A Performance Study of Virtual Output Queued Switches with Heterogeneous Bursty Traffic

Itamar Elhanany
Department of Electrical & Computer Engineering
Ben-Gurion University, Beer-Sheva, Israel
Email: itamar@ieee.org

Michael Kahane
Department of Electrical & Computer Engineering
Ben-Gurion University, Beer-Sheva, Israel
Email: michaelk@bgumail.bgu.ac.il

Dan Sadot
Department of Electrical & Computer Engineering
Ben-Gurion University, Beer-Sheva, Israel
Email: sadot@ee.bgu.ac.il

Abstract—**Virtual output queueing (VOQ) is a widely deployed buffering scheme in high-performance input-queued switches and routers. While there has been extensive investigation of the performance of switch architectures employing VOQ, the majority of the work addresses traffic that is uncorrelated and uniformly distributed among the outputs. This paper presents analysis for discrete-time virtual output queued switches with incoming traffic governed by a Markov modulated ON/OFF process, whereby bursts are non-uniformly distributed among the various destinations. Under the assumption of geometrically distributed interservice times, we utilize the probability generating functions of the interarrival times to obtain per-queue closed-form expressions for the mean queue occupancy and mean delay. The validity of the analytical inference is established through simulation results.**

***Index Terms*— Markov modulated ON/OFF process, Virtual Output Queueing, Input Queued Switches, Traffic Modeling, Performance analysis.**

I. INTRODUCTION

INPUT-queued switching architectures, commonly realized using crossbar technologies, are widely deployed in high-performance switches and routers [1], [2], [3]. In such systems, each input port may transmit to at most one output port at any given time. A scheduler, whether centralized or distributed, governs the switching process by determining the configuration of the crossbar matrix at any given time, thus enabling data cells to traverse the switch fabric. A well-known phenomenon named head-of-line blocking [4] occurs in single queued ingress architectures where contention on a specific egress path prevents non-blocked queued cells/packets from being transmitted, and in doing so limits the overall throughput of the switch. A common technique for overcoming the head-of-line blocking phenomena is virtual output queueing (VOQ) [5]. In VOQ a separate queue is maintained at the ingress port for each of the output destinations.

Markov modulated ON-OFF models have been repeatedly incorporated as building blocks for constructing more complex, pragmatic traffic scenarios. Multimedia traffic, which by nature tends to be correlated on several levels, is commonly modeled by a superposition of several ON-OFF sources [6]. The majority of the work presented in the literature, which provides analysis for scheduling algorithms under Markov modulated arrivals, addresses the case of deterministic (constant) inter-service times [7]. For many other applications a more accurate model for the inter-service times is that of geometrical distribution where in each time slot there is an independent probability of service which is typically derived from the scheduling algorithm employed. However, as the service discipline reflects on the scheduling algorithm deployed, even when under the assumption of a memoryless server it becomes quite intricate to accurately portray the queueing behavior.

In this paper we present analysis for a VOQ system with non-uniform destination distribution of Markov modulated arrivals and geometrically distributed interservice times. The paper provides a generic analytical framework, rather than precise deductions, for exploring the performance of input queues under different scheduling schemes. By further modeling the number of available destinations within each switching interval, the methodology presented in this paper may be applied to pragmatic scheduling schemes. Based on the probability generating function for the interarrival distribution, we obtain per-queue closed form expressions for the mean queue size and mean queue latency. Validation of the results is demonstrated through very good matching between analytical and simulation results.

II. QUEUEING MODEL FORMULATION

We begin by examining a discrete-time queueing system with a single-server and infinite buffer capacity, in which all events occur at fixed time slot intervals. Within each time slot, at most a single arrival and a single service event may occur. A late arrival model is considered, for reasons of convenience, such that within a time slot boundary a departure will always precede an arrival event. We observe the queue size at the instances following the arrival phase, such that the time slot boundaries are delimited by the observation instances. The service discipline is assumed to be governed by an i.i.d. Bernoulli process, resulting in geometrically distributed service times. Let μ denote the independent homogeneous probability of service at any given time slot. The basic stability condition dictates that $\lambda/\mu<1$, where λ is the normalized arrival rate. Let $Q(n)$ denote the queue occupancy at time slot

n, such that

$$Q(n) = Q(n-1) + A(n)\text{-}D(n) \ , \tag{1}$$

where $A(n) \in \{0,1\}$ and $D(n) \in \{0,1\}$ are the number of arrivals and departures during time slot n, respectively. In a stable system, the arrival rate must converge to the departure rate, such that

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} A(i)\right) = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} D(i)\right). \tag{2}$$

If this equation doesn't hold, the queue occupancy either grows to infinity or, alternatively, converges to zero. It has been shown in the literature [8] that in a *GI/Geo/1* discrete-time queueing system (general interarrival process and geometrically distributed service times), if f_n ($n \geq 1$) is the interarrival time distribution, with a p.g.f. $F(z) = \sum_{n=1}^{\infty} f_n z^n$, and the service times are geometrically distributed with parameter μ, then the stationary queue size distribution <u>*viewed by an arriving cell*</u>, π_m, will always be in the form

$$\pi_m = (1-\rho)\rho^m \qquad m \geq 0 \quad , \tag{3}$$

where ρ is a unique root of the equation

$$z = F\left(\mu z + (1-\mu)\right) \tag{4}$$

that lies in the region (0,1). Let $\pi_m = \Pr\{Q = m | arrival\}$ denote the probability of an arriving cell viewing m cells occupying the queue, while γ_m corresponds to the stationary probability of the queue size being m, regardless of arrivals. Interpreting the balance equation stated in (2) for a generic queueing system, we equate the mean arrival rate to the mean departure rate by writing

$$\begin{aligned} &E[arrival] = \lambda = E[departure] = E[Service \cap (Q>0)] = \\ &E[Q>0] \cdot E[Service | Q>0] = (1-\gamma_o)\mu \end{aligned} \ , \tag{5}$$

from which we isolate the stationary probability of the queue being empty, $\gamma_o = 1 - \lambda / \mu$. Using (3) the mean queue size, as viewed by an arriving cell, is

$$E[Q | arrival] = \sum_{m=1}^{\infty} m\pi_m = \frac{\rho}{1-\rho}. \tag{6}$$

From Little's result [8] we derive the mean latency experienced by arriving cells as

$$E[\tau] = E[Q | arrival] / \lambda = \frac{\rho}{\lambda(1-\rho)}, \tag{7}$$

where λ is the mean rate of arrival. A fundamental performance metric, from a system implementation perspective, is the VOQ buffer memory size. Given that γ_m are observed at the beginning of each time slot, they can be found based on an early arrival model for a *GI/Geo/1* queueing system [8]. Accordingly, the steady-state mean queue occupancy for the queue is in the form

$$\gamma_m = \begin{cases} 1-\xi & m = 0 \\ \xi(1-\rho)\rho^{m-1} & m \geq 1 \end{cases} \ , \tag{8}$$

where ρ is the same as in (4) and ξ is found using $\xi = 1 - \gamma_o = \lambda / \mu$. The latter allows us to describe the stationary behavior of the queue size as well as determine the mean queue occupancy,

$$E[Q] = \sum_{m=1}^{\infty} m\gamma_m = \frac{\xi}{(1-\rho)} = \frac{\lambda}{\mu(1-\rho)}. \tag{9}$$

The mean queue occupancy provides a good indication of the required VOQ buffering memory at each of the ports.

III. A Single Queue with ON/OFF Arrivals and Geometric Service Times

Consider a discrete-time, two-state Markov chain generating arrivals modeled by an ON-OFF source which alternates between the ON and OFF states. Let the parameters p and q denote the probabilities that the Markov chain remains in states ON and OFF, respectively. An arrival is generated for each time slot that the Markov chain spends in the ON state. The result is a stream of correlated bursts of arrivals and silent periods both of which are geometrically distributed in duration. It can easily be shown that the parameters p and q are interchangeable with the mean arrival rate, $\lambda = (1\text{-}q)/(2\text{-}q\text{-}p)$, and mean burst size, $B = 1/(1\text{-}p)$. Consequently, the offered load is identical to the steady-state portion of the time the chain spends in state ON.

The probability of two consecutive arrivals, $f_1 = p$, is identical to the probability that following an arrival the Markov chain remains in state ON. Similarly, f_2 is the probability that following an arrival, the chain transitions to the OFF state and then returns to the ON state. For n>2, it is apparent that following a transition from the ON state to the OFF state, there are $n\text{-}2$ time slots during which the chain remains in the OFF state before returning to state ON. As a result, we obtain the following general expression for f_n:

$$f_n = \begin{cases} p & n = 1 \\ (1-p)q^{n-2}(1-q) & n > 1 \end{cases}. \tag{10}$$

The corresponding p.g.f. is

$$\begin{aligned} F(z) &= \sum_{n=0}^{\infty} f_n z^n = \sum_{n=1}^{\infty} f_n z^n \\ &= pz + \sum_{n=2}^{\infty} q^{n-2}(1-p)(1-q)z^n \\ &= pz + (1-p)(1-q)\frac{z^2}{1-qz} \end{aligned} . \tag{11}$$

Next, we solve the equation $z = F(z\mu+(1\text{-}\mu))$ to find that the root in the region (0,1) is

$$\rho = \frac{(1-\mu)}{\mu}\left[\left(\mu(1-p-q)+q\right)^{-1}-1\right]. \tag{13}$$

IV. Virtual Output Queueing under Bursty Traffic

A. Homogeneous bursty arrivals

We extend the analytical foundations presented in section III to investigate the case of a switch model deploying virtual output queues with uniformly distributed Markov modulated arrivals. Letting N denote the number of ports, a burst is defined as a sequence of consecutive arrivals destined to the same output. The probability of the entire VOQ receiving service is μ, while within each VOQ service event all non-empty queues have an equal probability of being serviced. We label this arbitration discipline *random selection*, as the selection of a queue for transmission is done at random and does not consider any queue-state information.

This form of arbitration pertains to a generic study which is not directly applicable to scheduling packets in an input-buffered switch. By further limiting the number of available outputs in each time slot we may obtain a more accurate analysis for a given scheduling scheme.

We construct a Markov chain corresponding to the behavior of the investigated bursty arrival process, as shown in figure 1. The chain consists of N+1 states, N of which represent arrivals going to the N queues, while the last state is the OFF state. We label the ON states as $\kappa_1,\kappa_2,\ldots\kappa_N$, and the OFF state as κ_0. The probability of remaining in the OFF state is q while the probability of remaining in each of the ON states is p. To complement the latter, the probability of returning from any ON state to the OFF state is (1-p) while a transition from the OFF state to any of the ON states equals (1-q)/N. Thus, we can represent the Markov chain as an $(N+1)\times(N+1)$ transition probability matrix P where each element, p_{ij}, denotes the probability of transitioning from the i^{th} state to the j^{th} state among the states. The first row of P, with the exception of its first element, consists of the probabilities of transitioning from the OFF state to each of the ON states, signifying a beginning of a burst. The first column, with the exception of its first element, contains the probability of returning from each of the ON states to the OFF state (i.e. terminating of a burst). The first element on the diagonal is the probability of remaining in the OFF state while the rest of the diagonal elements are the probabilities of remaining in the ON states. Accordingly, the examined transition probability matrix is

$$P = \begin{pmatrix} q & \frac{1-q}{N} & \frac{1-q}{N} & \cdots & \frac{1-q}{N} \\ 1-p & p & 0 & \cdots & 0 \\ 1-p & 0 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1-p & 0 & 0 & \cdots & p \end{pmatrix} \tag{14}$$

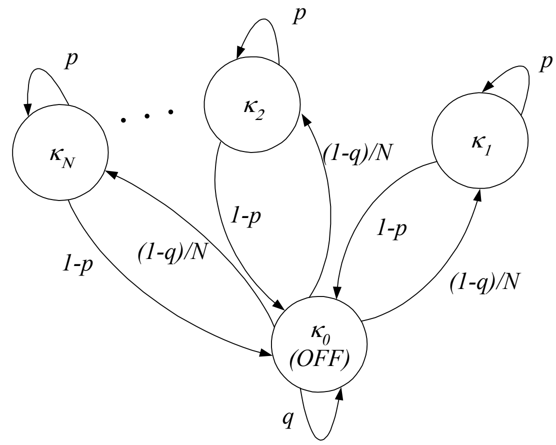

Fig. 1. Markov chain governing the behavior of a uniformly distributed ON/OFF arrival process to a VOQ with N ports. Each port receives an offered load of λ/N.

As with the single queue case, we would like to find, for each queue, the p.g.f. of the interarrival time distribution. The latter is done by utilizing the k-step transition matrix, $P^{(k)}$, in which each element, $p_{ij}^{(k)}$, represents the probability of transitioning from the i^{th} state to j^{th} state in precisely k-steps, with no restrictions made on passing through state j in any of the intermediate steps. In accordance with the Chapman-Kolmogorov equation [8] we have $P^{(k)} = P^k$ (k≥1), for which the p.g.f. is

$$P(z) = \sum_{n=0}^{\infty}(zP)^n = [I - zP]^{-1} \tag{15}$$

where $|z|<1$. We next define the k-step *first passage time* probability matrix, $F^{(k)}$, the elements of which, $f_{ij}^{(k)}$, are the probabilities of transitioning from state i to state j in *precisely* k-steps with the constraint that prior to the k^{th}-step the process has not visited state j. In other words, $f_{ij}^{(k)}$ denotes the probability of the first transition from state i to state j occurs in precisely k steps. It can be shown [8] that $f_{ij}^{(k)}$ is

$$f_{ij}^{(k)} = \sum_{s_1 \neq j}\sum_{s_2 \neq j}\cdots\sum_{s_{k-1} \neq j} p_{is_1} p_{s_1 s_2} \cdots p_{s_{k-2}s_{k-1}} p_{s_{k-1}j} \tag{16}$$

with $f_{ij}^{(1)} = p_{ij}$ and the s terms are the intermediate states between i and j.

Since the diagonal element, $f_{ii}^{(k)}\big|_{i>1}$, is by definition the probability of k steps separating two consecutive arrivals to queue i, it is identical to the definition of the inter-arrival time distribution of the i^{th} queue. It has been shown that the following relationship exists between $P_{ii}(z)$ and $F_{ii}(z)$ [8]:

$$F_{ii}(z) = 1 - \frac{1}{P_{ii}(z)} \,. \tag{17}$$

Accordingly, as a first step in finding F_{ii}(z), we need to find

$$P(z) = (I - zP)^{-1} = \begin{pmatrix} 1-zq & -z\frac{1-q}{N} & -z\frac{1-q}{N} & \dots & -z\frac{1-q}{N} \\ -z(1-p) & 1-zp & 0 & \dots & 0 \\ -z(1-p) & 0 & 1-zp & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -z(1-p) & 0 & 0 & \dots & 1-zp \end{pmatrix}^{-1} \quad (18)$$

However, we are exclusively interested in finding the diagonal elements of the inverted matrix, $P_{ii}(z)\big|_{i>1} = |M_{ii}| \cdot \left(|I - zP|\right)^{-1}$, where $|M_{ij}|$ denotes the determinant of a minor matrix of $[I - zP]$ with i^{th} row and j^{th} column removed. By multiplying each row in $[I - zP]$, except for the first, by $z(1-q)/(1-zp)/N$ and adding the result to the first row, we reduce $[I - zP]$ to an equivalent triangular matrix in the form

$$\begin{pmatrix} \frac{z^2(1-p)(1-q)}{(1-zp)N} & 0 & 0 & \dots & 0 \\ -z(1-p) & 1-zp & 0 & \dots & 0 \\ -z(1-p) & 0 & 1-zp & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -z(1-p) & 0 & 0 & \dots & 1-zp \end{pmatrix} \quad (19)$$

the determinant of which is simply the multiplication of all its diagonal elements. Applying the same rule to find the determinant of the minor matrix, $|M_{ii}|$, yields

$$P_{ii}(z)\big|_{i>1} = \frac{1 - zq + \frac{z(N-1)(1-q)}{N(1-zp)}}{1 - zq + \frac{z(1-q)}{1-zp}}. \quad (20)$$

Substituting the appropriate matrix elements, we obtain the p.g.f. of the inter-arrival time distribution as

$$F_{ii}(z)\big|_{i>1} = \frac{\left[p^2 q - p(1-p)(1-q)\right]z^3 + \left[(1-p)(1-q)/N - pq - p^2\right]z^2 + pz}{\left[pq - (N-1)(1-p)(1-q)/N\right]z^2 - (p+q)z + 1} \quad (21)$$

To facilitate the completion of the analysis, we are left with solving the equation $F(z\mu' + 1 - \mu') = z$ where μ' denotes the probability of service to each queue given that the queue is non-empty. Due to the lack of inter-queue arrival correlation, the sum of stationary mean queue sizes of all VOQs must equal the mean queue size of the single-queue model with the same arrival and service patterns. Using (9) we express the stationary mean queue size of any given queue, $E[Q_k]$, as

$$E[Q_k] = E[Q]/N = \frac{\lambda}{\mu N(1-\rho)}, \quad (22)$$

where $E[Q]$ denotes the mean queue occupancy of a single-queue receiving all traffic and having a probability of service μ. Consequently, we also have the following identity for each queue

$$E[Q_k] = \frac{\lambda}{\mu' N(1-\rho')} \quad (23)$$

which when equated to (22) yields

$$\mu' = \frac{\mu(1-\rho)}{(1-\rho')}. \quad (24)$$

Next, we substitute the above in $F(z\mu' + 1 - \mu') = z$ to obtain

$$\rho' = F\left(1 - \mu(1-\rho)\right), \quad (25)$$

where ρ is found as shown in section III. Having found ρ' we can now derive the mean packet latency and other queue behavior metrics.

B. Heterogeneous Distributed Bursty Arrivals

We next consider the case where bursty traffic is non-uniformly distributed between the virtual output queues. We characterize the traffic for each queue by the portion of the offered load it receives, λ_k ($k=1,2,..N$), which can be arbitrary and a mean burst size, B. For the traffic to be admissible, we require that λ_k are selected such that they satisfy $\sum_{k=1}^{N} \lambda_k = \lambda < \mu$. Since two consecutive bursts are always separated by at least one time slot of no arrivals, the maximal achievable load for any queue is bounded by $B/(B+1)$. The probability of remaining in the same state (transition from state k back to state k) is p, while the probability of transitioning from the OFF state (κ_0) to state k is q_k. Letting U denote process state, for the Markov chain to be stable we observe that any pair (κ_0,κ_i) must satisfy

$$\Pr\{U = \kappa_i\}(1-p) = \Pr\{U = \kappa_0\}q_i \quad (26)$$

Since $\Pr\{U = \kappa_i\}$ represents the steady-state portion of the time the chain spends in state κ_i, we conclude that $\Pr\{U = \kappa_i\} = \lambda_i$. However, by definition $\Pr\{U = \kappa_0\} = 1 - \lambda$, which leads to

$$q_i = \begin{cases} 1 - \sum_{i=1}^{N} q_i & i = 0 \\ \frac{\lambda_i(1-p)}{1-\lambda} = \frac{\lambda_i}{B(1-\lambda)} & i = 1,2,\dots N \end{cases}, \quad (27)$$

from which we fully construct P.

Based on the rationalization provided above for the case of homogenous traffic, we note that the queue size in non-uniform distribution scenarios also obeys a geometric distribution. In the case of the latter, each queue is associated with a distinctive parameter ρ_k, reflecting on the observation that the traffic arriving to each queue is not identical. We practice the same technique deployed previously whereby using the generic transition probability matrix, P, we find $P(z)=(I-zP)^{-1}$. Algebraic exploration of the latter yields the following generic result for $p(z)_{ii}$,

$$p(z)_{ii}\big|_{i>1}=\frac{1-zp_{11}-\sum_{j=2,j\neq i}^{n+1}\frac{z^2p_{j1}p_{1j}}{1-zp_{jj}}}{\left[1-zp_{11}-\sum_{j=2}^{n+1}\frac{z^2p_{j1}p_{1j}}{1-zp_{jj}}\right](1-zp_{ii})}, \tag{28}$$

from which we find $f(z)_{ii}$. The latter offers the required interarrival time distribution p.g.f., $F_k(z)=f(z)_{kk}\big|_{k>1}$, for each of the N queues. Next, we are required to solve the unique equation $F_k(z\mu_k+1-\mu_k)=z$ for each queue. We define the roots of these equations, in the region (0,1), as the geometric distributions coefficients ρ_k. Due to the uncorrelated nature by which bursts are distributed among the different queues, we conclude that if the load to queue k is λ_k then the mean queue size of that queue is $E[Q_k]=\lambda_k E[Q]/\lambda$, where $E[Q]$ is the mean queue size of the single-queue aggregate model as described above. From the queue-size identity

$$E[Q_k]=\frac{\lambda_k}{\mu_k(1-\rho_k)}=\frac{\lambda}{\mu(1-\rho)} \tag{29}$$

we have

$$\mu_k=\frac{\mu(1-\rho)}{(1-\rho_k)}, \tag{30}$$

which when substituted in $F_k(z\mu_k+1-\mu_k)=z$ yields

$$\rho_k=F_k\left(1-\mu(1-\rho)\right). \tag{31}$$

Obtaining ρ, the queue size distribution coefficient of the single-queue aggregated model, requires taking into consideration the non-uniformity of the arriving traffic. Since the mean burst size of the aggregate-model is also B, we conclude that $\bar{p}=1-1/B$ and in order to achieve a mean offered load of λ, we have

$$\bar{q}=\frac{1-\lambda(2-\bar{p})}{1-\lambda}, \tag{32}$$

where $\bar{p}$ and $\bar{q}$ are the equivalent probabilities of remaining in the ON and OFF states of the aggregate model, respectively. Using these parameters, we conclude that

$$\rho=\frac{(1-\mu)}{\mu}\left(\left(\mu\left(1-\bar{q}-\bar{p}\right)+\bar{q}\right)^{-1}-1\right). \tag{33}$$

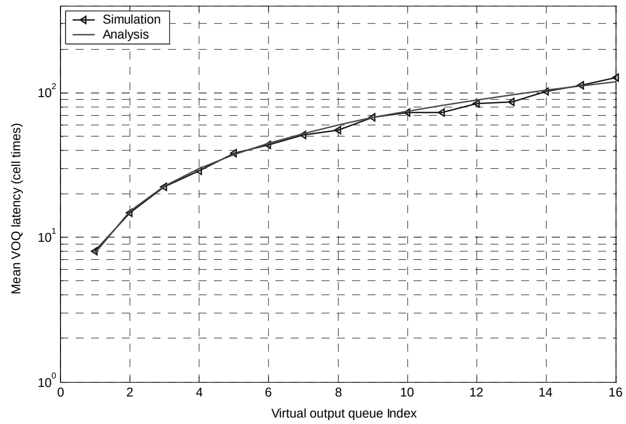

Fig. 2. Mean latency in a 16-port VOQ with λ = 0.7, μ=0.85, and arrivals distributed according to the $Zipf_{m=1}$ with mean burst sizes of 8 cells.

V. Simulation Results

As means of validating the analytical framework with simulation results, we employ a non-linear destination distribution model named Zipf's law [9]. The Zipf law states that the frequency of occurrence of some events, as a function of the rank (m) where the rank is determined by the above frequency of occurrence, is a power-law function: $P_k \sim 1/k^m$. A famous example of Zipf's law is the frequency of English words in a given text. Most common is the word "the", then "of", "to" etc. When the number of occurrence is plotted as the function of the rank (k=1 most common, k=2 second most common, etc.), the resulting form is a power-law function with exponential order typically close to 1. The probability that an arriving cell is heading to destination k is given by

$$\lambda_k=k^{-m}\left(\sum_{j=1}^{N}j^{-m}\right)^{-1}. \tag{34}$$

While m=0 corresponds to uniform distribution, and as m increases the distribution becomes more biased towards a preferred destination. Figure 2 illustrates the mean queueing latency for N=16, μ=0.85 and λ=0.7. Arriving traffic is distributed between the queues according to a $\text{Zipf}_{m=1}$ distribution with a mean burst size of 8 cells.

VI. Conclusions

In this paper we present an analytical technique for evaluating the queueing behavior of a virtual output queued switch with non-uniformly distributed Markov modulated ON/OFF arrivals and geometrically distributed interservice times. Exploiting the unique relationship existing between the probability of transition and the k-step first passage time distribution, closed-form expressions for the mean queue occupancy and mean delay are obtained. The presented methodology can be extended to further analyze scheduling algorithms introduced with correlated traffic patterns.

REFERENCES

[1] J. Nir, I. Elhanany, and D. Sadot, "A new Tbit/sec switching scheme for ATM/WDM networks*,"* *Electronics Letters*, 35, no. 1, pp. 30-31, January 1999.

[2] N. McKeown, "The iSLIP Scheduling Algorithm for Input-Queued Switches," *IEEE/ACM Trans. on Networking*, Vol. 7, No. 2, pp. 188-201, April 1999.

[3] R. Rojas-Cessa, E. Oki, H. J. Chao**, "**CIXOB-k: Combined Input-Crosspoint-Output Buffered Packet Switch," *Proc. IEEE GLOBECOM 2001*, Vol. 4, pp. 2654-2660, Nov. 2001.

[4] M. J. Karol, M. G. Hluchyj, and S. P. Morgan, "Input versus Output Queueing in a Space Division Switch," *IEEE Trans. Communications*, Vol. COM-35, pp. 1347-1356, Dec. 1987.

[5] Y. Tamir, G. Frazier, "Higher Performance Multiqueue Buffers for VLSI Communication Switches," Proc. Of 15th Annual Symposium on Computer Architecture, pp. 343-354, June 1988

[6] A. La Corte, A. Lombardo, G. Schembra, "Modeling Superposition of ON-OFF Correlated Traffic Sources in Multimedia Applications," *Proc. INFOCOM 1995*, pp. 993-1000.

[7] S. Wittevrongel, H. Bruneel, "Discrete-Time Queues with Correlated Arrivals and Constant Service Times," *Computers & Operations Research*, 1999, vol. 26, no. 2, pp. 93-108.

[8] J. J. Hunter, *Mathematical Techniques of Applied Probability: Discrete Time Models*: *Techniques and Applications*, Vol. 2, Academic Press, 1983.

[9] G. K. Zipf, *Psycho-Biology of Languages*, Houghton-Miffin, MIT Press, 1965.