# On the performance of output queued cell switches with non-uniformly distributed bursty arrivals

I. Elhanany and B. Matthews

**Abstract:** A novel performance analysis of output queued cell switches that are introduced with general independent heterogeneous traffic is presented. Random arbitration is employed whereby non-empty queues compete equally for service within each switching interval. In particular, the case of bursty two-state Markov-modulated arrivals is studied in which input ports generate bursty streams that are non-uniformly distributed. Under the assumption of a memoryless server, the probability generating function of the interarrival process is utilised to derive an approximation for the queue size distribution. The methodology established forms a flexible tool in obtaining bounds on the behaviour and expected performance characteristics of output queued switches under a wide range of correlated traffic scenarios. The validity of the analytical inference is established through simulation results.

## 1 Introduction

Output queued switches have been extensively studied in the literature. To a large extent, they represent the theoretical limit on the performance that can be achieved in any space-division switching fabric [1]. Consequently, performance analysis of input queued switches is commonly carried out in comparison to that of an output queued switch [1–3]. Pragmatic output queued switching fabrics, such as those based on shared memory architectures [4], have been deployed in switches and routers. However, the majority of the studies performed on these systems considers traffic that obeys a Bernoulli (uncorrelated) process. Moreover, in most cases uniform destination distribution is assumed such that all input ports offer the same load intensity to all output ports.

In a basic output queued switch architecture, arriving cells traverse the switching fabric directly to their designated outputs, without being queued or delayed in any way at the ingress stage. Such a scheme requires that a dedicated link, be it logical or physical, exist between each input port and each output port. In other words, $N^2$ such links are needed for an $N \times N$ switch. A key advantage of output queued switches is that of minimal latency and controllable QoS provisioning, both a result of the fact that arriving cells progress towards their destination without any impediment at the ingress. In practical implementations, a shared-memory architecture is employed, whereby multiple arriving cells are stored in a single physical memory unit. If a single memory unit is utilised at each output port, an $O(NR)$ memory bandwidth requirement results, where $R$ denotes the cell arrival rate. Figure 1 depicts a typical model for queueing architectures in output queued switches, whereby each port generates cells at a different mean arrival rate.



*IEE Proceedings* online no. 20050331

doi:10.1049/ip-com:20050331

Paper first received 6th July and in final revised form 2nd December 2005

The authors are with the Dept. of Electrical and Computer Engineering, 401-A Ferris Hall, University of Tennessee, Knoxville, TN 37996-2100, USA

E-mail: itamar@ieee.org

This paper presents analysis for output queued switches introduced with non-uniformly distributed bursty arrivals that are generated using a multitude of ON/OFF arrival processes. We apply random arbitration between the queues such that non-empty queues compete equally for service during each time slot. Random arbitration is considered primarily since it represents, from a hardware implementation perspective, a simple, scalable approach to arbitrating between multiple queues. More sophisticated arbitration schemes, such as those considering information regarding the queues' states, are expected to at least match, if not exceed, the performance of random arbitration. In this respect, random arbitration provides a lower bound on the attainable performance of the generic switch architecture considered. Based on the per-queue probability generating functions of the interarrival times distribution, it is shown that accurate depiction of the queues' behaviour can be obtained.

## 2 Queueing model with bursty arrivals

### 2.1 Notation and formulation

Packets may vary in size as they arrive at the switch ports. In typical switching platforms, a segmentation module partitions packets into fixed-size cells that are later

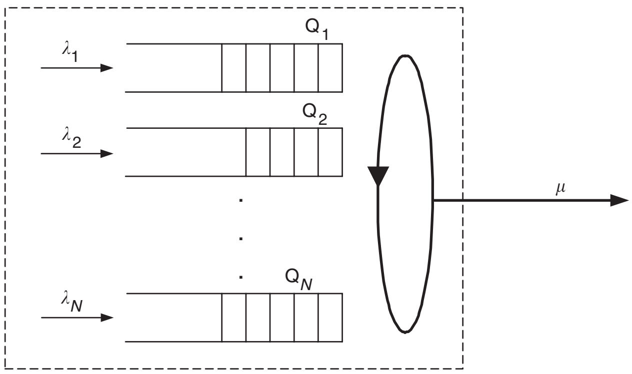


**Fig. 1** *A basic logical model of the queueing architecture employed at an egress port of an output queued switch*

reassembled at the egress modules prior to departing the switch. Processing fixed-size data units has proven both practical and easier to study. To that end, all data units traversing the switch fabric are assumed to be of fixed size. We consider a discrete-time queueing system with $N$ queues of infinite buffer capacity and a single server, in which all events occur at fixed time slot intervals. Within each time slot, at most $N$ arrivals may occur, originating from the $N$ inputs. Consequently, at most a single departure occurs servicing one of the non-empty queues at each time slot.

We model the service distribution as a memoryless process for which there is a constant per-queue probability of service, $\mu_k$, during each time slot. The aggregate service rate is, respectfully, defined as

$$\mu = \sum_{k=1}^{N} \mu_k \tag{1}$$

In order for the system to be stable, we require that the service rate, $\mu$, exceed the aggregate steady-state rate of arrivals such that

$$\sum_{j=1}^{N} \lambda_j < \mu \tag{2}$$

For a stable system, we can state that for each queue, the probability of arrival, $\lambda_k$, must converge to, or even equal, the departure rate such that

$$\lambda_k = \mu_k(1 - \pi_0^{(k)}) \tag{3}$$

where $\mu_k$ denotes the probability that queue $k$ is serviced given that the queue is non-empty, and $\pi_0^{(k)}$ is the probability that the queue is empty.

### 2.2 ON/OFF arrivals with geometric service times

It has been shown in the literature [5] that in a GI/Geo/1 discrete-time queueing system (independent arrivals times and geometrically distributed service times), if $f_n$ $(n \geq 1)$ is the interarrival time distribution, with a p.g.f. $F(z) = \sum^{\infty} f_n z^n$, and queue service times are geometrically distributed with parameter $\mu_k$, then the stationary queue size distribution, $\pi_m = (1-\gamma)\gamma^m$ $m \geq 0$, as viewed by an arriving cell will always be in the form

$$\pi_m = \begin{cases} 1 - \xi & m = 0 \\ \xi(1-\gamma)\gamma^m & m \geq 1 \end{cases} \tag{4}$$

where $\gamma$ is a unique root of the equation $z = F(\mu z + (1-\mu))$ that lies in the region $(0, 1)$ and $\xi$ is constant. The latter is, by definition, independent of arrivals. Hence, utilising (4) to derive $\xi$ yields the first moment

$$E[Q] = \sum_{m=1}^{\infty} m\pi_m = \frac{\xi}{(1-\gamma)} = \frac{\lambda}{\mu(1-\gamma)} \tag{5}$$

which provides us with the mean queue occupancy. Employing Little's results [5, 6], the mean latency is given by

$$E[W] = \frac{1}{\mu(1-\gamma)} \tag{6}$$

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 instances following the arrival phase, such that time slot boundaries are delimited by the observation instances. Consider a discrete-time, two-state Markov chain generating arrivals modelled by an ON/OFF source that alternates between the ON and OFF states. While in the ON state, a single cell is generated (per time slot). 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 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, $\lambda = (1-q)/(2-q-p)$, and mean burst size, $B = 1/(1-p)$. Consequently, the offered load is identical to the steady-state portion of the time the chain spends in state ON. Recalling the notation $f_n$ for the interarrival times distribution, the probability of two consecutive arrivals occurring is identical to the probability that following an arrival the Markov chain remains in state ON, i.e. $f_1 = p$. 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 $n-2$ time slots during which the chain remains in the OFF state before returning to the ON state. Accordingly, we obtain the following general expression for $f_n$

$$f_n = \begin{cases} p & n = 1 \\ (1-p)q^{n-2} \cdot (1-q) & n > 1 \end{cases} \tag{7}$$

The corresponding p.g.f. is

$$F(z) = pz + (1-p)(1-q)\frac{z^2}{1-qz} \tag{8}$$

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

$$\gamma = \frac{(1-\mu)}{\mu}\left[\frac{1}{\mu(1-p-q)+q} - 1\right] \tag{9}$$

Examining the condition $\gamma < 1$, which must be satisfied for stability, yields the anticipated inequality

$$\mu > (1-q)/(2-p-q) = \lambda \tag{10}$$

## 3 Output queued switch model with bursty arrivals

In the investigated system, random arbitration is employed, suggesting that during each time slot for which a service event occurs, one of the non-empty queues is randomly selected for transmission in an unbiased manner. The latter implies that the service discipline to each queue is also memoryless since during each time slot no information regarding previous service cycles is considered. As such, we will exploit the results for the GI/Geo/1 queueing system to derive approximate behavioural analysis of the individual output queues.

We focus our analysis on queue $k$ observing that three conditions must be met during each time slot for the queue to be serviced: (1) service must be granted to the port (i.e. output link not congested), (2) the queue must be non-empty and, finally, (3) the queue must prevail when equally contending against the other non-empty queues. While the first two conditions are rather straightforward, the third condition requires some elaboration. Assuming that queue $k$ is non-empty $(Q_k > 0)$, it has an equal probability of being selected for transmission as any other non-empty queue. The mean number of non-empty queues, excluding queue $k$, can be approximated by $\sum_{j \neq k}(1 - \pi_0^{(j)})$ where $\pi_0^{(j)}$ denotes the stationary probability that queue $j$ is empty, provided that queue $k$ is non-empty. To obtain the mean size of the contending set, we then add one to the mean number of non-empty queues. By multiplying the expressions for the three conditions stated above, we find the probability of departure from queue $k$ to be

$$\mu_k \cdot (1 - \pi_0^{(k)}) = \frac{\mu(1 - \pi_0^{(k)})}{\sum_{j \neq k} (1 - \pi_0^{(j)}) + 1} = \frac{\mu(1 - \pi_0^{(k)})}{N - \sum_{j \neq k} \pi_0^{(j)}} \quad (11)$$

Since the arrival rate should converge to, or even equal, the departure rate, we equate (11) to the rate of arrivals for each queue, yielding

$$\lambda_k = \frac{\mu(1 - \pi_0^{(k)})}{N - \sum_{j \neq k} \pi_0^{(j)}} \quad (12)$$

The latter holds when we assume that, for large values of $N$, the conditional probability of each queue being empty, given the size of other queues, converges to the unconditional probability of that queue being empty. Hence, we have $N$ linear equations for the $N$ variables $\pi_0^{(j)}$ $(j = 1, 2, \ldots, N)$. Solving for $\pi_0^{(j)}$ we directly obtain $\mu_k$, the probability of service to each of the queues.

We next turn our attention to the case where each input produces a stream of bursty arrivals modelled by a unique ON/OFF process, with respective parameters $p_k$ and $q_k$ pertaining to traffic originating at input (source) $k$. In view of the fact that queue sizes are geometrically distributed, based on (9) the respective parameters of these distributions are

$$\gamma_k = \frac{(1 - \mu_k)}{\mu_k} \left[ \frac{1}{\mu_k(1 - p_k - q_k) + q_k} - 1 \right] \quad (13)$$

We note that, as expected, $\gamma_k$ is a function of both the arrival model parameters and the rate at which the output queues are serviced. Fundamental performance metrics, such as the mean cell latency, are directly derived for each of the queues as shown in (6).

## 4 Simulation results

Our simulations pertain to a scenario where traffic is both bursty and non-uniformly distributed between the inputs. Arrivals are generated by an ON/OFF model which is independently operated for each input. As means of validating the analytical deductions with simulation results, we employ a nonlinear destination distribution model named Zipf's law [7, 8]. 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: $P_k \sim 1/k^m$. Accordingly, the probability that an arriving cell is heading to destination $k$ is given by

$$\lambda_k^{(m)} = k^{-m} \left( \sum_{j=1}^{N} j^{-m} \right)^{-1} \quad (14)$$

While $m = 0$ corresponds to a uniform distribution, as $m$ increases the distribution becomes more biased towards preferred destinations. There has been recent evidence that the Zipf model accurately portrays web caching and access statistics, in particular when the parameter $m$ is close to unity. Recent studies have illustrated the presence of Zipf law characteristics in Internet traffic patterns [9].

Previous work [1] has shown that the mean queue size in an output queued switch with first-in-first-out (FIFO) arbitration and traffic arriving uncorrelated and uniformly distributed, is

$$E[Q]_{\substack{FIFO \\ Bernoulli}} = \frac{\lambda^2(N - 1)}{2(1 - \lambda)N} \quad (15)$$

where $N$ denotes the number of ports in the switch. Uniformly distributed ON/OFF arrivals presented to an output queued switch employing FIFO arbitration yield the following mean queue size

$$E[Q]_{\substack{FIFO \\ ON/OFF}} = \frac{\lambda B(N - 1)}{(1 - \lambda)N} \quad (16)$$

where $B$ is the mean burst length. Numerous simulations indicate that the above is accurate, however at present we do not have an analytical proof of this result. Formalism that provides closely related foundations for the above can be found in [10]. By dividing each of the above mean queue sizes by the normalised offered loads, we obtain the corresponding mean cell latencies. We shall refer to these assertions when comparing the latency of the FIFO discipline to that of random arbitration.

Figure 2 illustrates the mean queue occupancy for each of the output queues in a 16-port switch employing random arbitration, where $\mu = 0.8$ and $\lambda = 0.7$. Arriving traffic is distributed between the queues according to a $\text{Zipf}_{m=0.5}$ distribution with a fixed mean burst size of 6 cells. As can be observed, the latency for each queue is well correlated with its share of the offered load. The two curves correspond to the simulated model and analytical approximation, as described in Section 3. The simulation results clearly validate the accuracy of the proposed analytical inference.

In Fig. 3, the queue size distribution is shown, for a 16-port switch that is introduced with uniformly distributed bursty traffic with aggregate normalised offered load of 0.75. The mean burst size is 8 cells, while the probability of service ($\mu$) is 0.9. The latter can reflect, for example, on a system that has a 10% congestion (no-service) time. As discussed in Section 3, the analytical results were obtained using (3), (11), (12) and (13).

In Fig. 4 the mean queue latency is presented as a function of the offered load for both random and FIFO arbitration. Traffic is assumed to be governed by an ON/OFF model with a mean burst size of 8 cells. Results for the case of random arbitration are shown for different

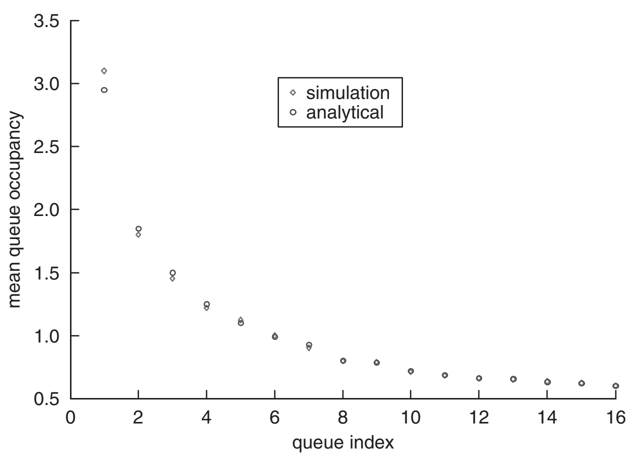


**Fig. 2** *Mean queue occupancy for each output queue in a 16-port switch employing random arbitration, where $\lambda = 0.8$ and $\mu = 1$*
Arriving traffic is distributed according to the $\text{Zipf}_{k=0.5}$ with mean burst sizes of 6 cells

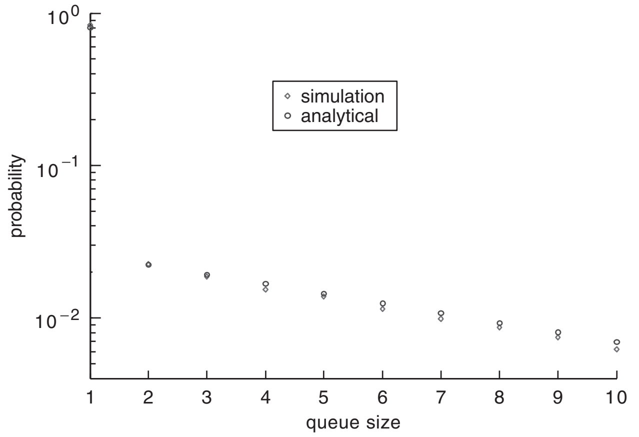


**Fig. 3** *Queue size distribution for a 16-port switch with normalised offered load of 0.75, mean burst size of 8 cells, and a probability of service of 0.9*

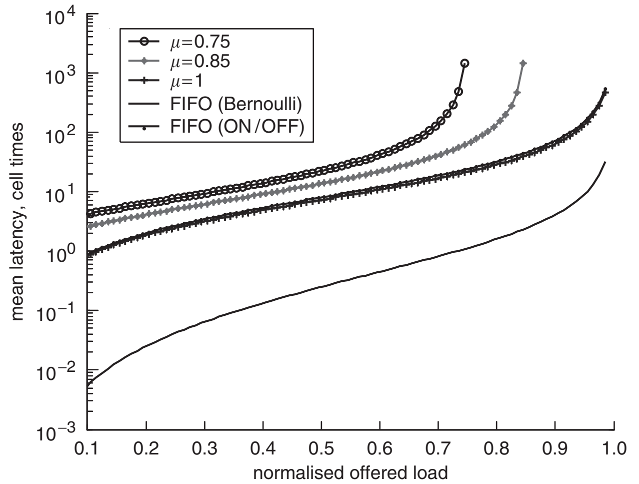


**Fig. 4** *Mean queueing latency against the offered load for random arbitration and FIFO schedulers*

probability of service values, the impact of which is clearly observable in particular as the offered load approaches the service rate. Also shown are the mean latency attributes of FIFO arbitration, which in many cases constitute the theoretical lower limit on the delay through a space division switch. For the case of $\mu = 1$, FIFO and random arbitration yield identical results, suggesting that the presented methodology for obtaining the latency under generic traffic conditions provides valuable approximation to a pure FIFO system, as would be expected.

## 5 Conclusions

In this paper we present an analytical framework for evaluating the queueing behaviour of an output queued switch that is introduced with non-uniformly distributed independent arrival processes. Under the assumption of a memoryless service discipline, the probability generating functions of the interarrival times distributions are utilised to derive per-queue approximations for the queue size distribution. The results offer an upper bound on the mean latency, as any arbitration scheme that considers additional system state information, such as queue sizes and the cell waiting times, is bound to yield higher performance. Arbitration schemes that prioritise queues (e.g. strict priority) do not necessarily improve performance on a per-queue level, however they tend to lead to more desirable system behaviour. To demonstrate the applicability of the proposed technique, detailed analysis of the case of non-uniformly distributed ON/OFF traffic is provided. Simulations results are offered, validating the accuracy of the approach taken.

## 6 Acknowledgments


This work has been partially supported by the Department of Energy (DOE) under research grant DE-FG02-04ER25607, and by the Woodrow W. Everett, Jr. SCEEE Development Fund in cooperation with the Southeastern Association of Electrical Engineering Department Heads.


## 7 References


1 Karol, M.J., Hluchyj, M.G., and Morgan, S.P.: 'Input versus output queueing in a space division switch', *IEEE Trans. Commun.*, 1987, **COM-35**, pp. 1347–1356

2 Elhanany, I., and Sadot, D.: 'DISA: a robust scheduling algorithm for scalable crosspoint-based switch fabrics', *IEEE J. Sel. Areas Commun.*, 2003, **21**, (4), pp. 535–545

3 McKeown, N.: 'The iSLIP scheduling algorithm for input-queued switches', *IEEE/ACM Trans. Netw.*, 1999, **7**, (2), pp. 188–201

4 Arpaci, M., and Copeland, J.: 'Buffer management for shared-memory ATM switches', *IEEE Commun. Surv. Tutor.*, 2000, **3**, (1), pp. 2–10

5 Hunter, J.J.: 'Mathematical techniques of applied probability: discrete time models: techniques and applications' (Academic Press, 1983), Vol. 2

6 Chaudhry, M.L., Gupta, U.C., and Templeton, J.G.C.: 'On the relations among the distributions at different epochs for discrete-time GI/Geom/1 queues', *Oper. Res. Lett.*, 1996, **18**, pp. 247–255

7 Breslau, L., Cao, P., Fan, L., Phillips, G., and Shenker, S.: 'On the implications of Zipf's law for web caching'. Proc. of IEEE INFOCOM'99, New York, March 1999

8 Yu, F., Zhang, Q., Zhu, W., and Zhang, Y.-Q.: 'QoS-adaptive proxy caching for multimedia streaming over the Internet', *IEEE Trans. Circuit Syst. Video Technol.*, 2003, **13**, (3), pp. 257–269

9 Söderqvist, M., and Gunnar, A.: 'Performance of traffic engineering using estimated traffic matrices'. Proc. of Radio Sciences and Communication RVK'05, Linköping, Sweden, June 2005

10 Bruneel, H.: 'Queueing behavior of statistical multiplexers with correlated inputs', *IEEE Trans. Commun.*, 1988, **COM-36**, (12), pp. 1339–1341