ANALYSIS OF TWO-CLASS DISCRETE PACKET QUEUES WITH HOMOGENOUS ARRIVALS AND PRIORITIZED SERVICE*

HAMED NASSAR† AND YASSER FOUAD‡

Abstract. In this article we analyze the system occupancy of a discrete time queue of packets, where each packet is either of class-1 or of class-2, with class-1 receiving higher service priority than class-2. Such a queue is naturally formed in many computer and digital communications systems, e.g. multiuser computers, multiprocessing computers, file servers, ATM multiplexers, and ATM switches, when the packet sources are multimedia applications. The analysis considers both priority disciplines, nonpreemptive and preemptive, and identifies the relation between them. It demonstrates mathematically the intuitive fact that when service time is deterministically 1 slot, both priority disciplines result in the same system occupancy. The analysis is carried out under the assumption that the service time is geometric, and that the packets arrive in batches of general size, at the rate of one batch per slot. These batches are homogeneous in the sense that each batch is either totally of class-1 packets or totally of class-2 packets. Two special cases are given at the end where the batch size is assumed once binomial and once Poisson.

Keywords. Multimedia, ATM, nonpreemptive and preemptive priority, discrete queueing, system occupancy

1. Introduction. Discrete time two-class queues of packets are formed in many computer and digital communications systems when the sources of these packets are of multimedia applications. In computer systems they are found in such installations as multiprogramming computers, multiuser computers, and multiprocessors. In digital communications systems they are found in such installations as asynchronous transfer mode (ATM) multiplexers and switches. Due to the multimedia nature of the sources, the packets are usually of two classes: class-1 and class-2. Class-1 packets, typically representing real time traffic, e.g. live audio and video, require higher service priority over class-2 packets, typically representing nonreal time traffic, e.g. file transfers.

Given their timely nature, class-1 packets need to be served more rapidly than class-2. This can be done by implementing a priority scheme for the queue, where class-1 packets are given higher priority over class-2. Two disciplines may be used if such a scheme is implemented, based on what is supposed to happen to a class-2 packet being served when a class-1 packet arrives. In the *nonpreemptive* discipline, the arriving class-1 packet will have to wait until the class-2 packet is served. In the *preemptive* discipline, on the other hand, the class-1 packet will enter service in the next slot, ejecting the class-2 packet back to the buffer. When there are no more class-1 packets in the queue, the ejected class-2 packet will reenter service. It can be

---

seen that the preemptive discipline is favorable to class-1 packets, since they *always* find the server available. However, this discipline is time wasteful since it throws away partial service time already expended on class-2 packets. The nonpreemptive discipline, by contrast, is still favorable to class-1 packets, albeit to a lesser extent, but is time conservative in that not even a single slot already expended by the server is thrown away.

Analyses of single class discrete time queueing system are available in the literature in abundance (see e.g. [1], [2], [3], [4], and [5]). By contrast, analyses of two-class discrete queues are less available. Only in recent years have such analyses begun to proliferate. For example, [6], [7], [8], and [9] study two-class discrete-time systems, with service time assumed deterministically 1 slot. Also, [10], [11], and [12] study systems with nonpreemptive priority and general service times with nonhomogeneous packet arrivals. Additionally [13], [14], and [15] study systems with preemptive priority. In [16], a preemptive priority system was considered but with two service rates, one geometric and one deterministically 1 slot. This work is generalized in [17] where the two service rates are made geometric.

Our analysis in this article is characterized by the two assumptions of homogeneous arrivals, in the sense that each arriving batch contains only one class and never the two classes, and geometric service time. In particular, the homogeneous arrival assumption cannot be generated as a special case of the above cited works. Furthermore, our analysis is characterized by the comparison between the system occupancy under the preemptive and nonpreemptive priority disciplines, and by quantifying this comparison through the derivation of relationships between the expected occupancy under both disciplines.

The article is organized as follows. The exact assumptions of our system are stated in the next section. In section 3 we analyze the joint PGF and in section 4 we analyze the marginal PGF. In section 5 we analyze the expected occupancies. A special case is considered in Section 6, where the size of the arriving batch is assumed once Binomial and once Poissonian, while in Section 7 some conclusions are drawn.

2. Model Assumptions and Notations. The queueing system to be analyzed in this article has the following assumptions.

1. The time axis is divided into equal intervals called slots. Each slot is exactly equal to the transmission time of one packet. Nonnegative integers $k = 0$, 1, ..., are assigned to individual slot boundaries. Time interval $[k, k+1)$ is referred to as slot $k + 1$.

2. The system has an infinite waiting room.

3. The system has a single server which serves packets at the rate of s packets per slot. Stated differently, provided the system is nonempty, its departure rate is s packets per slot. This implies that a packet being served in a certain

slot will depart the server by the end of that slot with probability s and will not depart with probability $\overline{s} = 1 - s$. It also implies that the service time of each packet is geometrically distributed with parameter s. There should be no confusion between the fact that the transmission time of one packet is 1 slot and that its service time is geometrically distributed. It is true that in some applications, the service time of a packet is just its transmission time, but this is not necessarily always the case. There can be some preprocessing for the packet before it is transmitted (e.g. encryption or error detection and correction calculations). Also, in acknowledged communications systems, a packet may need to be transmitted repeatedly if the output communications channel is unreliable (e.g. in wireless communications), making the service time many times the transmission time [16].

4. Arriving packets are of fixed size. These packets are queued in the buffer until they enter the server.

5. Packets arriving into the system are of two classes, class-1 and class-2. Class-1 packets have service priority over class-2. That is, no class-2 packet can enter service while a class-1 packet is present in the buffer. Thus we can look at the system as having two queues, one of class-1 packets and one of class-2 packets.

6. The discipline of service priority is nonpreemptive.

7. In each slot k, *either* a batch of $A_1^k = 0, 1, \ldots$, class-1 packets arrives with probability λ, *or* a batch of $A_2^k = 0, 1, \ldots$, class-2 packets arrives with probability $\overline{\lambda} = 1 - \lambda$. In other words, λ and $\overline{\lambda}$ are the batch arrival rates for class-1 and class-2, respectively. Both A_1^k and A_2^k are random variables (RVs) with arbitrary distributions. The A_1^k are independent and identically distributed (iid) and so are the A_2^k. Speaking about the packets, let r denote the total packet arrival rate. Then the class-1 packet arrival rate is $r_1 = \lambda E[A_1^k]$ and the class-2 packet arrival rate is $r_2 = \overline{\lambda} E[A_2^k]$, where the $E[\cdot]$ operator denotes the expected value of the random variable between brackets.

8. Each arriving batch is placed at the end of its appropriate queue, i.e. class-1 or class-2, on a first come first serve (FCFS) basis. With respect to the packets inside the batch, the placement in the queue is of random order.

9. A packet may start service only at the beginning of a slot, and may end service only just before the end of a slot.

10. A packet can not enter queue or service in its arrival slot. This implies that a packet that arrives at an empty system in a certain slot will still not enter the server before the beginning of the next slot. It also implies that a packet that has arrived in a system in a given slot will not be seen by the system till the beginning of the next slot.

The RVs used in the analysis are all both nonnegative and integral valued.

3. Joint PGF. In this section we analyze the steady state system occupancy, i.e. the number of packets in the system at the end of an arbitrary slot during steady state. The analysis will be done by looking not at all points in time but rather at a select set of points, namely the set of departure instants. Our analysis then is to focus upon the number of packets left behind by a departing packet. We are therefore describing a semi Markov process [18], where we define the imbedded Markov chain to be the number of packets present in the system immediately following a departure.

To this end, let $H_1^k = 0, 1$, be a RV denoting the number of class-1 batches that arrive at the system in slot k. That is,

$$(1) \qquad \Pr\left[H_1^k = n\right] = \begin{cases} \lambda & \text{if } n = 1 \\ \overline{\lambda} & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, let $H_2^k = 0, 1$, be a RV denoting the number of class-2 batches that arrive at the system in slot k. That is,

$$(2) \qquad \Pr\left[H_2^k = n\right] = \begin{cases} \overline{\lambda} & \text{if } n = 1 \\ \lambda & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

From the homogeneity assumption, the RVs H_1^k and H_2^k are dependent with the following conditional distribution.

$$(3) \qquad \Pr\left[H_1^k = n \mid H_2^k = m\right] = \begin{cases} 1 & \begin{array}{l} \text{if } n = 1, \ m = 0 \\ \text{or } n = 0, \ m = 1 \end{array} \\ 0 & \text{otherwise} \end{cases}$$

Let V_1^n be the number of class-1 packets that arrive at the system during the service time of a packet that leaves the system at departure instant n, $n = 0, 1, \ldots$. Similarly, let V_2^n be the number of class-2 packets that arrive at the system during the service time of a packet that leaves the system at departure n. Clearly, V_1^n is given by

$$(4) \qquad V_1^n = \sum_{i=0}^{\sum_{j=1}^{X^n} H_1^{(j)}} A_1^{(i)}$$

where $H_1^{(j)}$ denotes the number of class-1 batches to enter the system in the jth slot of the service time X of the packet that leaves the system at departure instant n. Similarly, V_2^n is given by

$$(5) \qquad V_2^n = \sum_{i=0}^{\sum_{j=1}^{X^n} H_2^{(j)}} A_2^{(i)}$$

where $H_2^{(j)}$ denotes the number of class-2 batches to enter the system in the jth slot of the service time X of the packet that leaves the system at departure instant n.

Let $P_1^n = 0, 1, \cdots$, be a RV denoting the class-1 system occupancy just after departure n. Similarly, let $P_2^n = 0, 1, \cdots$, be a RV denoting the class-2 system occupancy just after departure n. Also, let $D_1^n = 0, 1, \cdots$, be a RV denoting the number of *class-1* packets that depart from the system at departure n. The conditional distribution of D_1^n is given by

$$(6) \qquad \Pr\left[D_1^{n+1} = i | P_1^n = j\right] = \begin{cases} 1 & \begin{array}{l} \text{if } i = 1, j > 0 \\ \text{or } i = 0, j = 0 \end{array} \\ 0 & \text{otherwise} \end{cases}$$

Similarly, let $D_2^n = 0, 1, \cdots$, be a RV denoting the number of *class-2* packets that depart from the system at departure n. The conditional distribution of D_2^n is given by

$$(7) \qquad \Pr\left[D_2^{n+1} = i | P_1^n = j, P_2^n = l\right] = \begin{cases} 1 & \begin{array}{l} \text{if } i = 1, j = 0, l > 0 \\ \text{or } i = 0, j = 0, l = 0 \\ \text{or } i = 0, j > 0, l \geq 0 \end{array} \\ 0 & \text{otherwise} \end{cases}$$

Using the RVs defined above, taking into account their stated conditional distributions and referring to the assumptions in section 2, the system occupancy at departure $n + 1$ is given by

$$(8) \qquad P_i^{n+1} = P_i^n - D_i^{n+1} + V_i^{n+1}, \qquad i = 1, 2.$$

Note that the pairs (P_1^n, P_2^n), $n = 0, 1, \cdots$, define a two-dimensional Markov chain [3], with P_1^n and P_2^n being dependent for each n. Now we will embark on deriving the PGF of the (P_1^n, P_2^n), from which the stationary PGF will be obtained.

Let $p_{i,j}^n$ be the joint distribution of the pair (P_1^n, P_2^n). That is, $p_{i,j}^n = \Pr\left[P_1^n = i, P_2^n = j\right]$. And let $P^n(z_1, z_2)$ be the PGF of $p_{i,j}^n$. That is,

$$(9) \qquad P^n(z_1, z_2) \triangleq \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{i,j}^n z_1^i z_2^j = E\left[z_1^{P_1^n} z_2^{P_2^n}\right],$$

where the notation $E[\cdot]$ denotes the expectation of the RV between the brackets. Using (8) in (9), we get

$$(10) \qquad P^{n+1}(z_1, z_2) = E\left[z_1^{V_1^{n+1}} z_2^{V_2^{n+1}}\right] E\left[z_1^{P_1^n - D_1^{n+1}} z_2^{P_2^n - D_2^{n+1}}\right].$$

The separation of the expectations in (10) is a consequence of the fact that both A_1^{n+1} and A_2^{n+1} are independent of P_1^n, D_1^{n+1}, P_2^n and D_2^{n+1}.

The first factor in (10) is just $V^{n+1}(z_1, z_2)$. The second factor can be evaluated using the conditional distributions given in (6) and (7), and employing some tedious,

but straightforward, first probability principles. After substituting for the first factor and evaluating the second, (10) becomes

$$P^{n+1}(z_1, z_2) = \frac{V^{n+1}(z_1, z_2) P^n(z_1, z_2)}{z_1} + V^{n+1}(z_1, z_2) P^n(0, z_2) \left(\frac{1}{z_2} - \frac{1}{z_1}\right)$$

$$\text{(11)} \qquad\qquad + V^{n+1}(z_1, z_2) p_{0,0}^n \left(1 - \frac{1}{z_2}\right)$$

If the arrival rate into the system is strictly less than the service rate, then the system will reach steady state after a large number of slots. That is if $r < s$, then as $n \to \infty$, the PGFs $P^n(z_1, z_2)$ will converge to the PGF $P(z_1, z_2)$, the PGFs $V^n(z_1, z_2)$ will converge to the common PGF $V(z_1, z_2)$, the functions $P^n(0, z_2)$ will converge to the function $P(0, z_2)$, and the sequence $p_{0,0}^n$ will converge to $p_{0,0}$. Thus, in steady state (11) becomes

$$\text{(12)} \qquad P(z_1, z_2) = \frac{V(z_1, z_2)(p_0 z_1 (z_2 - 1) + P(0, z_2)(z_1 - z_2))}{z_2(z_1 - V(z_1, z_2))}$$

where $p_0 = p_{0,0}$.

Now, we can find $V(z_1, z_2)$ in (12) by using (4) and (5) as follows.

$$V(z_1, z_2) \triangleq E\left[z_1^{V_1^n} z_2^{V_2^n}\right]$$

$$= E\left[z_1^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_1^{(j)}} A_1^{(i)}} z_2^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_2^{(j)}} A_2^{(i)}}\right].$$

By expanding the expectation into a conditional expectation, we get

$$V(z_1, z_2) = E\left[z_1^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_1^{(j)}} A_1^{(i)}} z_2^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_2^{(j)}} A_2^{(i)}} \;\middle|\; H_1^{(j)} = 1, H_2^{(j)} = 0\right]$$

$$\cdot \Pr[H_1^{(j)} = 1, H_2^{(j)} = 0]$$

$$+ E\left[z_1^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_1^{(j)}} A_1^{(i)}} z_2^{\sum_{i=0}^{\sum_{j=1}^{X^n} H_2^{(j)}} A_2^{(i)}} \;\middle|\; H_1^{(j)} = 0, H_2^{(j)} = 1\right]$$

$$\text{(13)} \qquad\qquad \cdot \Pr[H_1^{(j)} = 0, H_2^{(j)} = 1].$$

From (1), (2) and (3), we can see that $\Pr[H_1^{(j)} = 1, H_2^{(j)} = 0] = \lambda$ and $\Pr[H_1^{(j)} = 0, H_2^{(j)} = 1] = \overline{\lambda}$, for each $j = 1, 2, \cdots, X^n$. Using these probabilities in (13), we get after some straightforward manipulation that

$$\text{(14)} \qquad\qquad V(z_1, z_2) = \lambda X(A_1(z_1)) + \overline{\lambda} X(A_2(z_2)),$$

where $X(.)$ is the common PGF of the X^k, the RV denoting the service time of the packet that enters service in slot k, $A_1(.)$ the common PGF of the A_1^k, and $A_2(.)$ the common PGF of the A_2^k. It should be noted from the assumptions that $X(.)$ is

$$X(z) \triangleq \sum_{n=1}^{\infty} \Pr\left[X^k = n\right] z^n$$

$$\text{(15)} \qquad\qquad = \frac{sz}{1 - \overline{s}z}$$

Second, we can find p_0 in (12) by arguing as follows. The probability that a queueing system, regardless of whether one-class or two-class, is empty is always the complement of the probability that it is busy, and the latter is known (see e.g. [19]) to be just the utilization ρ, defined as the ratio of packet arrival rate to packet service rate. That is, the probability p_0 that a queueing system is empty is always given by

$$(16) \qquad p_0 = 1 - \frac{r}{s},$$

where r in our system is

$$r = \lambda A_1'(1) + \overline{\lambda} A_2'(1),$$

where $A_1'(1) = dA_1(z)/dz|_{z=1}$ denotes the expected class-1 batch size, and $A_2'(1)$ the expected class-2 batch size.

Substituting for $V(z_1, z_2)$ from (14) and for p_0 from (16) into (12), we get

(17)
$$P(z_1, z_2) = \frac{\left(\lambda X(A_1(z_1)) + \overline{\lambda} X(A_2(z_2))\right)\left[(s-r)z_1(z_2-1) + sP(0,z_2)(z_1-z_2)\right]}{sz_2\left(z_1 - \lambda X(A_1(z_1)) - \overline{\lambda} X(A_2(z_2))\right)},$$

To find the function $P(0, z_2)$ in (17), we proceed as follows. First, it can be easily shown using Rouche's Theorem [20, p.123] that the denominator of (17) has exactly one zero on the unit disk $|z_1| < 1$. Let this zero be denoted by ξ (a function of z_2). Then, since $P(z_1, z_2)$ is a generating function, it must be analytic on the unit disk. That is, ξ must also be a zero of the numerator of (17). As a consequence, setting $z_1 = \xi$ in that numerator makes the latter equal to zero, which enables us to get

$$(18) \qquad P(0, z_2) = -\frac{(s-r)(z_2-1)\xi}{s(\xi - z_2)}, \qquad \xi - z_2 \neq 0.$$

Using (18) in (17), we get

$$(19) \qquad P(z_1, z_2) = \frac{\left(\lambda X(A_1(z_1)) + \overline{\lambda} X(A_2(z_2))\right)(s-r)(z_2-1)(z_1-\xi)}{s\left(z_1 - \lambda X(A_1(z_1)) - \overline{\lambda} X(A_2(z_2))\right)(z_2 - \xi)}$$

The zero ξ can itself be found using Lagrange's Theorem [20, p.123] as follows.

$$(20) \qquad \xi = \sum_{k=1}^{\infty} \frac{1}{k!}\left[\frac{\partial^{k-1}}{\partial z_1^{k-1}}\{-\lambda X(A_1(z_1)) - \overline{\lambda} X(A_2(z_2))\}^k\right]_{z_1=0}$$

For comparison purposes, it should be noted that the corresponding joint PGF assuming preemptive priority can be found using the same procedure (but at the slot boundaries, rather than the departure instants) to be

$$(21) \qquad P(z_1, z_2) = \frac{\left(\lambda A_1(z_1) + \overline{\lambda} A_2(z_2)\right)(s-r)(z_2-1)(z_1-\xi)}{\left[z_1 - \left(\lambda A_1(z_1) + \overline{\lambda} A_2(z_2)\right)(s + \overline{s}z_1)\right](z_2 - \xi)}$$

It is worth noting that if we set $s = 1$ either in the nonpreemptive result (19) or the preemptive result (21) we get exactly the same occupancy function, namely

$$(22) \qquad P(z_1, z_2) = \frac{\bar{r}\left(\lambda A_1(z_1) + \bar{\lambda}A_2(z_2)\right)(z_2 - 1)(z_1 - \xi)}{\left(z_1 - \lambda A_1(z_1) - \bar{\lambda}A_2(z_2)\right)(z_2 - \xi)},$$

where $\bar{r} = 1 - r$. It is also worth noting that this is a mathematical demonstration of the fact that when the service time is deterministically 1 slot (or equivalently the service rate is 1 packet per slot), it does not make a difference whether the priority discipline used is preemptive or nonpreemptive. This fact is by the way usually justified intuitively as follows. The service of a class-2 packet cannot be preempted so as the packet is sent in the next slot back to the queue, since by the next slot that packet will have ended service and departed any way. It is further worth noting that (22) is just Equation 7 in [21] after substituting $c = 1$ (i.e. one server). It is lastly worth noting that the PGF $\mathcal{P}(z)$ of the occupancy $\mathcal{P}$ of a system equivalent to ours but with a single class of packets can be easily obtained from (19) by setting $\lambda = 1$ and $z_1 = z_2 = z$, which amounts to assuming that only class-1 packets arrive into the queue. Denoting the size of each batch by A (instead of A_1), we get the classical $Geo^A/Geo/1$ result [19].

$$
\begin{aligned}
\mathcal{P}(z) &= P(z_1, z_2)|_{\lambda=1, z_1=z_2=z, A_1=A} \\
&= \frac{A(z)(z-1)(s - A'(1))}{z - A(z)(s + \bar{s}z)}.
\end{aligned}
$$
(23)

4. Marginal PGFs. In this section, the joint PGF (17) will be used to obtain the marginal PGFs of class-1 and class-2 occupancies. First, the marginal PGF of the class-1 occupancy P_1 can be obtained from (17) as follows

$$
\begin{aligned}
P_1(z) &= P(z, 1) \\
&= \frac{\left(\lambda A_1(z) - \bar{s}A_1(z) + \bar{\lambda}\right)P(0,1)(z-1)}{z - \bar{\lambda} - A_1(z)(\lambda + \bar{s}z - \bar{s})}
\end{aligned}
$$
(24)

Noting that $P(0,1)$ denotes the probability that the system is class-1 empty, this probability can be obtained by employing the normalization condition $P_1(1) = 1$, getting

$$(25) \qquad P(0,1) = \frac{s - r_1}{s}$$

Substituting in (24), we get

$$(26) \qquad P_1(z) = \frac{\left(\lambda A_1(z) - \bar{s}A_1(z) + \bar{\lambda}\right)(s - r_1)(z - 1)}{s\left(z - \bar{\lambda} - A_1(z)(\lambda + \bar{s}z - \bar{s})\right)}.$$

For comparison purposes, the marginal PGF of the class-1 occupancy $\widehat{P}_1$ in a preemptive system equivalent to ours can be shown, using the same procedure, to be

$$(27) \qquad \widehat{P}_1(z) = \frac{\left(\bar{\lambda} + \lambda A_1(z)\right)(s - r_1)(z - 1)}{z - \left(\bar{\lambda} + \lambda A_1(z)\right)(s + \bar{s}z)}.$$

It is worth noting that if we set $\lambda = 1$ in either (26) to (27) we should get the same result since we would be talking about the same one-class system. Sure enough, in either case we get (23) above.

Similarly, the marginal PGF of the class-2 system occupancy $P_2(z)$ can be obtain by putting $z_1 = 1$ and $z_2 = z$ in (17), yielding

$$
(28) \quad
\begin{aligned}
P_2(z) &= P(1, z) \\
&= \frac{[\lambda - \bar{s}\lambda A_2(z) + s\bar{\lambda} A_2(z)](z-1)[(s-r) - sP(0,z)]}{sz\bar{\lambda}(1 - A_2(z))}.
\end{aligned}
$$

Substituting for $P(0, z)$ from (18), then (28) yields

$$
(29) \quad P_2(z) = \frac{\left(\lambda - \bar{s}\lambda A_2(z) + s\bar{\lambda} A_2(z)\right)(s-r)(1-\xi)(z-1)}{s\bar{\lambda}(1 - A_2(z))(z-\xi)}.
$$

For comparison purposes, the marginal PGF of the class-2 occupancy $\widehat{P}_2$ in a preemptive system equivalent to ours can be shown, using the same procedure, to be

$$
(30) \quad \widehat{P}_2(z) = \frac{\left(\lambda + \bar{\lambda} A_2(z)\right)(s-r)(1-\xi)(z-1)}{\bar{\lambda}(1 - A_2(z))(z-\xi)}.
$$

5. Expected Occupancies. In this section we obtain expressions for the expectation of the class-1 and class-2 occupancies. First, the expected class-1 occupancy can be obtained from (26) by evaluating the first derivative at 1 as follows.

$$
(31) \quad E[P_1] = P_1'(1) = \frac{r_1 - \bar{s}A_1'(1)}{s} + \frac{\lambda A_1''(1) + 2\bar{s}A_1'(1)}{2(s - r_1)}.
$$

For comparison purposes, the expected class-1 occupancy in the preemptive case can be obtained from (27) yielding

$$
(32) \quad E\left[\widehat{P}_1\right] = r_1 + \frac{\lambda A_1''(1) + 2\bar{s}r_1}{2(s - r_1)}.
$$

By comparing (31) with (32), we find that the relation between the class-1 occupancies in the nonpreemptive and preemptive cases to be as follows.

$$
(33) \quad E[P_1] = E\left[\widehat{P}_1\right] + \frac{\bar{s}r_1\bar{\lambda}A_1'(1)}{s(s - r_1)}
$$

That is, the expected number of class-1 packets in a nonpreemptive system is higher than that in an equivalent preemptive system. This result is intuitively true, as class-1 packets will have to wait in a nonpreemptive system until the class-2 packet currently in service departs, growing in number during the wait due to the continued arrivals of packets of the same class.

Second, the expected class-2 occupancy can be obtained indirectly as follows.

$$
(34) \quad E[P_2] = E[P] - E[P_1],
$$

where $E\left[P\right]$ denotes the expected total system occupancy and can be obtained from (17) as follows. First, the PGF $P\left(z\right)$ of total number of packets existing in the system in steady state can be obtained by putting $z_1 = z_2 = z$ in (17), to get

$$(35) \quad P\left(z\right) = \frac{\left[\lambda A_1\left(z\right) + \overline{\lambda} A_2\left(z\right) - \overline{s} A_1\left(z\right) A_2\left(z\right)\right]\left(s - r\right)\left(z - 1\right)}{z\left(1 - \overline{s} A_1\left(z\right)\right)\left(1 - \overline{s} A_2\left(z\right)\right) - s\lambda A_1\left(z\right) - s\overline{\lambda} A_2\left(z\right) + s\overline{s} A_1\left(z\right) A_2\left(z\right)}$$

With some effort it can be shown that

$$\begin{aligned} E\left[P\right] &= P'\left(1\right) \\ &= \frac{r - \overline{s}(A_1'(1) + A_2'(1))}{s} \\ &\quad + \frac{2s\overline{s}\left(A_1'\left(1\right) + A_2'\left(1\right)\right) + s\left(\lambda A_1''\left(1\right) + \overline{\lambda} A_2''\left(1\right)\right) - 2\overline{s} A_1'\left(1\right) A_2'\left(1\right)}{2s\left(s - r\right)} \end{aligned}$$
(36)

Substituting for $E\left[P_1\right]$ and $E\left[P\right]$ from (31) and (36) in (34), we get

$$\begin{aligned} E\left[P_2\right] &= \frac{r_2 - \overline{s} A_2'(1)}{s} - \frac{\lambda A_1''\left(1\right) + 2\overline{s} A_1'\left(1\right)}{2\left(s - r_1\right)} \\ &\quad + \frac{2s\overline{s} A_1'\left(1\right) + 2s\overline{s} A_2'\left(1\right) + s\lambda A_1''\left(1\right) + s\overline{\lambda} A_2''\left(1\right) - 2\overline{s} A_1'\left(1\right) A_2'\left(1\right)}{2s\left(s - r\right)} \end{aligned}$$
(37)

For comparison purposes, the expected class-2 occupancy $E\left[\widehat{P}_2\right]$ in the preemptive case can be obtained from (21), (32) and (34) using the same procedure getting

$$(38) \quad E\left[\widehat{P}_2\right] = r_2 + \frac{2\overline{s} r + \lambda A_1''\left(1\right) + \overline{\lambda} A_2''\left(1\right)}{2\left(s - r\right)} - \frac{\lambda A_1''\left(1\right) + 2\overline{s} r_1}{2\left(s - r_1\right)}$$

By comparing (37) with (38) we find the relation between the class-2 occupancies in the nonpreemptive and its counterpart in the preemptive case to be

$$(39) \quad E\left[P_2\right] = E\left[\widehat{P}_2\right] - \frac{\overline{s}}{s}\left(\frac{r_1 \overline{\lambda} A_1'\left(1\right)}{s - r_1} + \frac{A_1'\left(1\right) A_2'\left(1\right) - r\left(\overline{\lambda} A_1'\left(1\right) + \lambda A_2'\left(1\right)\right)}{s - r}\right)$$

That is, the expected number of class-2 packets in a nonpreemptive system is less than that in an equivalent preemptive system. This result is intuitively true, as class-2 packets will not be preempted during service by arriving class-1 packets, decreasing in number as more of them leave the system without returning back to the queue.

As a final observation in this section, it can be seen from the above that while the expected class-2 occupancy is a function of both the class-1 and the class-2 arrival rates, the expected class-1 occupancy is a function of only the class-1 arrival rate

6. Special Cases. In this section, we consider special cases for the distribution of the arriving batch sizes. Specifically, we will first assume that the batches have binomial distribution, then later we will assume that the batches have Poisson distribution.

6.1. Binomial distribution. Let us assume that the batch sizes A_1 and A_2 have Binomial distribution with parameters μ_1 and μ_2, respectively. That is

$$(40) \qquad \Pr\left[A_1 = i\right] = \binom{N}{i} \mu_1^i \overline{\mu}_1^{N-i}, \qquad i \le N$$

and

$$(41) \qquad \Pr\left[A_2 = i\right] = \binom{N}{i} \mu_2^i \overline{\mu}_2^{N-i}, \qquad i \le N$$

The PGF $A_1(z)$ is

$$(42) \qquad A_1(z) = \left(\overline{\mu}_1 + \mu_1 z\right)^N$$

Similarly, the PGF $A_2(z)$ is

$$(43) \qquad A_2(z) = \left(\overline{\mu}_2 + \mu_2 z\right)^N$$

From (42), (43) and (31), we get

$$(44) \qquad E\left[P_1\right] = \frac{r_1 - \overline{s}N\mu_1}{s} + \frac{r_1\mu_1\left(N-1\right) + 2\overline{s}N\mu_1}{2\left(s - r_1\right)}.$$

where $r_1 = \lambda N\mu_1$ is the class-1 arrival rate. In the preemptive case we find from (32) that

$$(45) \qquad E\left[\widehat{P}_1\right] = r_1 + \frac{r_1\left(\left(N-1\right)\mu_1 + 2\overline{s}\right)}{2\left(s - r_1\right)}.$$

For visual illustration, we now plot for these binomial arrivals some of the class-1 expected occupancies for both the nonpreemptive and preemptive disciplines. First in Figure (1) we plot the expected occupancy of class-1 packets both in the preemptive and nonpreemptive cases against the class-1 packet arrival rate, r_1. It can be seen that as the rate increases, both occupancies increase which is logical.

Note that we will not repeat the same plot against r_2 since, as we noted above, the class-1 occupancy is not a function of r_2. Lastly, in Figure (2) we plot the expected class-1 occupancy both in the preemptive and nonpreemptive cases against the service rate, s. As can be seen, both occupancies decrease as the service rate increases, which is logical. The interesting observation is that the two curves get ever nearer as s approaches 1, meaning that the difference between the two disciplines becomes immaterial at high service rates. When $s = 1$, the two curves should meet at a point as was proven above mathematically.

The average of class-2 system occupancy can be obtained from (37) as follows.

$$
\begin{aligned}
E\left[P_2\right] = {} & \frac{r_2 - \overline{s}N\mu_2}{s} - \frac{r_1\mu_1\left(N-1\right) + 2\overline{s}\mu_1 N}{2\left(s - Nr_1\right)} \\
& + \frac{2s\overline{s}N\left(\mu_1 + \mu_2\right) + s\left(N-1\right)\left(r_1\mu_1 + r_2\mu_2\right) - 2\overline{s}\mu_1\mu_2 N^2}{2s\left(s - r\right)}
\end{aligned}
$$

(46)

$N = 100$, $s = 0.75$, $\mu_2 = 0.002$, $\lambda = 0.5$

Nonpreemptive

Preemptive

Expected class-1 occupancy (packets)

Class-1 arrival rate, r_1 (packets per slot)
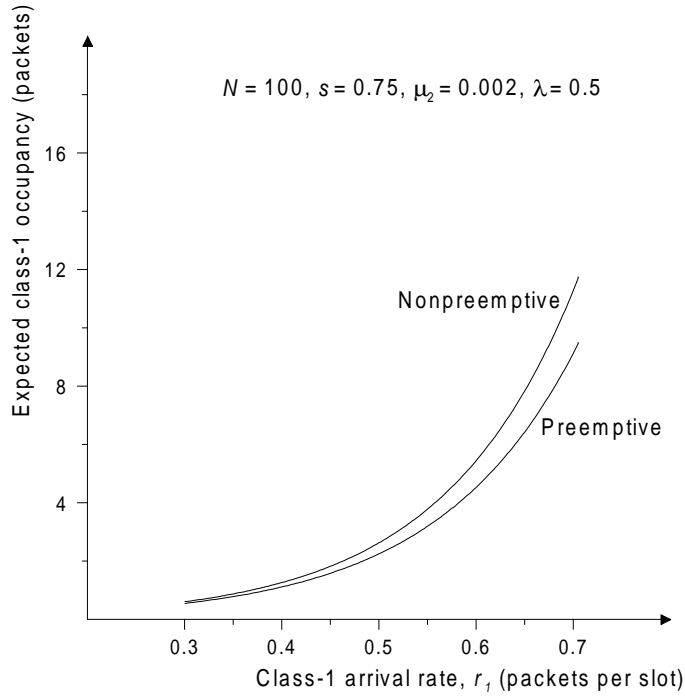
FIG. 1. *Class-1 system occupancy against class-1 packet arrival rate.*

$N = 100$, $\mu_1 = 0.006$, $\mu_2 = 0.002$, $\lambda = 0.5$

Nonpreemptive

Preemptive

Expected class-1 occupancy (packets)

Service rate, s (packets per slot)
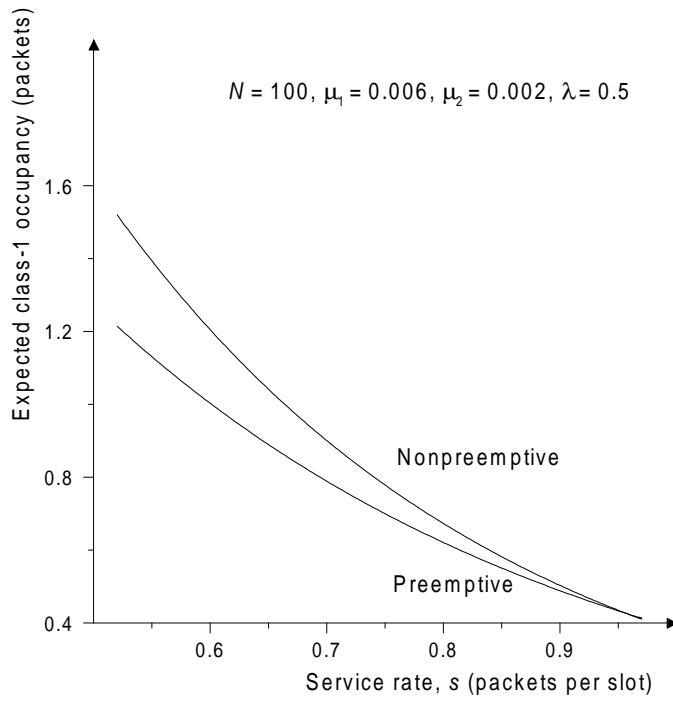
FIG. 2. *Class-1 system occupancy against service time*
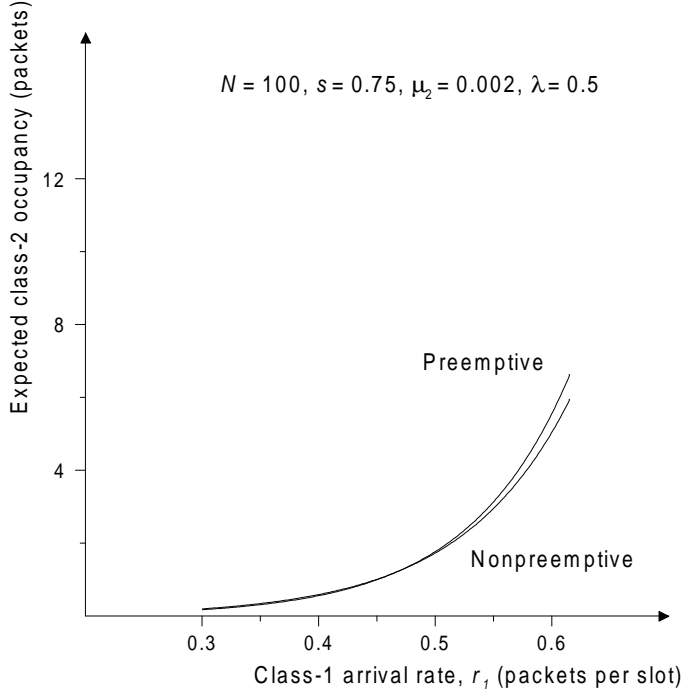
FIG. 3. *Class-2 system occupancy against class-1 arrival rate*

where $r_2 = \overline{\lambda} N \mu_2$ is the class-1 arrival rate. In the preemptive case we find from (38) that

$$E\left[\widehat{P_2}\right] = r_2 + \frac{2\overline{s}r + (N-1)(r_1\mu_1 + r_2\mu_2)}{2(s-r)} - \frac{r_1((N-1)\mu_1 + 2\overline{s})}{2(s-r_1)}.$$

For visual illustration, we now plot for these binomial arrivals some of the class-2 expected occupancies for both the nonpreemptive and preemptive disciplines. First in Figure (3) we plot the expected occupancy of class-2 packets both in the preemptive and nonpreemptive cases against the class-1 packet arrival rate, r_1.

It can be seen that as the rate increases, both occupancies increase which is logical. Second, in Figure (4) we plot the expected class-2 occupancy both in the preemptive and nonpreemptive cases against the class-2 arrival rate, r_2.

It can be seen that as the rate increases, both occupancies increase which is also logical. Lastly, in Figure (5) we plot the expected class-2 occupancy both in the preemptive and nonpreemptive cases against the service rate, s.

As was noted in the class-1 occupancy plots, it can be seen that both occupancies decrease as the service rate increase, with the observation that the two curves get ever nearer as s approaches 1, making the choice of either discipline irrelevant at high service rate.
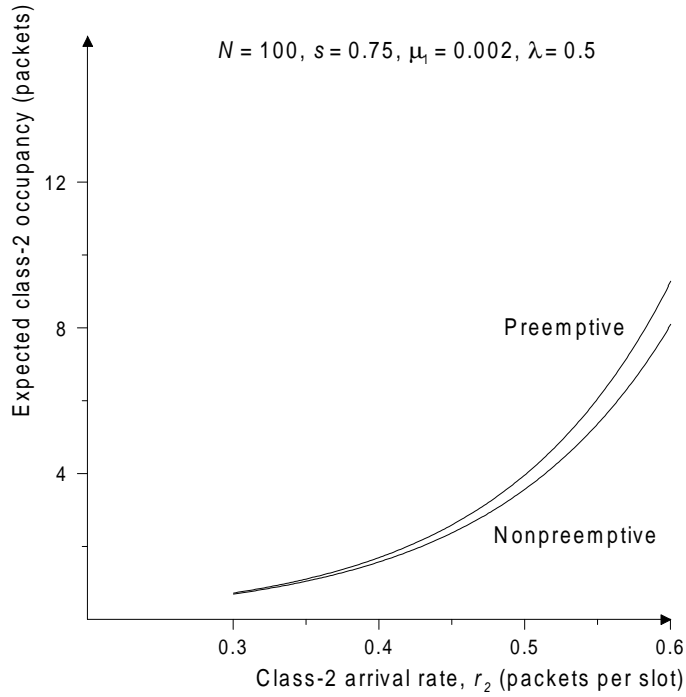
$N = 100$, $s = 0.75$, $\mu_1 = 0.002$, $\lambda = 0.5$

Preemptive

Nonpreemptive

Class-2 arrival rate, r_2 (packets per slot)

Expected class-2 occupancy (packets)

FIG. 4. *Class-2 system occupancy against class-2 arrival rate*

$N = 100$, $\mu_1 = 0.006$, $\mu_2 = 0.002$, $\lambda = 0.5$

Preemptive

Nonpreemptive

Service rate, s (packets per slot)

Expected class-2 occupancy (packets)
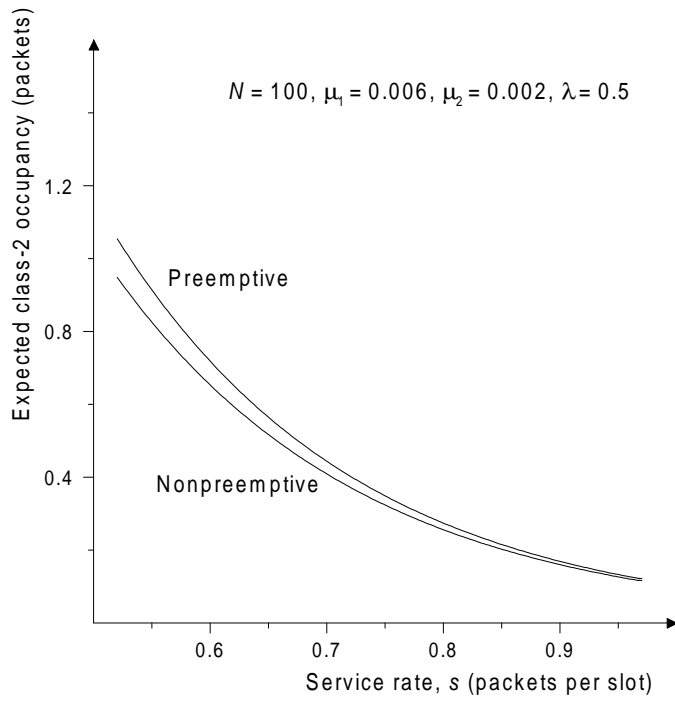
FIG. 5. *Class-2 system occupancy against service time*

6.2. Poisson distribution. Let us now assume that the batch sizes A_1 and A_2 have Poisson distribution with parameters μ_1 and μ_2 respectively. That is

$$\text{(47)} \qquad \Pr\left[A_1 = n\right] = \frac{(\mu_1)^n}{n!} e^{-\mu_1} \text{ for } n = 0,\, 1,\, \ldots$$

and

$$\text{(48)} \qquad \Pr\left[A_2 = n\right] = \frac{(\mu_2)^n}{n!} e^{-\mu_2} \text{ for } n = 0,\, 1,\, \ldots$$

The PGFs are then

$$\text{(49)} \qquad A_1\left(z\right) = e^{\mu_1(z-1)}$$

and

$$\text{(50)} \qquad A_2\left(z\right) = e^{\mu_2(z-1)}$$

Thus, from (49), (50) and (31), we get

$$\text{(51)} \qquad E\left[P_1\right] = \frac{r_1 - \bar{s}\mu_1}{s} + \frac{r_1\mu_1 + 2\bar{s}\mu_1}{2\left(s - r_1\right)}.$$

where $r_1 = \lambda\mu_1$ is the class-1 arrival rate.

In the preemptive case we find from (32) that

$$\text{(52)} \qquad E\left[\widehat{P}_1\right] = r_1 + \frac{r_1\mu_1 + 2\bar{s}r_1}{2\left(s - r_1\right)}.$$

The average of class-2 system occupancy can be obtained from (37), as follows

$$\text{(53)} \quad E\left[P_2\right] = \frac{r_2 - \bar{s}\mu_2}{s} + \frac{2s\bar{s}\left(\mu_1 + \mu_2\right) + s\left(r_1\mu_1 + r_2\mu_2\right) - 2\bar{s}\mu_1\mu_2}{2s\left(s - r\right)} - \frac{r_1\mu_1 + 2\bar{s}\mu_1}{2\left(s - r_1\right)}$$

where $r_2 = \bar{\lambda}\mu_2$ is the class-2 arrival rate. In the preemptive case we find from (38) that

$$\text{(54)} \qquad E\left[\widehat{P}_2\right] = r_2 + \frac{2\bar{s}r + r_1\mu_1 + r_2\mu_2}{2\left(s - r\right)} - \frac{r_1\mu_1 + 2\bar{s}r_1}{2\left(s - r_1\right)}.$$

7. Conclusion. We have analyzed the system occupancy of a two-class discrete queue suitable for modelling computers and digital communications systems handling packets whose sources are multimedia applications. The queue arrivals are batches of homogenous packets where the batch size is a random variable of arbitrary distribution, and the queue service time is geometric, with priority for class-1 packets over class-2 packets. Two priority disciplines are considered, the nonpreemptive and the preemptive. We have shown the relations between them, and also shown that as the service rate approaches 1 packet per slot, the two disciplines act identically the same.

This work can be extended in many ways as follows. The packet waiting room can be made finite. The service time can be assumed general. The server can serve

the packets with two service rates, one for each class. The arriving batches can be nonhomogeneous, i.e. of mixed classes. The arrival and service rates can be taken state dependent. In all these variations, the present work can serve as a basis.

REFERENCES

[1] C. KRUSKAL, M. SNIR, AND A. WEISS, *The Distribution of Queuing Times in Clocked Multi-stage Interconnection Networks*, IEEE Trans. on Computers, 37(1988), pp. 1347–1356.

[2] H. BRUNEEL, *Performance of Discrete-Time Queueing Systems*, Computers & Operations Research, 20:3(1993), pp. 303–320.

[3] B. STEYAERT AND Y. XIONG, *Analysis of a Discrete-Time Queue with General Three-State Markovian Traffic Sources*, Performance Evaluation, 27(1996), pp. 277–294.

[4] I. CIDON AND M. SIDI, *Dispersed Messages in Discrete-time Queues: Delay, Jitter and Threshold Crossing*, Performance Evaluation, 29(1997), pp. 85–104.

[5] M. CHAUDHRY AND U. GUPTA, *Modeling and Analysis of Discrete-Time Multiserver Queues with Batch Arrivals: $GI^X/Geo/m$*, INFORMS, Journal on Computing, 13:3(2001), pp. 172–180.

[6] T. TAKINE AND B. SENGUPTA, *An analysis of a Discrete-Time Queue for Broadband ISDN with Priorities Among Traffic Classes*, IEEE Trans. On Communications, 42(1994), pp. 1837–1843.

[7] F. ISHIZAKI AND T. TAKINE, *Analysis of a Discrete-Time Queue with Gated Priority*, Performance Evaluation, 23(1995), pp. 121–143.

[8] J. LEE AND C. UN, *Performance Analysis of Space-Priority Mechanisms in an Input and Output Queueing ATM Switch*, IEE Proceedings on Communications, 144(1997), pp. 229–236.

[9] Y. LEE AND B. CHOI, *Queueing System with Multiple Delay and Loss Priorities for ATM Networks*, Information Sciences, An International Journal, 138:1-4(2001), pp. 7–29.

[10] M. TAKAHASHI, H. OSAWA, AND T. FUJISAWA, *$Geo^X/G/1$ Retrial Queue with Nonpreemptive Priority*, Asia Pacific Journal of Operational Research, 16:2(1999), pp. 215–234.

[11] J. WALRAEVENS, B. STEYAERT, AND H.BRUNEEL, *Performance Analysis of The System Contents in a Discrete-Time Nonpreemptive Priority Queue with General Service Times*, Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL), pp. 140-147, Jan-Feb 2000.

[12] H. NASSAR AND H. AL MAHDY, *Queueing Analysis of an ATM Multimedia Multiplexer with Nonpreemptive Priority*, IEE Proceedings on Communications, 150:3(2003), pp. 189–196.

[13] Y. LEE, D. SUNG, B. CHOI, AND D. CHOI, *Priority Queueing System with Fixed-Length Packet-Train Arrivals*, IEE Proceedings on Communications, 145(1998), pp. 331–336.

[14] Y. LEE, *Discrete-Time $Geo^X/G/1$ Queue with Preemptive Resume Priority*, Mathematical and Computer Modelling, 34:3-4(2001), pp. 243–250.

[15] H. NASSAR AND H. AL MAHDY, *A Priority Discrete Queueing Model for Multimedia Multiplexers*, Mathematical and Computer Modelling of Dynamical Systems, 8:2(2002), pp. 199–211.

[16] H. NASSAR AND M. ALI AHMED, *Performance Analysis of an ATM Buffered Switch Transmitting Two-Class Traffic over Unreliable Channels*, International Journal of Electronics and Communications (AEU), 57:3(2003), pp. 190–200.

[17] H. Nassar and M. Ali Ahmed, *Waiting Time Distribution of an ATM Switch Routing Two-class Traffic with Two Service Rates*, International Journal of Communication Systems, 16:7(2003), pp. 663–675.

[18] H. Nassar, *A Markov Model for Multibus Multiprocessor Systems under Asynchronous Operation*, Information Processing Letters, 54:4(1998), pp. 11–16.

[19] M. Woodward, *Communication and Computer Networks*. IEEE Computer Society Press, 1994.

[20] H. Takagi, *Queueing Analysis*, vol. 1. Elsevier Science Publishers B. V., 1991.

[21] H. Bruneel and K. Laevens, *Discrete-Time Multiserver Queues with Priorities*, Performance Evaluation, 33:4(1998), pp. 249–275.