Queueing analysis of an ATM multimedia multiplexer with non-pre-emptive priority

H. Nassar and H. Al Mahdi

Abstract: A queueing-theoretic analysis of an ATM multiplexer handling two-class multimedia traffic is described. Specifically, it assigns class-1 cells, constituting real-time traffic, high service priority, and class-2 cells, constituting non-real time traffic, low service priority. The priority discipline used is non-pre-emptive. For the purpose of the analysis, the multiplexer is modelled as a priority, discrete time, single-arrival, single-server queueing system with an infinite buffer and geometric service time. The model dynamics are extracted by a rather complex difference equation whose solution is sought using a generating function technique. This difference equation approach is a major contribution of the paper. Unlike the prevalent stochastic equation approach, it makes the physical details of the system present and visible during the analysis. Results are obtained for the multiplexer occupancy and cell waiting time. These results are verified analytically by producing from them some previously published results as special cases. They are also verified by applying them to numerical examples and obtaining intuitively acceptable values.

1 Introduction

An ATM multiplexer of the structure shown in Fig. 1 is used in digital communications systems to improve the

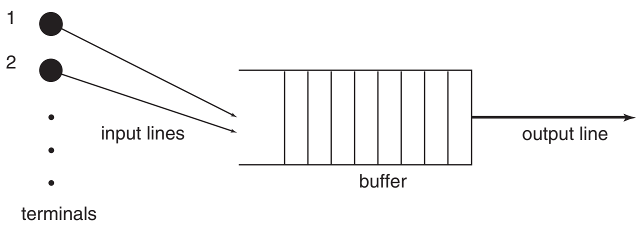

Fig. 1 *ATM multiplexer with cells waiting for service*

efficiency of communications lines. Cells (i.e. packets of equal length) arrive onto its input lines, are stored in a buffer, and then are transmitted one at a time onto the output line. The overall effect of the multiplexer is to concentrate bursty traffic, thereby improving the efficiency of the output line.

When the multiplexer is used in a multimedia environment, as is typically the case in a broadband integrated services digital network (B-ISDN) [1], the traffic it handles can be conveniently divided into two distinct types: real time, e.g. live audio and live video cells, and non-real time, e.g. file transfer cells. The quality of service (QoS) [1] of each type is different from that of the other. Specifically, real-time cells are loss-insensitive but delay-sensitive. This entails that cells of this type should be served so rapidly by the multiplexer that they get to their destinations in the shortest time possible, even if some of them are lost. On the other hand, non-real-time cells are delay-insensitive but loss-sensitive. This entails that cells of this type should be served so carefully that no cell is lost, even if the cells incur a longer delay in the multiplexer. The multiplexer, therefore, has to solve this problem.

A multiplexer equipped with a priority scheme can solve this QoS problem easily. Namely, it assigns real-time cells, henceforth called class-1 cells, high service priority, and non-real-time cells, henceforth called class-2 cells, low service priority. If a cell of class-1 and a cell of class-2 arrive at the multiplexer simultaneously, the multiplexer serves the former first.

If a priority scheme is adopted, one of two disciplines may be used, concerning what happens when a class-2 cell being served is treated upon the arrival of a class-1 cell. In the pre-emptive discipline, the arriving cell enters service immediately in the next slot, ejecting the class-2 cell back to the buffer. Later, when there are no more class-1 cells to serve, the ejected class-2 cell enters service again. In the non-pre-emptive discipline, on the other hand, the arriving cell waits until the class-2 cell finishes service and then takes its place. It can be seen that the pre-emptive discipline is favourable to class-1 cells, whereas the non-pre-emptive is favourable to class-2 cells.

Non-priority multiplexers handling single-medium traffic have received much research attention since the advent of digital communications. Buffered, they have typically been modelled as a discrete time queueing system. A large number of these models are available in the literature, e.g. [2–6].

Multiplexers handling multimedia traffic, on the other hand, have gained research attention only in recent years due to the ever proliferating multimedia applications. They have typically been modelled as a discrete priority queueing system. A number of these models are available in the literature [7–12]. The model in this paper is similar to these models in that it is a discrete time priority queueing system. However, it is different in that it incorporates three

IEE Proceedings online no. 20030218
doi:10.1049/ip-com:20030218
Paper first received 12th July and in revised form 3rd December 2002

H. Nassar is with the Computer Science Department, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt
H. Al Mahdi is with the Mathematics and Sciences Department, Faculty of Petroleum Engineering, Suez Canal University, Suez, Egypt

distinguishing features. First, it uses a difference, rather than stochastic, equation approach, making for better tracking and visualisation. The visualisation is further enhanced by supplying a transition diagram for the multiplexer states. Second, it uses non-pre-emptive, rather than pre-emptive, priority. Unlike pre-emptive priority, non-pre-emptive priority has the advantage of not wasting any service time already expended by the multiplexer. Third, it uses a geometric, rather than deterministic, service time. The geometric service time, in addition to accommodating the deterministic service time as a special case, can accommodate wider application possibilities. For instance, it accommodates the possibility that prior to transmission the cells undergo some geometric processing time, for activities such as error encoding and decoding or encryption and decryption. Also, it accommodates the possibility that the multiplexer uses an acknowledgment system, whereby it will not dispose of a cell it has transmitted before receiving an acknowledgment from the remote end, retransmitting the cell continually in each intervening slot. Needless to say, the transmission time in this case is geometric.

2 Model assumptions

The most basic assumption in this analysis is that the multiplexer operates in a discrete time manner. That is, the time axis is divided into slots, each exactly equal to the transmission time of one cell. Non-negative integers $k = 0, 1, \cdots$, are assigned to the individual slot boundaries. Time interval $[k, k+1)$ is referred to as slot $k+1$.

In the following, we formally state the remaining assumptions, which are largely reflected by the diagram in Fig. 1.

(i) The multiplexer has $N \leq \infty$ input lines and 1 output line.

(ii) The multiplexer has an infinite-capacity buffer to host the arriving cells until they are moved, one by one, to the transmission (service) phase where they go to the server. The time a cell spends in the buffer is called the queueing time, that in the server is called the service time, and the sum of both is called the waiting time.

(iii) Cells arrive into the multiplexer as a Bernoulli process. That is, every slot a cell will arrive (from all the inputs) with probability r and will not arrive with probability $\bar{r} = 1 - r$. This implies that the cell arrival rate at the multiplexer is r cells per slot, and that the cell interarrival time is geometrically distributed with expectation $1/r$ slots.

(iv) Given that a cell has arrived at the multiplexer, it is either of class-1 with probability λ or of class-2 with probability $\bar{\lambda} = 1 - \lambda$. This implies that the class-1 batch arrival rate is λ and the class-2 batch arrival rate is $\bar{\lambda}$, and that the batch interarrival times of class-1 and class-2 cells are geometrically distributed with expectations $1/\lambda$ and $1/\bar{\lambda}$, respectively. It also implies that the class-1 cell arrival rate is $r_1 = \lambda r$ and the class-2 cell arrival rate is $r_2 = \bar{\lambda} r$, and that the cell interarrival times of class-1 and class-2 cells are geometrically distributed with expectations $1/r_1$ and $1/r_2$, respectively. It is clear that the cell arrival rate r, regardless of class, is related to r_1 and r_2 through the relation

$$r = r_1 + r_2 \tag{1}$$

(v) The arriving cells are stored in the buffer in the form of a first-come-first-served (FCFS) queue.

(vi) The multiplexer has a single server, e.g. a register, to host the cell under transmission.

(vii) A cell may enter a queue or service only at the beginning of a slot. This implies that a cell arriving at the multiplexer after the beginning of a given slot is not considered to be in the multiplexer throughout that slot.

(viii) A cell being served in a certain slot will end service by the end of that slot with probability s and will not end service by the end of the slot with probability $\bar{s} = 1 - s$. This implies that the cell service rate of the multiplexer is s cells per slot, and that the service time is geometrically distributed with expectation $1/s$ slots. A cell may end service only at the end of a slot.

(ix) Class-1 cells have priority over class-2 cells. This implies that the multiplexer has two logical queues, one of class-1 cells and one of class-2 cells. Each queue operates on a FCFS basis and no class-2 cell can enter service unless the class-1 queue is empty.

(x) The type of priority is non-pre-emptive. That is, a class-1 cell that arrives while the class-1 queue is empty and a class-2 cell is being served will wait until the latter finishes service.

In the next Section we analyse the occupancy of the multiplexer under the above assumptions. Most of the variables considered in the analysis are random variables (RVs), all of them non-negative and integral-valued.

3 Multiplexer occupancy

Let $P_1^k = 0, 1, \cdots$, be a RV denoting the class-1 occupancy in slot k, i.e. the number of class-1 cells in the multiplexer at the end of slot k. Similarly, let $P_2^k = 0, 1, \cdots$, be a RV denoting the class-2 occupancy in slot k. Clearly, the state of the multiplexer in each slot k is fully determined by the pair (P_1^k, P_2^k), with the pairs (P_1^k, P_2^k), for all k, forming a two-dimensional Markov chain [13]. Let $p_{m,n}^k = \Pr\ [P_1^k = m, P_2^k = n]$ be the transient distribution of that chain. It is the aim of this Section to seek the stationary distribution $p_{m,n}$ of that chain, defined as the limit as k tends to ∞ of $p_{m,n}^k$.

To facilitate finding the distribution of the chain (P_1^k, P_2^k), we will identify another Markov chain by considering the type of cell being served. To this end, let $L^k = 0, 1, 2$, be a RV denoting the type of the cell being served in slot k, where $L^k = 1$ indicates a class-1 cell, $L^k = 2$ a class-2 cell, and $L^k = 0$ indicates that no cell is being served (multiplexer empty). Clearly, the triplets (P_1^k, P_2^k, L^k), for all k, form a three-dimensional Markov chain. Let $p_{m,n,l}^k = \Pr[P_1^k = m, P_2^k = n, L^k = l]$ be the transient distribution of that chain. It turns out that the transitions of the chain (P_1^k, P_2^k, L^k) are easier to track down than those of the chain (P_1^k, P_2^k). First, they can be readily evaluated from the assumptions given in Section 2. Second, they can be visualized by a state transition diagram for convenience as is done in Fig. 2.

The transition diagram has nodes which represent the states the multiplexer can be in, and directed arcs which represent the transitions between the states. Written next to each transition is the probability that this transition is made as time advances from one slot to the next. To illustrate this, consider state (2,1,1), which means that the system will have two class-1 cells, one class-2 cell and the cell in service will be of class-1. This state will be reached in a certain slot $k+1$ if the multiplexer in the previous state k is in any one of the following eight states and some event occurs.

(i) State (2,1,1) itself. This takes place if, by the end of slot k: (the cell in service does not depart AND no cell arrives) OR (the cell in service departs AND one class-1 cell arrives). The probability of this event is $\bar{s}\bar{r} + sr_1$.

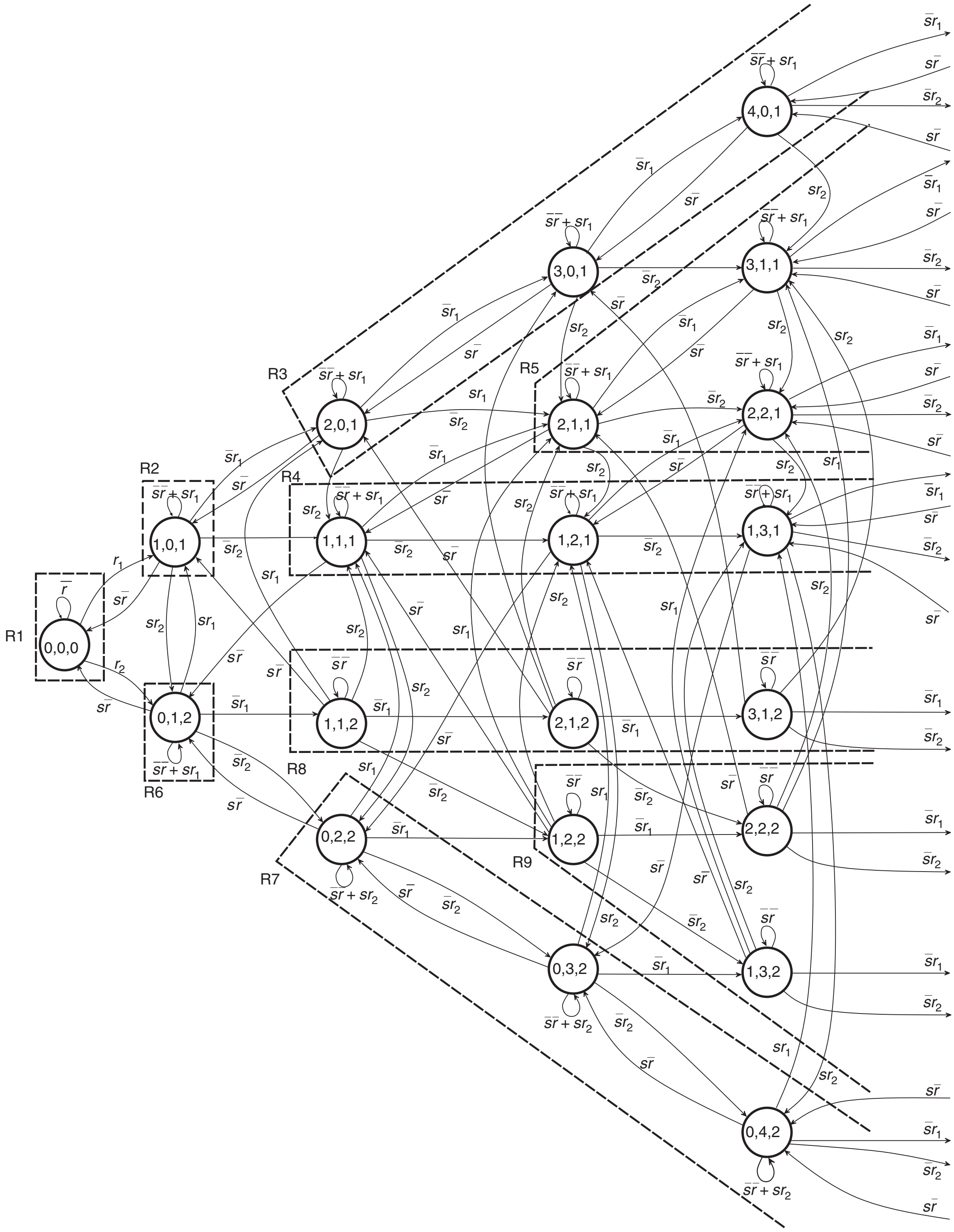

Fig. 2 *Multiplexer state transition diagram*

(ii) State (2,0,1). This takes place if, by the end of slot k: the cell in service does not depart AND one class-2 cell arrives. The probability of this event is $\bar{s}r_2$.

(iii) State (3,0,1). This takes place if, by the end of slot k: the cell in service departs AND one class-2 cell arrives. The probability of this event is sr_2.

(iv) State (2,1,2). This takes place if, by the end of slot k: the cell in service departs AND one class-2 cell arrives. The probability of this event is sr_2.

(v) State (3,1,1). This takes place if, by the end of slot k: the cell in service departs AND no cells arrive. The probability of this event is $s\bar{r}$.

(vi) State (2,2,2) itself. This takes place if, by the end of slot k: the cell in service departs AND no cells arrive. The probability of this event is $s\bar{r}$.

(vii) State (1,1,1) itself. This takes place if, by the end of slot k: the cell in service does not depart AND one class-1 cell arrives. The probability of this event is $\bar{s}r_1$.

(viii) State (1,2,2) itself. This takes place if, by the end of slot k: the cell in service departs AND one class-2 cell arrives. The probability of this event is sr_1.

The above enumeration explains why in Fig. 2 there are eight arcs going *into* node (2,1,1). It also explains how the difference equation is derived. Actually, the eight states and events above are generalised and used to construct the eight terms of (2). Equations (3)–(10) are constructed analogously. Each equation in the system of (2)–(10) represents a region in the transition diagram of Fig. 2. In fact, a region in the diagram is identified by being summarisable by a single difference equation. Accordingly, the transition diagram is divided into nine such regions, numbered $R1$ to $R9$ consecutively.

In what follows, we derive a set of subequations that collectively form a difference equation defining the distribution $p^k_{m,n}$. First, we focus on the case where there is a class-1 cell in the server. This case appears in regions $R2$, $R3$, $R4$, and $R5$ of the transition diagram. From region $R5$ one can write the principal equation

$$\begin{aligned} p^{k+1}_{m,n,1} = &(sr_1 + \bar{s}\,\bar{r})\,p^k_{m,n,1} \\ &+ \bar{s}r_2\,p^k_{m,n-1,1} + sr_2\,p^k_{m+1,n-1,1} \\ &+ sr_2\,p^k_{m,n,2} + s\bar{r}\,p^k_{m+1,n,1} + s\bar{r}\,p^k_{m,n+1,2} + \bar{s}r_1\,p^k_{m-1,n,1} \\ &+ sr_1\,p^k_{m-1,n+1,2},\ m = 2,3,\cdots,n = 1,2,\cdots \end{aligned} \tag{2}$$

This equation has three boundary conditions as follows. From region $R3$ we obtain

$$\begin{aligned} p^{k+1}_{m,0,1} = &(sr_1 + \bar{s}\,\bar{r})\,p^k_{m,0,1} + \bar{s}r_1\,p^k_{m-1,0,1} + sr_1\,p^k_{m-1,1,2} \\ &+ s\bar{r}\,p^k_{m+1,0,1} + s\bar{r}\,p^k_{m,1,2},\ m = 2,3,\cdots \end{aligned} \tag{3}$$

From region $R4$ we obtain

$$\begin{aligned} p^{k+1}_{1,n,1} = &(sr_1 + \bar{s}\,\bar{r})\,p^k_{1,n,1} \\ &+ \bar{s}r_2\,p^k_{1,n-1,1} + sr_1\,p^k_{0,n+1,2} + sr_2\,p^k_{2,n-1,1} + sr_2\,p^k_{1,n,2} \\ &+ s\bar{r}\,p^k_{2,n,1} + s\bar{r}\,p^k_{1,n+1,2}, \qquad n = 1,2,\cdots \end{aligned} \tag{4}$$

Then, from region $R2$ we obtain

$$\begin{aligned} p^{k+1}_{1,0,1} = &(sr_1 + \bar{s}\,\bar{r})\,p^k_{1,0,1} + s\bar{r}\,p^k_{2,0,1} \\ &+ s\bar{r}\,p^k_{1,1,2} + r_1\,p^k_{0,0,0} + sr_1\,p^k_{0,1,2} \end{aligned} \tag{5}$$

Second, we focus on the case where there is a class-2 cell in the server. This case appears in regions $R6$, $R7$, $R8$, and $R9$ of the transition diagram. From region $R9$ one can write this principal equation

$$\begin{aligned} p^{k+1}_{m,n,2} = &\bar{s}\,\bar{r}\,p^k_{m,n,2} + \bar{s}r_1\,p^k_{m-1,n,2} + \bar{s}r_2\,p^k_{m,n-1,2}, \\ m = &1,2,\cdots,n = 2,3,\cdots \end{aligned} \tag{6}$$

This equation has three boundary conditions obtained as follows. From region $R7$ we obtain

$$\begin{aligned} p^{k+1}_{0,n,2} = &(sr_2 + \bar{s}\,\bar{r})\,p^k_{0,n,2} + \bar{s}r_2\,p^k_{0,n-1,2} + sr_2\,p^k_{1,n-1,1} \\ &+ s\bar{r}\,p^k_{0,n+1,2} + s\bar{r}\,p^k_{1,n,1},\ n = 2,3,\cdots \end{aligned} \tag{7}$$

From region $R8$ we obtain

$$p^{k+1}_{m,1,2} = \bar{s}\,\bar{r}\,p^k_{m,1,2} + \bar{s}r_1\,p^k_{m-1,1,2}, \qquad m = 1,2,\cdots \tag{8}$$

Then from region $R6$ we obtain

$$\begin{aligned} p^{k+1}_{0,1,2} = &(sr_2 + \bar{s}\,\bar{r})\,p^k_{0,1,2} + r_2\,p^k_{0,0,0} + s\bar{r}\,p^k_{0,2,2} \\ &+ s\bar{r}\,p^k_{1,1,1} + sr_2\,p^k_{1,0,1} \end{aligned} \tag{9}$$

Finally, we focus on the case where there is no cell in the server. For this case, from region $R1$ of the transition diagram we have the boundary condition

$$p^{k+1}_{0,0,0} = \bar{r}\,p^k_{0,0,0} + s\bar{r}\left(p^k_{1,0,1} + p^k_{0,1,2}\right) \tag{10}$$

Assuming, the cell arrival rate at the multiplexer is strictly less than the cell service rate, then the queue in the buffer will be stable (i.e. will not blow up) and the multiplexer will reach steady state after a sufficiently large number of slots. That is, if $r<s$ then the limit

$$\lim_{k\to\infty} p^k_{m,n,l} = p_{m,n,l}$$

exists. Thus, in steady state, (2)–(10) become

$$\begin{aligned} &(sr_1 + \bar{s}\,\bar{r} - 1)\,p_{m,n,1} + \bar{s}r_2\,p_{m,n-1,1} \\ &\quad + sr_2\,p_{m+1,n-1,1} + sr_2\,p_{m,n,2} \\ &\quad + s\bar{r}\,p_{m+1,n,1} + s\bar{r}\,p_{m,n+1,2} \\ &\quad + \bar{s}r_1\,p_{m-1,n,1} + sr_1\,p_{m-1,n+1,2} = 0 \\ &\quad m = 2,3,\cdots,n = 1,2,\cdots \end{aligned} \tag{11a}$$

$$\begin{aligned} &(sr_1 + \bar{s}\,\bar{r} - 1)\,p_{m,0,1} + \bar{s}r_1\,p_{m-1,0,1} + sr_1\,p_{m-1,1,2} \\ &\quad + s\bar{r}\,p_{m+1,0,1} + s\bar{r}\,p_{m,1,2} = 0,\ m = 2,3,\cdots \end{aligned} \tag{11b}$$

$$\begin{aligned} &(sr_1 + \bar{s}\,\bar{r} - 1)\,p_{1,n,1} + \bar{s}r_2\,p_{1,n-1,1} + sr_1\,p_{0,n+1,2} \\ &\quad + sr_2\,p_{2,n-1,1} + sr_2\,p_{1,n,2} + s\bar{r}\,p_{2,n,1} \\ &\quad + s\bar{r}\,p_{1,n+1,2} = 0, \quad n = 1,2,\cdots \end{aligned} \tag{11c}$$

$$\begin{aligned} &(sr_1 + \bar{s}\,\bar{r} - 1)\,p_{1,0,1} + s\bar{r}\,p_{2,0,1} + s\bar{r}\,p_{1,1,2} \\ &\quad + r_1\,p_{0,0,0} + sr_1\,p_{0,1,2} = 0 \end{aligned} \tag{11d}$$

$$\begin{aligned} &(\bar{s}\,\bar{r} - 1)\,p_{m,n,2} + \bar{s}r_1\,p{+}_{m-1,n,2} \\ &\quad + \bar{s}r_2\,p_{m,n-1,2} = 0,\ m = 1,2,\cdots,n = 2,3,\cdots \end{aligned} \tag{11e}$$

$$\begin{aligned} &(sr_2 + \bar{s}\,\bar{r} - 1)\,p_{0,n,2} + \bar{s}r_2\,p_{0,n-1,2} + sr_2\,p_{1,n-1,1} \\ &\quad + s\bar{r}\,p_{0,n+1,2} + s\bar{r}\,p_{1,n,1} = 0,\ n = 2,3,\cdots \end{aligned} \tag{11f}$$

$$(\bar{s}\,\bar{r} - 1)\,p_{m,1,2} + \bar{s}r_1\,p_{m-1,1,2} = 0,\ m = 1,2,\cdots \tag{11g}$$

$$\begin{aligned} &(sr_2 + \bar{s}\,\bar{r} - 1)\,p_{0,1,2} + r_2\,p_{0,0,0} + s\bar{r}\,p_{0,2,2} + s\bar{r}\,p_{1,1,1} \\ &\quad + sr_2\,p_{1,0,1} = 0 \end{aligned} \tag{11h}$$

$$r\,p_{0,0,0} - s\bar{r}\left(p_{1,0,1} + p_{0,1,2}\right) = 0 \tag{11i}$$

Now, let $P\,(z_1, z_2)$ be the probability generating function (PGF) of the distribution $p_{m,n}$. That is,

$$\begin{aligned} &P(z_1,z_2) = P_1(z_1,z_2) + P_2(z_1,z_2) + p_0, \\ &|z_1|<1, |z_2|<1 \end{aligned} \tag{12}$$

where

$$P_1(z_1,z_2) = \sum_{m=1}^{\infty}\sum_{n=0}^{\infty} p_{m,n,1} z_1^m z_2^n$$

$$P_2(z_1,z_2) = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} p_{m,n,2} z_1^m z_2^n$$

and $p_0 = p_{0,0,0}$. In what follows, we use (11) to obtain $P\,(z_1, z_2)$.

We start by multiplying (11) by $z_1^m z_2^n$ and summing over the range over which the constituent subequations are defined. Thus, we get

$$(sr_1+\bar{s}\,\bar{r}-1)\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m,n,1}z_1^m z_2^n+\bar{s}r_2\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m,n-1,1}z_1^m z_2^n$$
$$+sr_2\left(\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m+1,n-1,1}z_1^m z_2^n+\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m,n,2}z_1^m z_2^n\right)$$
$$+s\bar{r}\left(\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m+1,n,1}z_1^m z_2^n+\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m,n+1,2}z_1^m z_2^n\right)$$
$$+\bar{s}r_1\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m-1,n,1}z_1^m z_2^n+sr_1\sum_{m=2}^{\infty}\sum_{n=1}^{\infty}p_{m-1,n+1,2}z_1^m z_2^n$$
$$=0 \tag{13a}$$

$$(sr_1+\bar{s}\,\bar{r}-1)\sum_{m=2}^{\infty}p_{m,0,1}z_1^m$$
$$+\bar{s}r_1\sum_{m=2}^{\infty}p_{m-1,0,1}z_1^m+sr_1\sum_{m=2}^{\infty}p_{m-1,1,2}z_1^m$$
$$+s\bar{r}\sum_{m=2}^{\infty}p_{m+1,0,1}z_1^m+s\bar{r}\sum_{m=2}^{\infty}p_{m,1,2}z_1^m=0 \tag{13b}$$

$$(sr_1+\bar{s}\,\bar{r}-1)\sum_{n=1}^{\infty}p_{1,n,1}z_2^n z_1$$
$$+\bar{s}r_2\sum_{n=1}^{\infty}p_{1,n-1,1}z_2^n z_1+sr_1\sum_{n=1}^{\infty}p_{0,n+1,2}z_2^n z_1$$
$$+sr_2\left(\sum_{n=1}^{\infty}p_{2,n-1,1}z_2^n z_1+\sum_{n=1}^{\infty}p_{1,n,2}z_2^n z_1\right)$$
$$+s\bar{r}\left(\sum_{n=1}^{\infty}p_{2,n,1}z_2^n z_1+\sum_{n=1}^{\infty}p_{1,n+1,2}z_2^n z_1\right)=0 \tag{13c}$$

$$(sr_1+\bar{s}\,\bar{r}-1)p_{1,0,1}z_1+s\bar{r}(p_{2,0,1}z_1+p_{1,1,2}z_1)$$
$$+r_1p_{0,0,0}z_1+sr_1p_{0,1,2}z_1=0 \tag{13d}$$

$$(\bar{s}\,\bar{r}-1)\sum_{m=1}^{\infty}\sum_{n=2}^{\infty}p_{m,n,2}z_1^m z_2^n$$
$$+\bar{s}r_1\sum_{m=1}^{\infty}\sum_{n=2}^{\infty}p_{m-1,n,2}z_1^m z_2^n$$
$$+\bar{s}r_2\sum_{m=1}^{\infty}\sum_{n=2}^{\infty}p_{m,n-1,2}z_1^m z_2^n=0 \tag{13e}$$

$$(sr_2+\bar{s}\,\bar{r}-1)\sum_{n=2}^{\infty}p_{0,n,2}z_2^n$$
$$+\bar{s}r_2\sum_{n=2}^{\infty}p_{0,n-1,2}z_2^n+sr_2\sum_{n=2}^{\infty}p_{1,n-1,1}z_2^n$$
$$+s\bar{r}\left(\sum_{n=2}^{\infty}p_{0,n+1,2}z_2^n+\sum_{n=2}^{\infty}p_{1,n,1}z_2^n\right)=0 \tag{13f}$$

$$(\bar{s}\,\bar{r}-1)\sum_{m=1}^{\infty}p_{m,1,2}z_1^m z_2+\bar{s}r_1\sum_{m=1}^{\infty}p_{m-1,1,2}z_1^m z_2=0 \tag{13g}$$

$$(sr_2+\bar{s}\,\bar{r}-1)p_{0,1,2}z_2+r_2p_{0,0,0}z_2$$
$$+s\bar{r}(p_{0,2,2}z_2+p_{1,1,1}z_2)+sr_2p_{1,0,1}z_2=0 \tag{13h}$$

$$rp_{0,0,0}-s\bar{r}(p_{1,0,1}+p_{0,1,2})=0 \tag{13i}$$

The function $P_1(z_1, z_2)$ can be obtained by summing (13*a*)–(13*d*) to obtain

$$P_1(z_1,z_2)=\frac{\begin{array}{r}[z_1[sA(z_1,z_2)P_2(z_1,z_2)-sA(0,z_2)P_2(0,z_2)\\-z_2sA(0,z_2)P_{11}(z_2)+z_1z_2r_1p_0]]\end{array}}{z_2(z_1-sr_1z_1-\bar{s}z_1A(z_1,z_2)-sA(0,z_2))} \tag{14}$$

where $P_{11}(z_2)=\sum_{n=0}^{\infty}p_{1,n,1}z_2^n$, and $A(z_1,z_2)=\bar{r}+r_1z_1+r_2z_2$. The function $P_2(z_1, z_2)$ can be obtained by summing (13*e*)–(13*i*) to obtain

$$P_2(z_1,z_2)=\frac{\begin{array}{r}[sA(0,z_2)z_2P_{11}(z_2)+sA(0,z_2)P_2(0,z_2)\\+(A(0,z_2)-1)z_2p_0]\end{array}}{(1-\bar{s}A(z_1,z_2))z_2} \tag{15}$$

The function $P_{11}(z_2)$ can be obtained using (13*a*) and (13*i*) to obtain

$$P_{11}(z_2)=\frac{(z_2-sA(0,z_2)-\bar{s}z_2A(0,z_2))P_2(0,z_2)-z_2(A(0,z_2)-1)p_0}{z_2sA(0,z_2)} \tag{16}$$

The probability p_0 can be obtained as follows. Since the probability that a queueing system, regardless of whether uniclass or biclass, is empty is just the complement of the probability that it is busy, and since the latter is known [14] to be equivalent to the utilisation factor ρ, defined to be the ratio of cell arrival rate to cell service rate, then in our case we have

$$p_0=1-\frac{r}{s} \tag{17}$$

Combining (14)–(17), it follows that

$$P(z_1,z_2)=A(z_1,z_2)\times\frac{\begin{array}{l}[s(1-\bar{s}A(0,z_2))(z_1-z_2)P_2(0,z_2)\\ \quad+z_2(z_1-1)(s-r)(1-\bar{s}A(z_1,z_2))]\end{array}}{z_2(z_1-A(z_1,z_2)(s+\bar{s}z_1))(1-\bar{s}A(z_1,z_2))} \tag{18}$$

Equation (18) does not quite give the final form of $P(z_1, z_2)$ since $P_2(0, z_2)$ is still unknown. To find $P_2(0, z_2)$ we use the same methodology as in ([16], p. 131). It is based on noting that the PGF $P(z_1, z_2)$ is analytic on the unit disc $|z_1|\le 1$, for any fixed value of z_2 for which $|z_2|<1$ ([15], p. 394) and [10]. The analyticity implies that whenever the denominator in (18) has zeros on the unit disc, the numerator must have the same zeros on it. So, by finding those zeros and plugging them into the numerator, where the unknown function exists, we can find the unknown. So, it is time now to find in our case the zeros of the denominator in (18) and identify their location regarding the unit disc.

Recalling that the arrival and service rates (which are probabilities at the same time) are such that $0\le r<s\le 1$, and that $|z_2|<1$, it can first be seen that the last factor

$$1-\bar{s}(\bar{r}+r_1z_1+r_2z_2)=0$$

has exactly one zero, namely $z_1 = (1 - \bar{s}\bar{r} - \bar{s}r_2z_2)/\bar{s}r_1$. This zero lies outside the unit disc since

$$\begin{aligned}
|z_1| &= \left|\frac{1}{\bar{s}r_1} - \frac{\bar{r}}{r_1} - \frac{r_2z_2}{r_1}\right| \\
&\geq \left|\frac{1}{\bar{s}r_1} - \frac{\bar{r}}{r_1}\right| - \left|\frac{r_2z_2}{r_1}\right| \\
&= \left|\frac{1}{\bar{s}r_1} - \frac{1}{r_1} + \frac{r}{r_1}\right| - \frac{r_2}{r_1}|z_2| \\
&> \left|\frac{1}{\bar{s}r_1} - \frac{1}{r_1} + 1 + \frac{r_2}{r_1}\right| - \frac{r_2}{r_1} \\
&> \left|1 + \frac{r_2}{r_1}\right| - \frac{r_2}{r_1} \\
&= 1
\end{aligned}$$

Second, for the factor

$$z_1 - (\bar{r} + r_1z_1 + r_2z_2)(s + \bar{s}z_1) = 0 \tag{19}$$

it can be shown using Rouché's theorem ([15], p. 20) that it has exactly one zero on the unit disc $z_1 \leq 1$. Since the factor is quadratic in z_1, it has two zeros, say ξ_1 and ξ_2, both functions of z_2. These zeros can be found by employing the quadratic formula yielding

$$\xi_1 = \frac{[s + r\bar{s} - sr_1 - \bar{s}r_2z_2 - \sqrt{(s + r\bar{s} - sr_1 - \bar{s}r_2z_2)^2 - 4\bar{s}r_1(s\bar{r} + sr_2z_2)}\,]}{2\bar{s}r_1}$$

and

$$\xi_2 = \frac{[s + r\bar{s} - sr_1 - \bar{s}r_2z_2 + \sqrt{(s + r\bar{s} - sr_1 - \bar{s}r_2z_2)^2 - 4\bar{s}r_1(s\bar{r} + sr_2z_2)}\,]}{2\bar{s}r_1}$$

It is clear that $|\xi_1| < |\xi_2|$, implying that ξ_1 is the zero that lies on the disc.

The above zero-finding analysis indicates that the denominator of (18) has only one zero of z_1 on the unit disc, namely ξ_1. Thus, for $P(z_1, z_2)$ to be analytic on the unit disc, ξ_1 must also be a zero of the numerator of (18). That is, if we substitute $z_1 = \xi_1$ in that numerator we should get a 0, which enables us to find $P\ (0, z_2)$ as follows:

$$P_2(0, z_2) = \frac{z_2(1 - \xi_1)(s - r)(1 - \bar{s}(A(0, z_2) + r_1\xi_1))}{s(1 - \bar{s}A(0, z_2))(\xi_1 - z_2)}$$

completing the derivation of the PGF $P\ (z_1, z_2)$ in (18).

One means of verifying (18) is to force the multiplexer to collapse to single class, with arrival rate r cells per slot and service rate s cells per slot. Denoting the occupancy of that single class multiplexer by P, its PGF $P(z)$ can be obtained from (18) using any of the following three methods.

(i) Substitute for $z_2 = z_1 = z$ in (18). This is equivalent to assuming that priorities among the cells are abolished.

(ii) Substitute for $r_2 = 0$ in (18), noting that the coefficient of z_2 there will vanish, and that $z_2 = z_1 = z$. This is equivalent to assuming that only class-1 cells arrive at the multiplexer.

(iii) Substitute for $r_1 = 0$ in (18), noting that the coefficient of z_1 there will vanish, and that $z_2 = z_1 = z$. This is equivalent to assuming that only class-2 cells arrive at the multiplexer.

All three methods yield the same classical result [11, 14]:

$$P(z) = \frac{(s - r)(rz + \bar{r})}{s\bar{r} - rz\bar{s}} \tag{20}$$

Before closing, it is worth noting that (18) can be used to obtain the PGF of several interesting multiplexer occupancies. For example, the marginal PGF $P_1\ (z)$ of class-1 occupancy P_1 can be obtained as follows:

$$\begin{aligned}
P_1(z) &= P(z, 1) \\
&= \frac{(1 - r_1 + r_1z)(r_2 + (1 - \bar{s}(1 - r_1 + r_1z))(s - r))}{(s - r_1s - r_1\bar{s}z)(1 - \bar{s}(1 - r_1 + r_1z))}
\end{aligned} \tag{21}$$

Also, the marginal PGF $P_2\ (z)$ of class-2 occupancy can be obtained as follows:

$$\begin{aligned}
P_2(z) &= P(1, z) \\
&= \frac{s(1 - r_2 + r_2z)(1 - \bar{s}(\bar{r} + r_2z))(1 - z)P_{02}(z)}{z(1 - (1 - r_2 + r_2z))(1 - \bar{s}(1 - r_2 + r_2z))}
\end{aligned} \tag{22}$$

These two results conform well with those obtained by the stochastic equation approach [11, 12].

4 Expectations and numerical results

In this Section we obtain expressions for the expectations of the multiplexer occupancy. These expressions are then used to find expressions for the expectations of cell waiting time.

First, the expected class-1 multiplexer occupancy can be obtained from (21) by evaluating the first derivative at 1 as follows:

$$E[P_1] = P_1'(1) = r_1\frac{s\bar{r} + r_2}{s(s - r_1)} \tag{23}$$

Second, the expected multiplexer occupancy regardless of cell class, i.e. of the RV $P_1 + P_2$, is obtained from (20) as follows:

$$E[P] = P'(1) = \frac{r\bar{r}}{s - r} \tag{24}$$

Finally, we can get the expected class-2 multiplexer occupancy as follows:

$$E[P_2] = E[P] - E[P_1] = r_2\frac{s^2\bar{r} - r_1(s - r)}{s(s - r)(s - r_1)} \tag{25}$$

It should be noted that the expected class-1 and class-2 waiting times can be obtained, respectively, by dividing $E[P_1]$ by r_1 and $E\ [P_2]$ by r_2, making use of Little's formula [14]. Namely, we get

$$E[W_1] = \frac{s\bar{r} + r_2}{s(s - r_1)}$$

and

$$E[W_2] = \frac{s^2\bar{r} - r_1(s - r)}{s(s - r)(s - r_1)}$$

Now, we use (23)–(25) to plot the expected occupancy for some example multiplexers. Fig. 3 shows the expected multiplexer occupancy against the class-1 arrival rate, for constant class-2 arrival rate $r_2 = 0.2$ and constant service rate $s = 0.95$. Three occupancies are shown: class-1, $E[P_1]$, class-2, $E[P_2]$, and total occupancy, $E[P]$ (regardless of class, that is). As anticipated, both class-1 and class-2 occupancies increase as the class-1 arrival rate increases. However, the rate of increase of class-1 is higher (and, incidently, almost linear). This is understandable since class-1 cells increasingly dominate the arrivals as r_1 goes up.

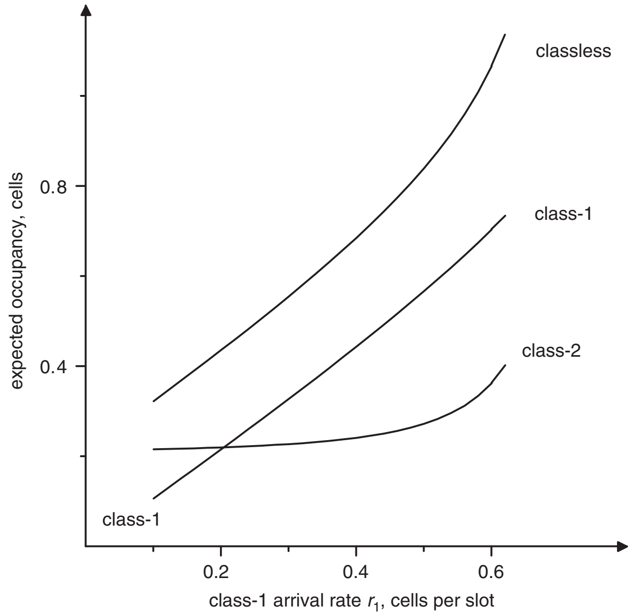

Fig. 3 *Expected occupancy against class-1 arrival rate, for $s = 0.95$ and $r_2 = 0.2$*

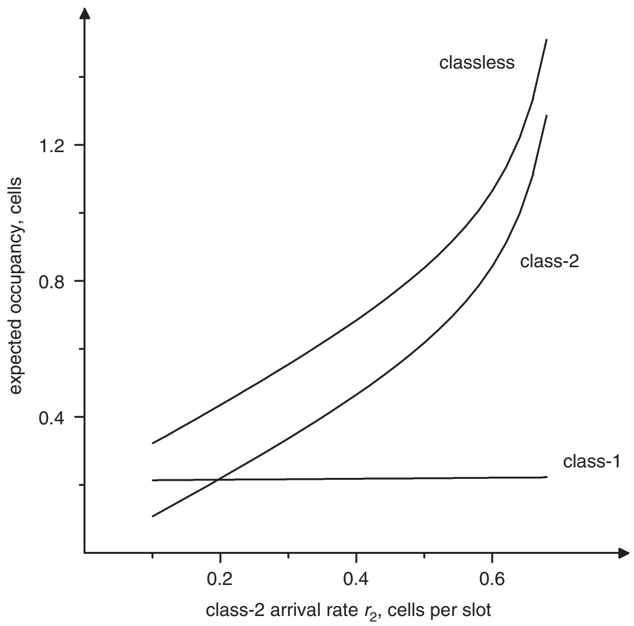

Fig. 4 *Expected occupancy against class-2 arrival rate, for $s = 0.95$ and $r_1 = 0.2$*

Note the intersection of class-1 and class-2 occupancies, which arises as follows. At low class-1 arrival rate, i.e. at $r_1 < r_2 = 0.2$, the arrivals of class-2 are dominant, hence class-2 occupancy is higher than class-1 occupancy. As r_1 increases, class-1 occupancy approaches, becomes equal to, and then goes beyond class-2 occupancy, hence the intersection at the equality point.

It can easily be seen that the total occupancy curve stands above that of class-1. This shows the difference between a multiplexer without priority (real-time cell embedded in a long queue, thus incurring a harmful, long waiting time) and one with priority (real-time cell separated in a shorter queue, thus incurring a shorter waiting time). It is thus recommended to use priority when there are cells to be given faster service than others.

Fig. 4 shows the expected multiplexer occupancy against the class-2 arrival rate, for constant class-1 arrival rate $r_1 = 0.2$ and constant service rate $s = 0.95$. The Figure is similar to Fig. 3, and thus similar comments apply. The notable observation here, however, is that class-1 occupancy is almost constant. This is due to the fact that class-1 occupancy is not affected by class-2 arrival rate except in one obviously rare circumstance, namely when a class-1 cell arrives at a class-1 empty multiplexer finding a class-2 cell in service.

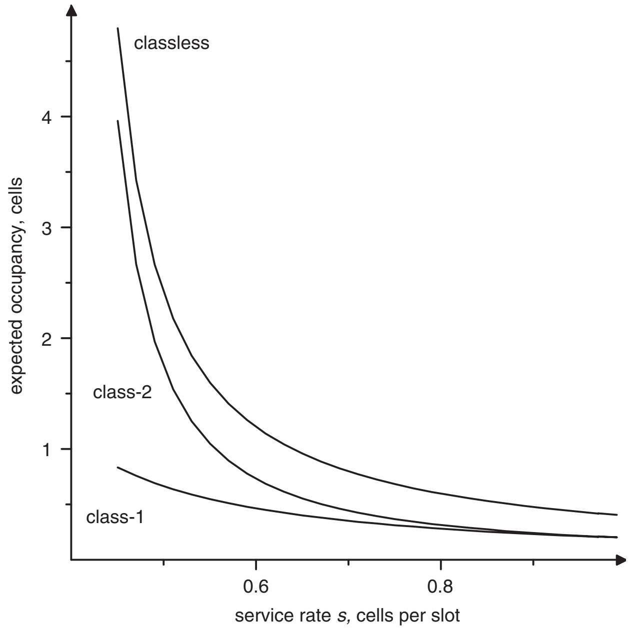

Fig. 5 *Expected occupancy against service rate, for $r_1 = r_2 = 0.2$*

Fig. 5 shows the expected occupancy against the multiplexer service rate s for constant class-1 and class-2 arrival rates $r_1 = r_2 = 0.2$. The noticeable feature here is that the three occupancies decrease with increase of s, albeit at different rates. This difference can easily be justified along the same lines given above.

5 Conclusions

In this paper the occupancy of a multiplexer handling two class traffic has been analysed using a non-pre-emptive priority scheme. For the purpose of the analysis, the multiplexer is modelled as a priority, discrete-time, single-server, single-arrival queueing system with infinite buffer. The major contribution of the paper is use of a difference equation approach rather than a stochastic equation approach (as has typically been the case in the published literature) to perform the analysis. The advantage of the difference equation approach is that the physical details of the system are always present and visible during the analysis, so much so that one can even construct a diagram showing the transitions among the system states.

We have obtained results for the multiplexer occupancy and, as an aside, for the cell waiting time. The results have been verified by showing that they conform with previously published results. They have also been verified by showing that they generate intuitively acceptable graphs when translated into numerical values.

6 Acknowledgment

This work was carried out under the auspices of the Math Department, Faculty of Science, Suez Canal University. The authors wish to thank the faculty members of the Department, and in particular Dr Sherif Zaki and

Dr Nagwa Sherif. In addition, they wish to thank the anonymous referee who pinpointed a mathematical glitch which, had it gone unrecognised, would have devalued the paper.

7 References

1 REN, J.-F., MARK, J. W., and WONG, J. W.: 'Dynamic priority queueing approach to traffic regulation and scheduling in B-ISDN'. Proceedings of IEEE Global Telecommunications Conference, San Francisco, CA, USA, Nov.–Dec. 1994, pp. 612–618

2 KOBAYASHI, H., and KONHEIM, A.: 'Queueing models for computer communications multiplexer analysis', *IEEE Trans. Commun.*, 1977, **25**, (1), pp. 1–27

3 KRUSKAL, C., SINR, M., and WEISS, A.: 'The distribution of queuing times in clocked multistage interconnection networks', *IEEE Trans. Comput.*, 1988, **37**, (11), pp. 1347–1356

4 GRACIA, J., and CASALS, O.: 'Discrete time queueing model to study the packet delay variation in an ATM network', *Perform. Eval.*, 1994, **25**, (11), pp. 3–22

5 STEYAERT, B., and XIONG, Y.: 'Analysis of a discrete-time queue with general three-state Markovian traffic sources', *Perform. Eval.*, 1996, **27**, (4), pp. 277–294

6 CIDON, I., and SIDI, M.: 'Dispersed messages in discrete-time queues: delay, jitter and threshold crossing', *Perform. Eval.*, 1997, **29**, (1), pp. 85–104

7 TAKINE, T., and HASEGAWA, T.: 'An analysis of a discrete-time queue for broadband ISDN with priorities among traffic classes', *IEEE Trans. Commun.*, 1994, **42**, (2), pp. 1837–1843

8 TAKEMORI, K., and SASASE, I.: 'Performance analysis of an input and output queueing packet switch with multiple priority classes', *Electron. Commun.*, 1996, **79**, (6), pp. 38–50

9 LEE, J., and UN, C.: 'Performance analysis of space-priority mechanisms in an input and output queueing ATM switch', *IEE Proc., Commun.*, 1997, **144**, (4), pp. 229–236

10 CHOI, B. *et al.*: 'Priority queueing multiplexer with fixed-length packet-train arrivals', *IEE. Proc., Commun.*, 1998, **145**, (5), pp. 331–336

11 NASSAR, H., and AL MAHDY, H.: 'A priority discrete queueing model for multimedia multiplexers', *Math. Comput. Model. Dyn. Syst.*, 2002, **2**, (8), pp. 199–211

12 NASSAR, H., and ALI, M.: 'Performance analysis of an ATM buffered switch transmitting two-class traffic over unreliable channels', *AEU- Int. J. Electron. Commun.*, 2002, **56**, (7), pp. 1–11

13 NASSAR, H.: 'A Markov model for multiprocessor multiplexers under asynchronous operation', *Inf. Process. Lett.*, 1995, **54**, pp. 11–16

14 WOODWARD, M.: 'Communication and computer networks' (IEEE Comput. Society Press, New York, 1994)

15 TAKAGI, H.: 'Queueing analysis, Vol.1: Vacation and priority systems' (Elsevier Science, Amsterdam, 1991)

16 GROSS, D., and HARRIS, C.: 'Fundamentals of queueing theory' (John Wiley, New York, 1985 2nd edn.)