Online ISSN 1848-3380, Print ISSN 0005-1144
ATKAFF 56(2), 196–206(2015)

Fayza A. Nada, Medhat A. Rakha

Performance Evaluation of Space Division Output Buffered Switches

DOI 10.7305/automatika.2015.07.763
UDK 004.62; 004.42

Original scientific paper

Discrete-time multiserver queues have been successfully used in the performance evaluation of computer and communication systems in various contexts such as TDM, voice-data integration, deflection routing and more recently, ATM-switching technology. In digital communication systems, buffers are used for the temporary storage of digital information awaiting transmission via some communication channel(s). Usually, information presents itself in the form of fixed-size packets, and synchronous (slotted) transmission is used so that the buffers can be modelled as discrete-time queuing systems. In this article, performance analysis of space division output buffered switches operating in an ATM multimedia environment is presented. Fixed size packets arrive onto the switch inputs in each time slot. The switch is modelled as a discrete-time, batch arrival, multiserver queuing system, with infinite buffer and geometric service times. The main contribution of this work is using multiserver case and geometric service time. Many performance measures such as System occupancy, Queuing time, Unfinished work and Waiting time are analysed. Probability Generating Functions (PGFs) and expectations for the corresponding random variables are derived. The results of the analysis have been verified in many ways. First, they have been used to generate the results of some previous analyses as special cases. Second, they obviously seem to preserve classical queuing relations (e.g. the famous Little's formula). Last, they have been shown to generate intuitively acceptable graphs, when translated into numerical values.

Key words: Data Communications, Performance Analysis, Output Buffered Switches, System Occupancy, Multiserver queues

Procjena učinkovitosti prospajanja komutatora. Diskretni višeserverski redovi uspješno su korišteni kod analize učinkovitosti računalnih i komunikacijskih sustava u različitim okolnostima poput TDM-a, glasovno-podatkovne integracije, ATM komutatora. U sustavima digitalnih komunikacija bufferi se koriste za privremeno pohranjivanje digitalnih informacija koje čekaju na prijenos preko komunikacijskih kanala. Uobičajeno se informacije prikazuju u formi paketa fiksne veličine, a sinkronizirani prijenos koristi se kako bi bufferi mogli biti modelirani kao diskretni sustav s čekanjem. U ovome članku prikazana je analiza učinkovitosti komutatora u ATM multimedijskom okruženju. Paketi fiksne veličine stižu na ulaze preklopnika u svakom vremenskom okviru. Preklopnik je modeliran kao diskretni višeserverski sustav s batch dolascima, s beskonačnim bufferom i geometrijskom razdiobom vremena servisa. Glavni doprinos ovog rada je korištenje višeserverskog slučaja i geometrijske razdiobe vremena servisa. Analizirane su mnoge mjere učinkovitosti kao što je zauzetost sustava, vrijeme čekanja, nezavršen posao. Izvedena je funkcija izvodnica i očekivanje za odgovarajuće slučajne varijable. Rezultati analize provjereni su na mnoge načine. Prvo su korišteni za generiranje rezultata nekih prethodnih analiza kao posebnih slučajeva. Zatim je utvrđeno da očito zadržavaju klasične relacije (npr. poznata Littlova formula). Konačno, pokazano je da generiraju intuitivno prihvatljive grafove kada se prenesu u numeričke vrijednosti.

Ključne riječi: prijenos podataka, analiza učinkovitosti, komutatori, zauzetost sustava, višeserverski redovi

1 INTRODUCTION

Performance analysis is the study of the performance and behaviour of computer systems in order to make choices in the design, selection or procurement of these systems and their components that balances computer system performance with cost. Thus, systems such as: processors,operating systems, computer architecture and organizations, network configurations,languages and databases are all subjects of investigation and optimization [1]. There are many different ways to measure the performance of a network, as each network is different in nature and design. There are two basic areas of performance analysis one measurement and analysis of systems and the other modelling.. For measurement and analysis we need statistics and data analysis methods, mathematical expressions and measures and experimental design tools, while for modelling we need probability, queuing theory, simulation techniques and state transition diagrams that allow the network planner to analyze how the network will perform in each state, ensuring that the network will be optimally designed [2].

An ATM network is made up of an ATM switch and ATM endpoints. Examples of ATM endpoints are workstations, routers, digital service units (DSUs), LAN switches, and video coder-decoders (CODECs). In the ATM switch, the ATM cells have to be transported from an inlet to one or more outlets. In principle, an ATM switch shall perform the following two basic functions: switching and queuing [3]. Queuing is the primary factor determining the performance of a switch. The queuing can be simply classified into four categories. These are determined by the physical location of the queues: at the inputs, the outputs, inputs and outputs, or shared queuing. In output queuing solution, every output port must be able to accept cells from every input port simultaneously during one time slot [4]. However, only a single cell may be served by an output port, thus causing possible output contetion. The possible output contetion is solved by queues which are located at each output of the switch fabric and allow it to store multiple cells which may arrive during one time slot.

Recently, with the advent of ATM-based multiservice networks, a renewed interest in discrete-time models becomes apparent. Usually, the quantities studied in these investigations are the buffer occupancy, system occupancy, and the delays (waiting times) experienced by the packets in the buffer. At many instances, discrete-time multiserver queues have been successfully used in the performance evaluation of computer and communication systems in various contexts, such as TDM [5], voice-data integration [6], deflection routing [7] and more recently, ATM-switching technology [8].

In [9], [10], [11] the delays in various multiserver systems are investigated, and in [12] packet delays in single server systems with various types of first order Markovian correlated arrival processes are studied. Also, the discrete-time queues with or without server interruptions have received great attention in the scientific literature. Both single server [13] and multiserver [14] have been analysed. One of the earlier papers is the analysis by [15], who investigated a finite multiserver queue without server interruptions. Most authors, however, analyse an infinite system [16]. Some make specific assumptions about the arrival process [17] while general independent arrivals are considered elsewhere [18]. Several models have been used to describe the interruption process. Models with correlation can be found in [19], [20].

While both analytical and numerical results have been obtained on many occasions with respect to performance measures related to the buffer contents distribution [21], [12] for the case of multiserver [14], [22] as well as for the single server case [23] the derivation of delay characteristics has received much less attention in the past. More specifically, analytic results have been limited primarily to the single server case, whereas the multiple server case has mainly been investigated by means of numerical methods.

In most cases, analytic results concerning the delay are limited to the mean value of the packet delay [16] which can be obtained by means of Little's Theorem, although other performance measures related to the delay, such as the variance and the tail distribution, are equally important for a wide range of applications, including system design in ATM-based B-ISDN networks [24]. In a number of papers [19] delays are analysed for single server queueing systems where packets arrive according to an correlated arrival process. The rest of this paper is organized as follows. Next section introduces the mathematical model assumptions. In sections 3 and 4, we calculate the PGF of system occupancy, the expected value of system occupancy, and introduce some numerical examples. Section 5, 6, and 7 are for calculation of PGFs of unfinished work, queueing time, and waiting time respectively. Section 8 gives expectations, numerical examples and final results. Last section is for conclusion.

All the computations in this paper have been verified with *Mathematica* [25].

2 MODEL ASSUMPTIONS

First of all, it is assumed that the switch operates in a discrete time manner. That is, the time axis is divided into slots, each equal to the transmission time of one packet. Non-negative integers $k = 0, 1, \cdots$, are assigned to the individual slot boundaries. Time interval $[k, k+1)$ is referred to as slot $k+1$. Furthermore, most of the quantities considered in the article are RVs, all of them non-negative and integral valued.

The switch has the following assumptions, largely reflected by Figure 1. There are N input ports and N buffered output ports, each with c channels. The arrivals at the input ports are Bernoulli processes. That is, monitoring an arbitrary input port, every slot a packet will arrive with probability r and will not arrive with probability $\overline{r} = 1 - r$. This implies that the arrival rate at any port is r packets per slot. Also, it implies that the packet interarrival time is geometrically distributed with parameter r.

A packet that has arrived at an input port is routed in the *same* slot to its requested output (a channel). The probability that the packet requests a particular output port i is

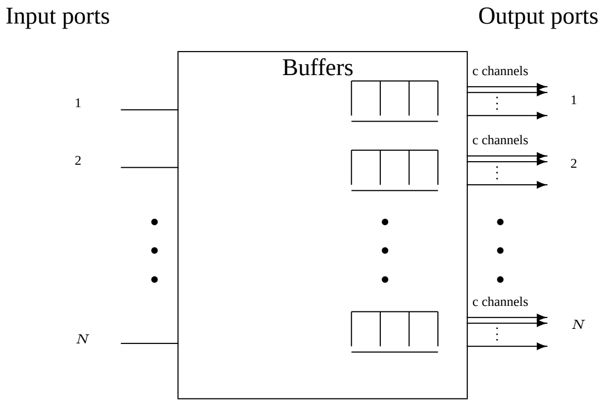

Fig. 1. An $N \times N$ output buffered space division switch with c channels at each output port.

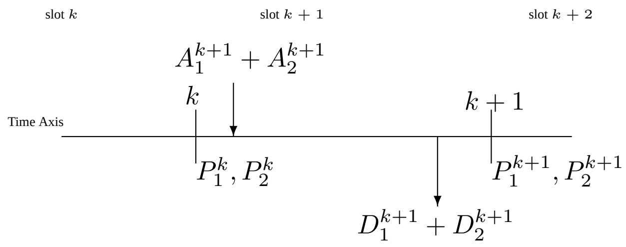

Fig. 2. The system occupancy in two successive slots

$\frac{1}{cN}$, for all $i = 1, 2, \cdots, N$. The packet request is independent of the input port it arrives into. It is clear that the probability of having a packet arrival in the output buffer, coming from any given input link, during any given slot, is equal to $\frac{r}{cN}$. As the arrival processes on different input links are independent, it is thus clear that the total number of packets arrivals during each slot has a binomial distribution with parameters N and $\frac{r}{cN}$, i.e. with expected value $\frac{r}{c}$. Also, the total numbers of packets arrivals in the buffer during different slots are independent RVs. So, the arrival process can be modelled by a binomial distribution

It can be seen from the above assumptions that the traffic into the switch, out of the switch, and inside the switch is uniform. As a consequence, modelling the switch reduces to modelling an arbitrarily 'tagged' output port. Unless otherwise indicated, the word 'the port' in the sequel will refer to this tagged output port. Buffered, the port can be conveniently modelled as a queueing system. In every slot, a batch of packets arrives at the port from the input ports. These packets wait in the buffer until they are served out of the port, hence out of the switch.

The port can be looked upon as made up of two parts: the buffer and the c channels. The buffer is of infinite capacity and is used to host packets arriving from the input ports. The time the packet spends in the buffer is called *queueing* time. The c channels are used to host the departing packets. Physically, they could be registers. The time the packet spends in the server (channel) is called *service* time. The sum of the queueing time and the service time is called *waiting* time. If a packet arrives into the port, it enters either service, if there is an available channel for it, or a queue, if there is no available channel for it. In all cases, entering service or queue takes place exactly at the beginning of the slot following the arrival slot. This implies that a packet is *not* considered to be in the port in its arrival slot. If a packet is being served during a particular slot, it may finish service only at the end of that slot. As indicated in Figure 2, in each slot k, $A^k = 0, 1, \cdots, N$, packets arrive from the input ports into the port. The A^k are independent and identically distributed (iid) RVs. Note that due to the uniformity of the traffic, the arrival rate into the port, r, is the same as those into an input port.

Let a_i and $A(z)$ be the *common* distribution and *common* PGF of the A^k. That is

$$a_i \triangleq \Pr\left[A^k = i\right], \qquad k = 0, 1, \cdots,$$

and

$$A(z) \triangleq \sum_{i=0}^{N} a_i z^i = E\left[z^{A^k}\right], \qquad |z| \leq 1,$$

where the notation $E[\cdot]$ denotes the expectation of the RV between brackets. It can be easily shown that

$$a_i = \binom{N}{i}\left(\frac{r}{cN}\right)^i \left(1 - \frac{r}{cN}\right)^{N-i}, \qquad i = 0, 1, ..., N,$$

$$A(z) = \left(1 - \frac{r}{cN} + \frac{r}{cN} z\right)^N. \tag{1}$$

$$A'(1) = \frac{r}{c}, \qquad A''(1) = \left(\frac{r}{c}\right)^2 \left(\frac{N-1}{N}\right). \tag{2}$$

Let D^{k+1}be the number of class 1 packets that will leave the port at the end of slot $k+1$ with distribution d_i. In each slot a packet leaves a server with probability s or does not leave with probability $\overline{s}$ then the number of departing packets per slot follows a binomial distribution. This implies that the service times of packets are geometrically distributed with parameters s.

Let $X^k = 1, 2, \ldots$, be the service time of the packet that arrives into the port in slot k. It is clear that the X^k are iid. Let x_i and $X(z)$ be the common distribution and common PGF of X^k. From the assumptions, it can be shown that $x_i = s\overline{s}^{i-1}$, and that

$$X(z) = \frac{sz}{1 - \overline{s}z}. \tag{3}$$

3 SYSTEM OCCUPANCY

Let $P^k = 0, 1, \cdots$, be a RV denoting the port occupancy in slot k, with distribution p_i^k and PGF $P^k(z)$. That is

$$P^k(z) = \sum_{i=0}^{\infty} p_i^k z^i = E\left[z^{P^k}\right]. \tag{4}$$

Looking at the existing packets P^k at slot k as independent trials where in each slot a packet leaves a server with probability s or does not leave with probability $\overline{s}$, then the number of packets served by the end of slot $k+1$ (D^{k+1}) is going to depend on P^k, with the following conditional distribution

$$\begin{aligned} &\Pr\left[D^{k+1} = i \,\middle|\, P^k = j\right] \\ =& \begin{cases} \binom{j}{i} s^i \overline{s}^{j-i} & \text{if } j < c, i \le j \\ \binom{c}{i} s^i \overline{s}^{c-i} & \text{if } j \ge c, i \le c \\ 0 & \text{otherwise} \end{cases}. \end{aligned} \tag{5}$$

The evaluation of the system occupancy in two successive slots k, $k+1$ can be described, in view of figure 2, by the following RV equation

$$P^{k+1} = P^k - D^{k+1} + A^{k+1}. \tag{6}$$

Using (6) and all the possible combinations of P^k, D^{k+1} in (4), then applying (5), and after some manipulation, we get

$$\begin{aligned} & P^{k+1}(z) \\ =\ & E\left[z^{P^k - D^{k+1} + A^{k+1}}\right] \\ =\ & E\left[z^{A^{k+1}} z^{P^k - D^{k+1}}\right] \\ =\ & A(z) \sum_{j=0}^{c-1} \left[(s + \overline{s}z)^j - (s + \overline{s}z)^c z^{j-c}\right] p_j^k \\ & + A(z)(s + \overline{s}z)^c z^{-c} P^k(z). \end{aligned} \tag{7}$$

At the steady state,

$$\begin{aligned} P(z) \ =\ & \frac{A(z) z^c}{z^c - A(z)(s + \overline{s}z)^c} \times \\ & \sum_{j=0}^{c-1} \left[(s + \overline{s}z)^j - (s + \overline{s}z)^c z^{j-c}\right] p_j. \end{aligned} \tag{8}$$

To define the PGF $P(z)$, given by (8) completely, we have to define the unknown probabilities $p_j, j = 0, 1, \ldots, c-1$. We will apply Rouche's Theorem to the denominator of (8), and proceed to consider points ξ_n, within the unit disk for which the denominator of (8) is equal to zero. Then

$$\xi_n^c = A(\xi_n)\,(s + \overline{s}\xi_n)^c. \tag{9}$$

For any such point ξ_n, $|\xi_n| \le 1, n = 0, 1, 2, \ldots, c-1$ we must have a simple zero. It can be shown that the function $P(z)$ is bounded within the unit disk $|z| \le 1$ therefore, both the numerator and the denominator of (8) must be zero for the same values of z. Then substituting with the zeros ξ_n in the numerator of (8) and provided that $A(\xi_n) \ne 0$, we get

$$\begin{aligned} & \xi_n^c \sum_{j=0}^{c-1} \left[(s + \overline{s}\xi_n)^j - (s + \overline{s}\xi_n)^c \xi_n^{j-c}\right] p_j \\ =\ & 0 \\ , n =\ & 1, 2, \ldots, c-1, \end{aligned} \tag{10}$$

which are $c-1$ equations in c unknowns. The equation number c needed to solve for the unknown probabilities comes from the normalization condition $P(1) = 1$ but let us first write (8), in the form

$$P(z) = \frac{A(z)\Phi(z)}{z^c - A(z)(s + \overline{s}z)^c}, \tag{11}$$

$$\begin{aligned} & \Phi(z) = z^c \times \\ & \sum_{j=0}^{c-1} \left[(s + \overline{s}z)^j - (s + \overline{s}z)^c z^{j-c}\right] p_j, \\ & \Phi(1) = 0. \end{aligned} \tag{12}$$

Taking the first derivative of (12) at $z = 1$, thus

$$\Phi'(1) = s \sum_{j=0}^{c-1} [c - j]\, p_j. \tag{13}$$

Applying the normalization condition to (11) after applying L'Hospital's rule, we get

$$\Phi'(1) = sc - \frac{r}{c}, \tag{14}$$

where $A'(1) = \frac{r}{c}$. Now, equating (13) and (14), hence

$$s \sum_{j=0}^{c-1} [c - j]\, p_j = sc - \frac{r}{c}. \tag{15}$$

Equations (10) and (15) can be written explicitly, as the following c equations

$$
\begin{aligned}
&\left[\xi_1^c-(s+\bar{s}\xi_1)^c\right]p_0\\
&+\left[\xi_1^c(s+\bar{s}\xi_1)-(s+\bar{s}\xi_1)^c\xi_1^1\right]p_1\\
&+\cdots+\left[\begin{array}{c}\xi_1^c(s+\bar{s}\xi_1)^{c-1}\\-(s+\bar{s}\xi_1)^c\xi_1^{c-1}\end{array}\right]p_{c-1}=0\\
&\left[\xi_2^c-(s+\bar{s}\xi_2)^c\right]p_0\\
&+\left[\xi_2^c(s+\bar{s}\xi_2)-(s+\bar{s}\xi_2)^c\xi_2^1\right]p_1\\
&+\cdots+\left[\begin{array}{c}\xi_2^c(s+\bar{s}\xi_2)^{c-1}\\-(s+\bar{s}\xi_2)^c\xi_2^{c-1}\end{array}\right]p_{c-1}=0\\
&\vdots\\
&\left[\xi_{c-1}^c-(s+\bar{s}\xi_{c-1})^c\right]p_0\\
&+\left[\begin{array}{c}\xi_{c-1}^c(s+\bar{s}\xi_{c-1})\\-(s+\bar{s}\xi_{c-1})^c\xi_{c-1}^1\end{array}\right]p_1\\
&+\cdots+\left[\begin{array}{c}\xi_{c-1}^c(s+\bar{s}\xi_{c-1})^{c-1}\\-(s+\bar{s}\xi_{c-1})^c\xi_{c-1}^{c-1}\end{array}\right]p_{c-1}\\
&=0\\
&scp_0+s[c-1]p_1+\cdots+p_{c-1}=sc-\frac{r}{c}\,,
\end{aligned}
\tag{16}
$$

which are c equations in the c unknowns $p_j,\ j=0,1,\ldots,c-1$, and can be solved numerically to find the unknown probabilities, as soon as the value of c is specified. Now, to continue our calculation and measurement we have to define the value of c, by taking some special cases for c.

Case 1 : PGF of the System Occupancy When $c=1$.Substituting for $c=1$ in (8), thus

$$P(z)=\frac{A(z)\left[z-(s+\bar{s}z)\right]p_0}{z-A(z)(s+\bar{s}z)}. \tag{17}$$

After finding p_0

$$P(z)=\left(1-\frac{r}{s}\right)\frac{A(z)\left[z-(s+\bar{s}z)\right]}{z-A(z)(s+\bar{s}z)}. \tag{18}$$

Case 2 : PGF of the System Occupancy When $c=2$.Substituting for $c=2$ in (8), thus

$$
\begin{aligned}
&P(z)=\frac{1}{z^2-A(z)(s+\bar{s}z)^2}\times\\
&\Big(A(z)\left[z^2-(s+\bar{s}z)^2\right]p_0\\
&+A(z)\left[z^2(s+\bar{s}z)-(s+\bar{s}z)^2z\right]p_1\Big)
\end{aligned}
\tag{19}
$$

After calculating p_0 and p_1, we get

$$
\begin{aligned}
P(z)=&\frac{A(z)}{z^2-A(z)(s+\bar{s}z)^2}\times\\
&\Big\{\quad\left[z^2-(s+\bar{s}z)^2\right]\times\\
&\left(-\tfrac{\beta_1(2s-r)}{\alpha_1\beta_2-\alpha_2\beta_1}\right)\\
&+\left[z^2(s+\bar{s}z)-(s+\bar{s}z)^2z\right]\times\\
&\left(\tfrac{\alpha_1(2s-r)}{\alpha_1\beta_2-\alpha_2\beta_1}\right)\quad\Big\}\,,
\end{aligned}
\tag{20}
$$

$$
\begin{aligned}
\alpha_1&=\left[\xi_1^2-(s+\bar{s}\xi_1)^2\right],\quad \alpha_2=2s\\
\beta_1&=\left[\begin{array}{c}\xi_1^2(s+\bar{s}\xi_1)\\-(s+\bar{s}\xi_1)^2\xi_1\end{array}\right],\beta_2=s.
\end{aligned}
\tag{21}
$$

4 EXPECTATIONS AND NUMERICAL EXAMPLES

4.1 Expected Value of the System Occupancy When $c=1$

To ease the calculation of the expected value of the system occupancy let us write (18), in the form

$$P(z)=\left(1-\frac{r}{s}\right)\frac{\Phi(z)}{\theta(z)}, \tag{22}$$

$$
\begin{aligned}
\Phi(z)&=A(z)\left[z-(s+\bar{s}z)\right],\\
\theta(z)&=z-A(z)(s+\bar{s}z).
\end{aligned}
\tag{23}
$$

Taking the first derivative of (22) at $z=1$, using L'Hospital's rule twice, and after some manipulation, we will get

$$
\begin{aligned}
E[P]&=\frac{A''(1)+2r-2r^2}{2(s-r)}\\
&=\frac{1}{2(s-r)}\left\{r^2\frac{(N-1)}{N}+2r-2r^2\right\}.
\end{aligned}
\tag{24}
$$

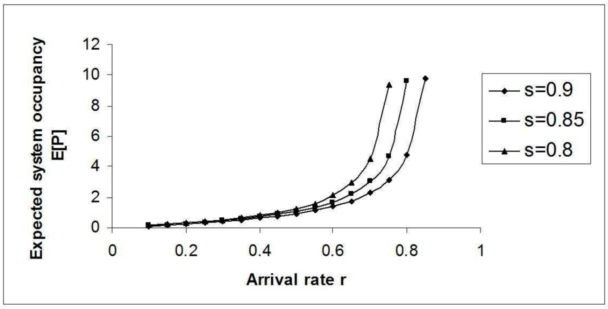

Fig. 3. Expected system occupancy E[P] vs the arrival rate r, $N=8, c=1$

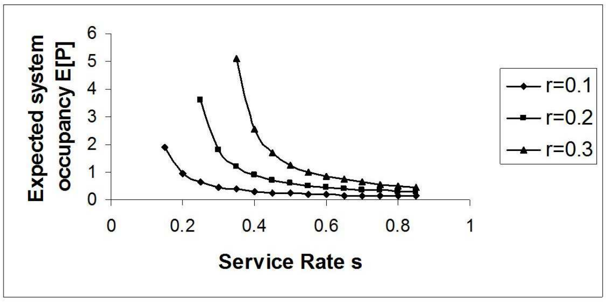

Fig. 4. Expected system occupancy E[P] vs the service rate s, $N=8, c=1$

4.2 Expected Value of the System Occupancy When $c=2$

Taking the first derivative of (20) at $z=1$, we can show that

$$\begin{aligned} E[P] &= \frac{1}{(4s-r)} \times \\ &\quad \begin{bmatrix} 2s[2-s](\sigma_1+\sigma_2) \\ +sr(2\sigma_1+\sigma_2) \end{bmatrix} \\ &\quad -\frac{(2\sigma_1+\sigma_2)s}{(4s-r)^2} \times \\ &\quad \left[4s[2-s]-4r\overline{s}-2A''(1)\right] \\ &= \frac{[2s[2-s](\sigma_1+\sigma_2)+rs(2\sigma_1+\sigma_2)]}{(4s-r)} \\ &\quad -\frac{(2\sigma_1+\sigma_2)s}{(4s-r)^2} \times \\ &\quad \left[4s[2-s]-4r\overline{s}-\frac{r^2}{2}\frac{(N-1)}{N}\right], \end{aligned} \tag{25}$$

where σ_1, σ_2 are given by

$$\begin{aligned} \sigma_1 &= -\frac{\beta_1(4s-r)}{2(\alpha_1\beta_2-\alpha_2\beta_1)}, \\ \sigma_2 &= \frac{\alpha_1(4s-r)}{2(\alpha_1\beta_2-\alpha_2\beta_1)}. \end{aligned} \tag{26}$$

Now we introduce many numerical examples which verify the results. From figure (3) we note that, the expected value of the system occupancy for $c=1$ increases when increasing the arrival rate of the packets for a given values of N, s. In the same figure, it is easy to note that the expected value of the system occupancy decreases also when the service rate of the system increases. Figure (4) gives us the relation between the expected value of the system occupancy and the service rate of the system for $r=0.1, 0.2, 0.3$, and $N=8$. As should be, the system occupancy decreases as the service rate increases and vice versa.

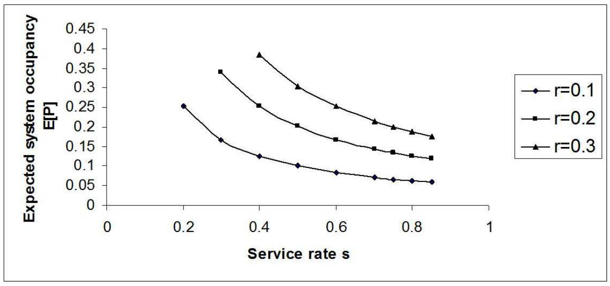

Fig. 5. Expected system occupancy $E[p]$ *vs the service rate* s, $N=8, c=2$

Again, we will focus on the expected value of the system occupancy but for $c=2$. In figure (5), the expected value of the system occupancy is plotted as a function in the service rate for arrival rate $r = 0.1$, 0.2, 0.3, and $N=8$. As clear in the figure, the expected value of the system occupancy increases when the service rate decreases for a given arrival rate. Moreover, as the arrival rate increases and the service rate increases, this gives the best $E[P]$.

As a final conclusion of the results, we introduce figures that show a comparison between the system measurements for both cases when $c=1, c=2$. It is very clear in figures (6) and (7) that the system measurements are improved by using multiserver ($c=2$) rather than single server, as we should expect.

5 UNFINISHED WORK

Before deriving the PGF of the unfinished work, let us define the following:

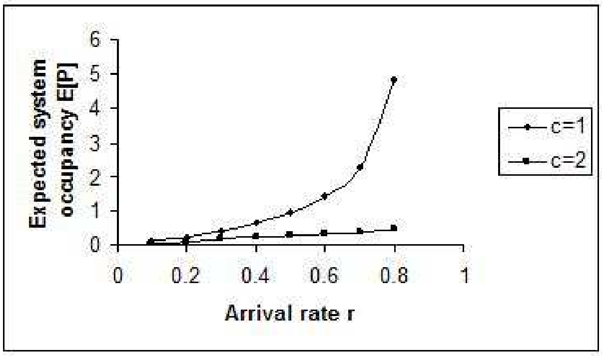

Fig. 6. Expected system occupancy $E[p]$ vs the arrival rate r, $N = 8, s = 0.9$

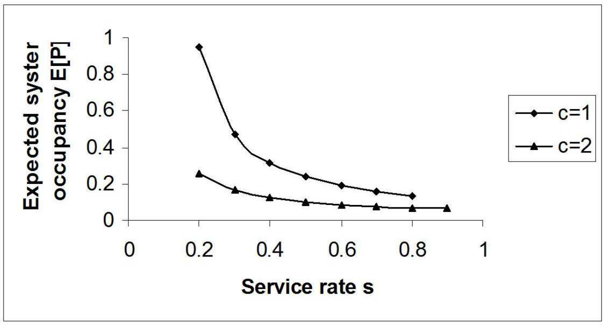

Fig. 7. Expected system occupancy $E[p]$ vs the service rate s, $N = 8, r = 0.1$

1. Let U^k be a RV representing the unfinished work at the end of slot k with distribution u_i^k and PGF $U^k(z)$. That is

$$U^k(z) = \sum_{i=0}^{\infty} u_i^k z^i = E\left[z^{U^k}\right]. \tag{27}$$

2. Let X be a RV representing the service time of an arbitrary packet, i.e. it represents the number of slots that an arbitrary packet spends in service, with geometric distribution x_i and PGF $X(z)$. That is

$$X(z) = \frac{sz}{1-\overline{s}z}, \text{ and} \tag{28}$$

$$E[X] = X'(1) = \frac{1}{s}, \quad X''(1) = \frac{2\overline{s}}{s^2}. \tag{29}$$

3. Let $X^{(i)}$ be a RV representing the service time of the ith packet of all the packets arrived in the same slot i.e. it represents the number of slots that this packet spends in the server, note that the RV $X^{(i)}$ has the same distribution as X.Let G^k be a RV representing the service times of the packets arriving together in slot k, with distribution g_i and PGF $G(z)$. That is

$$G(z) = E\left[z^{\sum_{i=1}^{A^k} X^{(i)}}\right] = A\left(X(z)\right). \tag{30}$$

4. Let L^{k+1} be a RV representing the amount of work done (measured in slots) during slot $k+1$, with distribution l_j . Clearly L^{k+1} depends on U^k with the following conditional distribution

$$l_{j|U^k=i} == \begin{cases} 1 & \text{if } i < c, j = i \\ & \text{if } i \geq c, j = c \\ 0 & \text{otherwise} \end{cases}. \tag{31}$$

The unfinished work in two successive slots k, $k+1$ can be described by the following RV equation

$$U^{k+1} = U^k - L^{k+1} + G^{k+1}. \tag{32}$$

Using (32) and all the possible combinations of U^k, L^{k+1} in (27) then applying (31), and in view of probability theorems, after some manipulation, we reach

$$\begin{aligned} U^{k+1}(z) \;&=\; A\left(X(z)\right) z^{-c} \times \\ &\sum_{i=0}^{c-1} \left[z^c - z^i\right] u_i^k \\ &+ A\left(X(z)\right) z^{-c} U^k(z). \end{aligned} \tag{33}$$

Taking the limit of (33) at the steady state, as $k \to \infty$, and solving for $U(z)$, so

$$U(z) = \frac{A\left(X(z)\right) \sum_{i=0}^{c-1} \left[z^c - z^i\right] u_i}{z^c - A\left(X(z)\right)}. \tag{34}$$

For defining unknown probabilities u_i, $0 \leq i \leq c-1$ consider points $\xi_i, i = 1, 2, \ldots, c-1$ within the unit disk for which the denominator vanishes. So $\xi_j^c = A\left(X(\xi_j)\right)$ .Then

$$\sum_{i=0}^{c-1} \left[{\xi_j}^c - \xi_j^i\right] u_i = 0, \;\; j = 1, 2, \ldots, c-1, \tag{35}$$

Also, it can be shown that

$$\sum_{i=0}^{c-1} \left[c - i\right] u_i = c - \frac{r}{cs}. \tag{36}$$

Equations (35) and (36) are c equations in the c unknowns u_i, $i = 0, 1, \ldots, c-1$ and they can be solved to find the unknown probabilities. Then

$$U(z) \;=\; \left(c - \frac{r}{cs}\right) \frac{A\left(X(z)\right)(z-1)}{z^c - A\left(X(z)\right)} \prod_{i=1}^{c-1} \frac{(z-\xi_i)}{(1-\xi_i)}. \tag{37}$$

6 QUEUING TIME

To derive the PGF of the queueing time let us focus on an arbitrary packet that arrived in slot k. The queuing time of this tagged packet consists of two parts. The unfinished work of the system at the end of slot $k-1$ and The service time for all the packets arriving in slot k but served before the tagged packet. Let us define the following: Let Q^k be a RV representing the queueing time of the tagged packet arriving at slot k, i.e. it represents the number of slots that this packet spends in the queue, with distribution q_i^k and PGF $Q^k(z)$. Let F^k be a RV representing the number of packets arriving in slot k but served before the tagged packet with distribution f_i^k and PGF $F^k(z)$. Let V^k be a RV representing the service time of the packets arriving in the same slot as the tagged packet but served before it, That is

$$F^k(z) = \sum_{i=0}^{\infty} f_i^k z^i = E\left[z^{F^k}\right]. \tag{38}$$

$$V^k = \sum_{i=1}^{F^k} X^{(i)}, \tag{39}$$

where $X^{(i)}$, represents the service time of the ith packet of the packets served before the tagged packet. Let L^k be a RV representing the amount of work done in slot k with distribution l_n.From the previous assumptions we can conclude that the queueing time can be expressed by, the following RV equation

$$Q^k = U^{k-1} - L^k + V^k. \text{ Then} \tag{40}$$

$$Q^k(z) = V^k(z) E\left[z^{U^{k-1}-L^k}\right]. \tag{41}$$

Computing the value of the factor $E\left[z^{U^{k-1}-L^k}\right]$ in (41), we realize that

$$Q^k(z) = V^k(z) z^{-c} \sum_{i=0}^{c-1} \left[z^c - z^i\right] u_i^k + V^k(z) z^{-c} U^k(z). \tag{42}$$

Using (39) we can derive an expression for $V^k(z)$, it is given by

$$V^k(z) = F^k\left(X(z)\right). \tag{43}$$

Substituting $V^k(z)$ from (43) in (42) and taking the limit of it at the steady state, as $k \to \infty$, thus we will have (42) in the form

$$Q(z) = F\left(X(z)\right) z^{-c} \left[\sum_{i=0}^{c-1} \left[z^c - z^i\right] u_i + U(z)\right]. \text{ Then} \tag{44}$$

$$\begin{aligned} Q(z) &= F\left(X(z)\right) z^{-c} \\ &\times \left[\left(c - \frac{r}{cs}\right)(z-1) \prod_{i=1}^{c-1} \frac{(z-\xi_i)}{(1-\xi_i)} + U(z)\right]. \end{aligned} \tag{45}$$

Now, substituting for $U(z)$ from (37) in (45), and after some manipulation, thus

$$\begin{aligned} Q(z) &= F\left(X(z)\right) \\ &\times \left(c - \frac{r}{cs}\right) \frac{(z-1)}{z^c - A\left(X(z)\right)} \prod_{i=1}^{c-1} \frac{(z-\xi_i)}{(1-\xi_i)}. \end{aligned} \tag{46}$$

After the derivation of $F(z)$, it is given by

$$F(z) = \sum_{i=0}^{\infty} f_i z^i = \frac{c}{r} \frac{A(z) - 1}{(z-1)}. \tag{47}$$

Substituting for $z = X(z)$ in (47), thus

$$F\left(X(z)\right) = \frac{c\left(A\left(X(z)\right) - 1\right)}{r\left(X(z) - 1\right)}. \tag{48}$$

Substituting for $F\left(X(z)\right)$ from (48) in (46), therefore

$$\begin{aligned} & Q(z) \\ = \; & \frac{c\left(A\left(X(z)\right) - 1\right)}{r\left(X(z) - 1\right)} \\ & \times \left(c - \frac{r}{cs}\right) \frac{(z-1)}{z^c - A\left(X(z)\right)} \prod_{i=1}^{c-1} \frac{(z-\xi_i)}{(1-\xi_i)}, \end{aligned} \tag{49}$$

where $X(z)$ is given by (28).

7 WAITING TIME

Let us define the RV W^k to represent the waiting time of an arbitrary packet "tagged packet" at slot k with distribution w_i^k and PGF $W^k(z)$. That is

$$W^k(z) = \sum_{i=0}^{\infty} w_i^k z^i = E\left[z^{W^k}\right]. \tag{50}$$

To derive an explicit formula for $W^k(z)$ we note that the waiting time of the tagged packets in the system at slot k consists of the queueing time of the packet added to the service time of this packet, so they are related by the relation

$$W^k = Q^k + X^k, \tag{51}$$

$$W^k(z) = Q^k(z) X^k(z), \tag{52}$$

where X^k denotes the service time of the tagged packet. At the steady state, as $k \to \infty$, (52) will be written in the form

$$W(z) = Q(z) X(z). \tag{53}$$

Substituting for $Q(z)$ from (49) in (53), we finally get

$$\begin{aligned} & W(z) \quad (54) \\ = \; & \frac{c\left(A\left(X(z)\right)-1\right)}{r\left(X(z)-1\right)} \\ & \times\left(c-\frac{r}{cs}\right)\frac{X(z)(z-1)}{z^{c}-A\left(X(z)\right)}\prod_{i=1}^{c-1}\frac{(z-\xi_i)}{(1-\xi_i)}, \end{aligned}$$

where $X(z)$, is given by (28).

8 EXPECTATIONS AND NUMERICAL RESULTS

8.1 Expected Value of the Unfinished Work

The expected value of the unfinished work of the system at the steady state can be found from (37) by evaluating the first derivative of $U(z)$ at $z = 1$. It can be shown that

$$\begin{aligned} & E\left[U\right] \\ = \; & \frac{1}{(2cs-r)} \\ & \times\left(r\left(c-\frac{r}{2s}\right)-cs(c-1)+\frac{r\overline{s}}{s}+\frac{A''(1)}{s}\right) \\ & +\sum_{j=1}^{c-1}\frac{1}{1-\xi_j}. \quad (55) \end{aligned}$$

8.2 Expected Value of the Waiting Time

To obtain an expression for the expected value of waiting time of a packet we will apply Little's Theorem [26]. When $c = 1$, from (54), we will have

$$E[W]=\frac{1}{r}\left(\frac{A''(1)+2r-2r^2}{2\left(s-r\right)}\right), \quad (56)$$

and when $c = 2$, using (54), we will have

$$\begin{aligned} & E[W] \\ = \; & \frac{2}{r}\left\{\begin{array}{c} \frac{[2s[2-s](\sigma_1+\sigma_2)+sr(2\sigma_1+\sigma_2)]}{(4s-r)} \\ -\frac{(2\sigma_1+\sigma_2)s\left[4s[2-s]-4r\overline{s}-2A''(1)\right]}{(4s-r)^2} \end{array}\right\}, \quad (57) \end{aligned}$$

$$\begin{aligned} \sigma_1 &= -\frac{\beta_1\left(4s-r\right)}{2\left(\alpha_1\beta_2-\alpha_2\beta_1\right)}, \quad (58) \\ \sigma_2 &= \frac{\alpha_1\left(4s-r\right)}{2\left(\alpha_1\beta_2-\alpha_2\beta_1\right)}. \end{aligned}$$

Case 1: $c = 1$. Calculating $A''(1)$ and Substituting in (56), hence

$$E[W]=\frac{1}{2r\left(s-r\right)}\left\{r^2\frac{(N-1)}{N}+2r-2r^2\right\}.$$

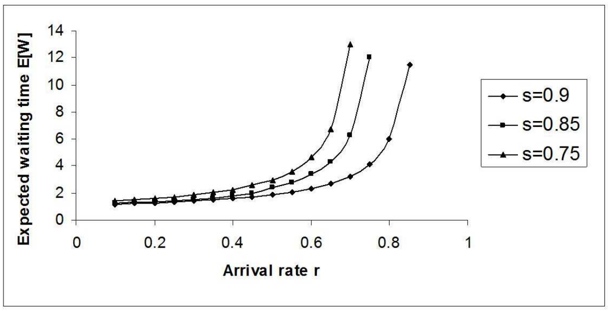

Fig. 8. Expected waiting time $E[W]$ vs the arrival rate r, $N = 8, c = 1$

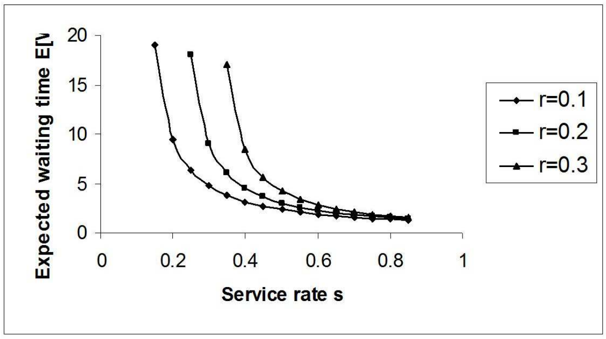

Fig. 9. Expected waiting time $E[W]$ vs the service rate s, $N = 8, c = 1$

As obvious in the figures (8) and (9), the expected value of the waiting time of a packet increases when the arrival rate of packets increases but it decreases when the service rate of the packets increases as it is clear in figure (9).

Case 2 : $c = 2$

Calculating $A''(1)$ and substituting in (56), hence

$$\begin{aligned} & E[W] \\ = \; & \frac{2}{r}\times \\ & \left\{\begin{array}{c} \frac{\left[2s\left[2-s\right]\left(\sigma_1+\sigma_2\right)+rs\left(2\sigma_1+\sigma_2\right)\right]}{(4s-r)} \\ -\frac{(2\sigma_1+\sigma_2)s}{(4s-r)^2}\times \\ \left[4s\left[2-s\right]-4r\overline{s}-\frac{r^2}{2}\frac{(N-1)}{N}\right] \end{array}\right\}, \end{aligned}$$

where σ_1, σ_2 are given by (58). Figure (10) shows that the expected value of the waiting time increases due to the increase in the arrival rate for a given service rate and as the service rate gets lower the expected value of the waiting time gets bigger, as expected.

8.3 Final Results

Figures (11) and (12) give comparisons between the expected waiting times for both cases when $c = 1, c = 2$. It

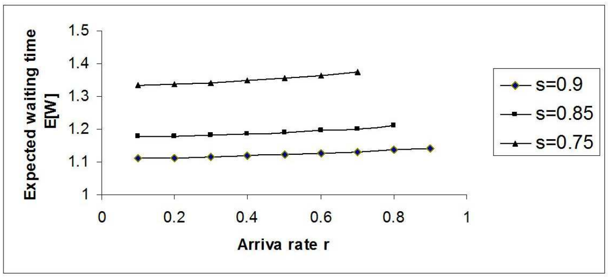

Fig. 10. Expected waiting time $E[W]$ vs the arrival rate r, $N = 8, c = 2$

is very clear in both figures how the expected waiting times are improved by using multiserver ($c = 2$) rather than one server as we should expect.

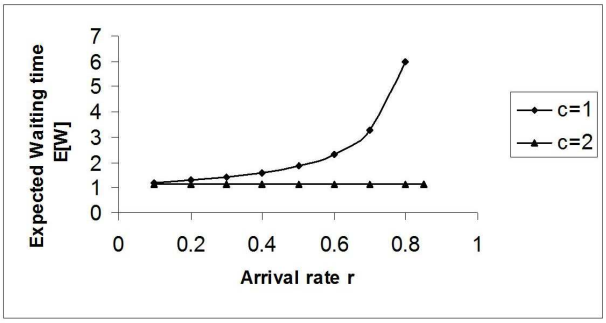

Fig. 11. Expected waiting time $E[W]$ vs the arrival rate r, $N = 8, s = 0.9$

9 CONCLUSION

In this article the main contributions have been using multiserver case and geometric service time. The switch as a whole has been modelled as a discrete time, multiserver server, batch arrival, infinite buffer queueing system, with geometric service time. We have obtained PGFs for four performance measures: occupancy, unfinished work, queuing time and waiting time. The PGFs have been used to derive the corresponding expectations. The results of the analysis have been verified in many ways. First, they have been used to generate the results of some previous analyses as special cases. Second, they obviously seem to preserve classical queueing relations (e.g. the famous Little's formula). Last, they have been shown to generate intuitively acceptable graphs, when translated into numerical values. All the computations in this paper have been verified with *Mathematica*

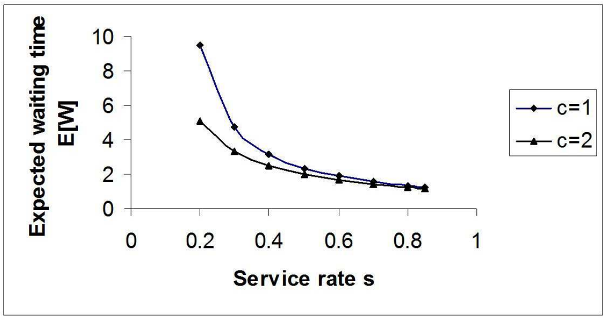

Fig. 12. Expected waiting time $E[W]$ vs the service rate s, $N = 8, r = 0.1$

REFERENCES

[1] B. R. Haverkort, *Performance of Computer Communication Systems*. J. Wiley and Sons, 1998.

[2] e. a. G. Bolch, *Queueing Networks and Markov Chains: Modelling and Performance Evaluation with Computer Science Applications*. J. Wiley and Sons, 1998.

[3] R. Y. Awdeh and H. T. Moufftah., "Survey of atm switch architectures," in *Computer Networks and ISDN Systems*, (Vol. 27), pp. 1567–1613, 1995.

[4] G. Kesidis, *ATM Network Performance*. Boston, MA: Kluwer Academic Publishers, 1996.

[5] W. Chu, "buffer behavior for poisson arrivals and multiple synchronous constant outputs," *IEEE Trans.*, vol. 19, pp. 530–534, 1970.

[6] S. Q. Li and J. Mark, "Performance trade-offs in an integrated voice/data services tdm system," *Performance Evaluation*, vol. 8, pp. 51–64, 1988.

[7] A. Bignell and T. Todd, "Analytic mode model for deflection routing networks," *Electron. Lett.*, vol. 26, pp. 67 – 69, 1990.

[8] J. C. H. M. Nassar and F. A. Nada, "Queueing analysis of an atm multichannel switch routing two-class multimedia traffic with two service rates," *IEICE Transactions*, vol. E86–A, no. 5, pp. 1505–1513, 2003.

[9] V. Goswami and U. Gupta, "Analyzing the discrete-time multiserver queues geom/geom/m queue with late and early arrivals," *Int. J. of Information & Managment Sciences*, vol. 9, no. 2, pp. 55–66, 1998.

[10] F. A. Nada, "Performance analysis of multiserver atm buffers routing multimedia traffics with geometric service time," *Automatika*, vol. 51, no. 3, pp. 293–301, 2010.

[11] F. A. Nada, "Unfinished work of queuing systems with two servers, geometric service time and priorities," *Int. J. of Algorithms, Computing, and Mathematics*, vol. 3, no. 4, pp. 67–77, 2010.

[12] S. Foss and N. Chernova, "On the optimality of the fcfs discipline in multiserver systems and queuing networks," *Siberian Mathematical Journal*, vol. 42, no. 2, pp. 372–385, 2001.

[13] H. Bruneel and K. Laevens, "Performance analysis of a discrete-time queue with a single server subjected to random interruptions," in *Book of abstracts, Eighth conference*

on quantitative methods for decision making, pp. 37–38, January 1994.

[14] H. Bruneel and K. Laevens, "Analytic derivation of delay characteristics for discrete-time multiserver queues with random server interruptions," in *Proceeding of the conference on applied probability in engineering*, p. 203, July 1993.

[15] G. Falin, "A multiserver retrial queue with a finite number of sources of primary calls," *Mathematical and Computer Modelling*, vol. 30, no. 3-4, pp. 33–49, 1999.

[16] H. Bruneel and B. Steyaert, "Finite versus infinite capacity multiserver atm buffers with correlated arrivals," in *Presentation at the fourth IFIP WG 6.3 Workshop on performance of communication systems*, Auguest 1999.

[17] H. Bruneel and B. Vinck, "Queuing analysis of discrete-time buffer systems with compound arrival process and variable capacity," in *Proceeding performance Tools'95+MMR'95*, pp. 41 – 55, September 1995.

[18] H. Bruneel and S. Wittevrongel, "Discrete-time atm queues with independent and correlated arrival streams," *Performance evaluation and applications of ATM networks*, pp. 387–412, 2000.

[19] S. W. H. Bruneel and S. D. Vuyst, "Mean buffer contents and mean packet delay in statistical multiplexers with correlated train arrivals," in *Proceedings of the seventh IFIP workshop on performance modelling and evaluation of ATM/IP networks*, pp. 1/12–12/12, June 1999.

[20] S. D. V. H. Bruneel and S. Wittevrongel, "Expected buffer contents and mean packet delay in statistical multiplexers with correlated train arrivals," in *Proceedings of the seventh IFIP Workshop on Performance Modelling and evaluation of ATM/IP networks, IFIP ATM '99*, pp. 1/12–12/12, 1999.

[21] H. Bruneel and B. Vinck, "Relationship between delay and buffer contents in atm queues," *Electronics Lett.*, vol. 31, no. 12, pp. 952–954, 1995.

[22] F. A. Nada, "A case study of multiserver atm buffers routing multimedia traffics with geometric service time," in *Proceeding of ICCEE*, pp. 453–457, 2009.

[23] H. Bruneel and B. Vinck, "General relationship between queue length and delay in discrete-time single-server queues," in *Proceedings of the cost 257 third managment committee meetting*, 1997.

[24] H. Bruneel and S. Wittevrongel, "Atm switches with nonindependent routing," *European Transactions on Telecommunications*, vol. 7, no. 6, pp. 575–579, 1996.

[25] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley; 2nd edition, 1991.

[26] M. Woodward, *Communication and Computer networks*. Los Alamitos, CA 90720-1264: IEEE Computer Society Press, 1994.

Fayza A. Nada Fayza A. Nada received her MSc and Ph. D. in computer science, both from College of Science, Suez Canal University, Egypt in 2002 and 2006, respectively. Her research interests include performance analysis, mathematical modelling of computer and communications systems, queuing theory, wireless networks, and using mobile IP protocols in multimedia traffic.

Medhat A. Rakha Medhat A. Rakha worked as A Professor of Pure Mathematics in the Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia - Egypt. He obtained his Ph. D. in Algebra and Special Functions from both Suez Canal University, Ismailia - Egypt and Texas A & M University, College Station, Texas - USA. Professor Rakha main research interest is the theory of hypergeometric and generalized hypergeometric functions. He is currently on sabbatical leave at the Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, Muscat, Oman.

AUTHORS' ADDRESSES

Asst. Prof. Fayza A. Nada,
Faculty of Computers and Information,
Suez Canal University,
Ismailia, Egypt
email: fayzanada.rus@cas.edu.om

Prof. Medhat A. Rakha
Permanent Address:
Department of Mathematics,
Faculty of Science,
Suez Canal University,
Ismailia, Egypt
email: medhat_ rakha@science.suez.edu.eg
Present Address:
Department of Mathematics and Statistics,
College of Science,
Sultan Qaboos University,
Muscat, Oman
email: medhat@squ.edu.om

Received: 2014-02-09
Accepted: 2015-01-09