate the first and second moments of the delay in an MMPP/G/1 queue [6].

A. *Computation of G_i for an m-state MMPP*

Initial Step : Define

$$G_i^0 = 0 \qquad H_i^{0,k} = I \qquad \text{for } k = 0, 1, 2, \cdots$$

$$\Theta = \max_j((R_i - Q_i)_{jj})$$

$$\gamma_n = \int_0^\infty e^{-\Theta x}\frac{(\Theta x)^n}{n!}dH(x) \qquad \text{for } n = 0, 1, \cdots, n^*$$

where n^* is chosen such that $\sum_{k=1}^{n^*}\gamma_k > 1 - \epsilon_1$, $\epsilon_1 \ll 1$.

Recursion : For $k = 0, 1, 2, \cdots$, do

$$H_i^{n+1,k} = \left[I + \frac{1}{\Theta}(Q_i - R_i + R_iG_i^k)\right]H_i^{n,k}$$

$$G_i^{k+1} = \sum_{n=0}^{n^*}\gamma_n H_i^{n,k}$$

Stopping Criterion :

$$\| G_i^{k-1} - G_i^k \| < \epsilon_2 \ll 1$$

Set $G_i = G_i^{k+1}$.

B. *Computation of γ_n*

The γ_n for Erlang-k and exponential service times are given by

1. *Erlang-k service*

$$\gamma_n = \int_o^\infty e^{-\Theta x}\frac{(\Theta x)^n}{n!}\mu^k\frac{x^{k-1}}{(k-1)!}e^{-\mu x}dx$$
$$= \frac{(n+k-1)!}{n!(k-1)!}\frac{\mu^k\Theta^n}{(\Theta+\mu)^{n+k}}$$

2. *Exponential service*

$$\gamma_n = \int_o^\infty e^{-\Theta x}\frac{(\Theta x)^n}{n!}\mu e^{-\mu x}dx$$
$$= \frac{\mu\Theta^n}{(\Theta+\mu)^{n+1}}$$

The γ_n for the service time distribution which is the summation of the phase-type distribution with Erlang-k service times and an exponential evacuation time is given by the weighted sum of the individual γ_n values. The weights are the probabilities of encountering each of the individual distributions, the $\pi_{ij}(k)$s.

C. *The MMPP/G/1 algorithm*

Step 1. Compute the matrix G for the given input port.
Step 2. Compute the steady state vector g which satisfies

$$g_iG_i = g_i \qquad g_i\epsilon = 1$$

Step 3. Compute the moments of the waiting time using Eqn 16.

REFERENCES

[1] A. T. Andersen and B. F. Nielsen, "A Markovian approach for modeling packet traffic with long-range dependence," *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 5, pp. 719-732, June 1998.

[2] C. Blondia, "The M/G/1 finite capacity queue," *Comm. Statist. Stochastic Models,* vol. 5, no. 2, pp. 273-294, 1989.

[3] P. J. Courtois, "The M/G/1 finte capacity queue with delays," *IEEE Trans.on Communications,* vol 28, pp. 165-172, 1980.

[4] A. Erramilli, O. Narayan and W. Wilinger, "Experimental Queuing Analysis with LRD Packet Traffic," *IEEE/ACM Trans on Networking,* vol 4, no 2, Apr 1996.

[5] A. Erramilli, R. P. Singh and P. Pruthi, "An application of deterministic chaotic maps to model packet traffic," *Queuing Systems,* vol. 20, pp. 171-206, 1995.

[6] W. Fischer and K. Meier-Hellstern, "The Markov modulated Poisson process (MMPP) cookbook," *Performance Evaluation,* vol. 18, no. 2, pp. 149-171, 1993.

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

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

[9] J. Keilson and L. D. Servi, "The M/G/1/K blocking formula and its generalizations," *Queueing Systems,* vol. 14, no. 1, 1993.

[10] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (Extended Version)," *IEEE/ACM Transactions on Networking,* vol. 2, no. 1, pp. 1-15, Feb 1994.

[11] S. Q. Li, "Performance of a non-blocking space division packet switch with correlated input traffic," *Proceedings of IEEE GLOBECOM 1989.*

[12] N. McKeown, "Scheduling algorithms for input-queued cell switches," *Ph.D. Thesis,* Univ of of California at Berkeley, 1995.

[13] N. Mckeown and A. Mekkittikul, "A starvation free algorithm for achieving 100% throughput in an input queued switch", *Proceeding of ICCCN 96,* Washington, DC, Oct.1996.

[14] N. McKeown and A. Mekkittikul, "A practical scheduling algorithm to achieve 100% throughput in input-queued switches", *IEEE INFOCOM 98,* pp. 792-799, Apr.1998.

[15] 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.

[16] S. C. Niu and R. B. Cooper, "Transform-Free Analysis of M/G/1/K and Related Queues," *Math of Oper Res,* vol. 18, no. 2, pp. 486-510, 1993.

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

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

[19] V. Paxson and S. Floyd, "Wide area traffic : The failure of Poisson modeling," *IEEE/ACM Transactions on Networking,* vol. 3, no. 3, pp. 226-244, June 1995.

[20] H. Perros, "Queuing Networks with Blocking," *Oxford University Press,* 1994.