Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Vol. 9, Issue 1 (June 2014), pp. 1-12

Applications and Applied Mathematics:
An International Journal
(AAM)

Delay Analysis of a Discrete-Time Non-Preemptive Priority Queue with Priority Jumps

Deepak C. Pandey and Arun K. Pal
Department of Mathematics
Statistics and Computer Science
G. B. Pant University of Agriculture and Technology
Pantnagar, U.S. Nagar, India - 263145
deechp2006@gmail.com; arun_pal1969@yahoo.co.in

Received: December 15, 2013; Accepted: April 8, 2014

Abstract

In this paper, we consider a discrete-time non-preemptive priority queueing model with priority jumps. Two classes, real-time (high priority) and non-real time (low priority), of traffic will be considered with providing jumps from lower priority traffic to the queue of high priority traffic. We derive expressions for the joint probability generating function of the system contents of the high and the low priority traffic in the steady state and also for some performance measures such as the mean value of the system contents and the packet delay. The behavior of the priority queues with priority jumps will be illustrated by using these results and is compared to the FIFO scheme.

Keywords: Discrete-time queueing theory; priority scheduling; generating functions; delay analysis

MSC 2010 No.: 90B22, 90B18, 68M20, 93A30

1. Introduction

In modern communication networks the different types of traffic require different QoS (Quality of Service) standards. In these networks we always deal with two types of traffic namely delay-sensitive traffic (i.e., voice and video) and delay-insensitive traffic (i.e., data). Since response time or delay is a crucial performance measure for delay-sensitive applications, time delays in priority queues have been studied extensively in recent years. Due to the sensitivity of the delay-sensitive traffic, it is a basic requirement for designing and constructing an efficient communication network with a very small mean delay and delay jitter. While the values of the important performance measures of the delay-insensitive traffic, namely loss ratio and throughput should also be small. Priority scheduling scheme is an important way by which we can achieve these requirements.

A lot of literature is available in the field of analysis of single class discrete time queueing system (see, for instance, Bruneel (1993); Cidon and Sidi (1997); Chaudhry and Gupta (2001); Steyaert and Xiong (1996), Haghighi *et al*. (2011) and the references therein) while the research work in this field of the analysis of two-class discrete queues have begun to proliferate only in recent years. Two-class discrete-time systems with service time assumed deterministically (1 slot) have been analyzed in Ishizaki and Takine (1995) and Lee and Choi (2001). Also, systems with non-preemptive priority and general service times with nonhomogeneous packet arrivals have been studied in Nassar and Al Mahdy (2003) and Takahashi *et al*. (1999). In a HOL non-preemptive priority scheduling, we assume that delay-sensitive traffic has priority over delay-insensitive traffic without preemption i.e., when a server becomes idle, a packet of delay-sensitive traffic, when available, will always be scheduled next but the service of a delay-insensitive packet which has already in service cannot be interrupted by a freshly arriving delay-sensitive packet. This priority scheduling scheme is not only very easy to implement but also provides relatively low delays for the delay sensitive traffic [see, for instance, Kleinrock (1975); Walraevens *et al.* (2003) and the references therein].

In the existing literature, there have been a number of contributions with respect to this priority scheme. The HOL non-preemptive priority queues have been widely discussed taking a variety of arrival and service time distributions in past [see, for instance, Rubin and Tsai (1989); Sugahara *et al.* (1995); Takine *et al*. (1994) and the references therein]. As a result priority given to the class-1 traffic the performance of class-2 traffic can be severely degraded. If the network is highly loaded and has a large proportion of the class-1 traffic, the HOL priority scheduling causes a large delay to the class-2 traffic. This condition known as the starvation problem can be solved with the help of dynamic priority schemes. The queue-length-threshold scheduling disciplines (QLT) is the one class which is studied in Choi and Lee (2001), Fratini (1990), Knessl *et al.* (2002) while the head-of-line with priority jumps (HOL-PJ) is the another class of dynamic priority scheme analyzed in Lim and Kobza (1990).

In this paper, we will consider a head-of-line priority scheme with priority jumps (HOL-PJ) in which class-2, i.e., delay-insensitive packets (at the HOL position) jump to (the end of) the class-1, i.e., delay-sensitive traffic queue. In literature, many jumping schemes with different criteria are discussed (see, for instance, Lim and Kobza (1990); Maertens *et al*. (2006a); Maertens *et al.* (2006b); Maertens *et al.* (2007); Maertens *et al.* (2008) and the references therein). We will consider and discuss a new jumping scheme in which class-2 packets jump to the high priority, class-1 queue, if there are no arrivals in the class-1 queue. Probability generating functions (pgfs) are used for the analysis, pgfs of the system contents of the class-1 and class-2 queue and the pgf of the delay of class-1 packet are derived. With the help of these pgfs higher moments can be easily obtained. The mean delay for class-2 packet is also obtained.

The outline of the paper is as follows. In section 2, the queueing model under study is described. In section 3, our focus is on the system contents and expressions for the pgf of the system contents are derived. The packet delay is analyzed in section 4, and again the pgf of the packet delay is obtained. In section 5, moments of the system contents and packet delay are calculated.

In section 6, some numerical examples are given and discussed. Finally, the paper is concluded in section 7.

2. Mathematical Model

We consider a discrete-time single-server queueing system with two queues having infinite buffer space. Two types of traffic are arriving in the system; viz. packets of class-1 that are stored in the first queue and packets of class-2 which are lined in the second. It is assumed that the time is slotted, where 1 slot equals the transmission time of a packet. We denote the number of arrivals of class-j during slot k by $a_{j,k}(j = 1,2)$. Both types of packet arrivals are assumed to be i.i.d. from slot-to-slot and are characterized by the joint probability mass function

$$a(m,n) \triangleq Prob\left[a_{1,k} = m, a_{2,k} = n\right],$$

and joint probability generating function (pgf) $A(z_1, z_2)$,

$$A(z_1, z_2) \triangleq E\left[z_1^{a_{1,k}} z_2^{a_{2,k}}\right] = \sum_{m,n=0}^{\infty} a(m,n)\, z_1^m\, z_2^n.$$

We denote the total number of arriving packets during slot k by $a_{T,k} \triangleq a_{1,k} + a_{2,k}$ and its pgf is given as $A_T(z) \triangleq E[z^{a_{T,k}}] = A(z,z)$. Further, we define the marginal pgf's of the number of arrivals from class-1 and class-2 during a slot by $A_1(z) \triangleq E[z^{a_{1,k}}] = A(z,1)$ and $A_2(z) \triangleq E[z^{a_{2,k}}] = A(1,z)$ respectively. We furthermore denote the arrival rate of class-j $(j = 1,2)$ by $\lambda_j = A_j^{'}(1)$ and the total arrival rate by $\lambda_T = A_T^{'}(1) = A_1^{'}(1) + A_2^{'}(1)$. The system has one server that provides the service of packets, at a rate of 1 packet per slot. It is assumed that the system is stable, i.e., $\lambda_T < 1$.

Due to the priority scheduling mechanism, it is as if class 1 packets (the high priority packets) are stored in front of class 2 packets (the low priority packets) in the queue. The low priority queue is provided service only in the case of being empty of the high priority queue. Class-1 packets are assumed to have non-preemptive priority over class-2 packets, and within one class the service discipline is FCFS. Since the priority scheduling is non-preemptive, service of a packet will not be interrupted by newly arriving packets.

Finally, the system is influenced by the following jumping mechanism: at the end of each slot in which a packet of class-1 queue is transmitted and in which class-2 packets arrive at the system, the packet at the HOL-position of the low priority queue jumps to the high priority queue.

3. System Contents

Let us denote the system contents of class-j at the beginning of slot k by $u_{j,k}$ $(j = 1,2)$ and the total system contents at the beginning of the slot k by $u_{T,k}$. The joint pgf of $u_{1,k}$ and $u_{2,k}$ denoted by $U_k(z_1, z_2)$ and given by

$$U_k(z_1, z_2) \triangleq E\left[z_1^{u_{1,k}}\, z_2^{u_{2,k}}\right].$$

The system contents of both types of packets can be obtained according to the following system equations:

If $u_{1,k} = 0$, then

$$\begin{cases} u_{1,k+1} = a_{1,k}\,, \\ u_{2,k+1} = [u_{2,k} - 1]^{+} + a_{2,k}. \end{cases} \tag{3.1}$$

If $u_{1,k} > 0$, then

If $a_{1,k} = 0$;

$$\begin{cases} u_{1,k+1} = u_{1,k}\,, \\ u_{2,k+1} = [u_{2,k} - 1]^{+} + a_{2,k}, \end{cases} \tag{3.2}$$

and if $a_{1,k} > 0$; then

$$\begin{cases} u_{1,k+1} = [u_{1,k} - 1]^{+} + a_{1,k}\,, \\ u_{2,k+1} = u_{2,k} + a_{2,k}\,, \end{cases} \tag{3.3}$$

where $[\cdot]^{+}$ denotes the maximum of the argument and zero. When the class 1 queue is empty at the beginning of slot k, a packet of class 2 queue (if any) is served during slot k (3.1). When the class 1 queue is non-empty at the beginning of the slot k a packet of class 1 queue is served. In the latter case, if $a_{1,k} > 0$, a class 2 i.e., low priority packet jumps at the end of slot k to the class 1 i.e., high priority queue [Equations (3.2) and (3.3)]. Using the above system equations the relationship between the joint pgf's of the system contents at the slots k and $k+1$ is obtained as follows:

$$U_{k+1}(z_1, z_2) = \frac{\begin{matrix} z_1(z_2-1)A(z_1,z_2)U_k(0,0)+z_1A(z_1,z_2)U_k(0,z_2)+z_1U_k(z_1,z_2)A(0,z_2) \\ -z_1U_k(0,z_2)A(0,z_2)+z_2U_k(z_1,z_2)A(z_1,z_2)-z_2U_k(z_1,z_2)A(0,z_2) \\ -z_2U_k(0,z_2)A(z_1,z_2)+z_2U_k(0,z_2)A(0,z_2) \end{matrix}}{z_1 z_2}. \tag{3.4}$$

The distribution of system contents, i.e., $U(z_1, z_2)$, in the case of steady-state is defined as

$$U(z_1, z_2) \triangleq \lim_{k\to\infty} U_k(z_1, z_2).$$

Applying the limit in Equation (3.4), we get the following formula for $U(z_1, z_2)$:

$$U(z_1, z_2) = \frac{z_1(z_2-1)A(z_1,z_2)U(0,0)+(z_1-z_2)\{A(z_1,z_2)-A(0,z_2)\}U(0,z_2)}{z_1z_2-z_2A(z_1,z_2)-(z_1-z_2)A(0,z_2)}. \tag{3.5}$$

For determining the two unknown quantities, namely $U(0, z_2)$ and $U(0, 0)$, we use Rouche's theorem and the normalization condition respectively (see, for instance, Maertens *et al.* (2007); Maertens *et al.* (2008); Walraevens *et al.* (2003) and the references therein) and finally obtains the joint pgf of the system contents at the beginning of a random slot in the steady state:

$$U(z_1,z_2)=\frac{(1-\lambda_T)(z_2-1)\{z_1A(z_1,z_2)\big(z_2-Y(z_2)\big)\big(z_2-A(0,z_2)\big)+z_2(z_1-z_2)A(Y(z_2),z_2)\big(A(z_1,z_2)-A(0,z_2)\big)\}}{\big(z_2-Y(z_2)\big)\big(z_2-A(0,z_2)\big)\big(z_1z_2-z_2A(z_1,z_2)-(z_1-z_2)A(0,z_2)\big)}, \tag{3.6}$$

with

$$Y(z)\triangleq A(Y(z),z)+\frac{(Y(z)-z)}{z}A(0,z). \tag{3.7}$$

The marginal pgfs $U_T(z)$, $U_1(z)$ and $U_2(z)$ of the total system contents and the system contents of class-j are obtained by putting appropriate values of z_1 and z_2, which are given as follows:

$$U_T(z)\triangleq \lim_{k\to\infty}E[z^{u_{T,k}}]=U(z,z)=\frac{(1-\lambda_T)A_T(z)(z-1)}{z-A_T(z)}, \tag{3.8}$$

$$U_1(z)\triangleq \lim_{k\to\infty}E[z^{u_{1,k}}]=U(z,1)=\frac{(1-\lambda_1)A_1(z)(z-1)}{A_1(0)\big(z-A_1(z)\big)}, \tag{3.9}$$

$$U_2(z)\triangleq \lim_{k\to\infty}E[z^{u_{2,k}}]=U(1,z)=\frac{(1-\lambda_T)(z-1)\{A_2(z)\big(z-Y(z)\big)\big(z-A(0,z)\big)+z(1-z)A(Y(z),z)\big(A_2(z)-A(0,z)\big)\}}{\big(z-Y(z)\big)\big(z-A(0,z)\big)\big(z-zA_2(z)-(1-z)A(0,z)\big)}, \tag{3.10}$$

where the quantity $Y(z)$ is obtained by using Equation (3.7).

4. Delay Analysis

The number of slots between the end point of the packet's arrival slot and the end point of its departure slot is defined as the packet delay. In other words it is the total amount of time that a packet spends in the system. In this section, the pgf expression of the packet delay of class 1 packets will be derived. Since a jump of the packets of class 2 to class 1 takes place at the end of the slot, the freshly arriving packets of class 1 are queued in front of packets that jump in the same slot. As a consequence, the packet delay of a tagged class 1 packet only depends on the system contents of queue 1 at the beginning of its arrival slot. This also means that the packet delay of a tagged class 1 packet can be treated as if it is the only type of packet in the system with only packets of class 1 arriving Bruneel & Kim (1993). Let the slot k be assumed to be the arrival slot of the tagged packet, the system contents of class 1 at the beginning of the slot k be denoted by $u_{1,k}$ and the total number of class 1 packets that arrive during slot k and which have to be served before the tagged packet are defined and denoted by $f_{1,k}$. Then the amount of time a tagged class 1 packet spends in the system is given by

$$d_1=[u_{1,k}-1]^+ + f_{1,k}+1. \tag{4.1}$$

Indeed, the tagged class 1 packet has to wait in queue 1 until all packets that were already in this queue at the moment of its arrival, are completely served. The number of these packets is obviously determined by all packets that were already present in queue 1 at the beginning of its arrival slot (potentially including class 2 packets which jumped to queue 1 before the tagged packet arrived) and all class 1 packets that arrived before the tagged packet in its arrival slot. The delay then equals this waiting time augmented with the service time of a packet, which equals 1. This leads to the above expression. Introducing pgf's yields

$$D_1(z) = E[z^{d_1}] = F_1(z)[U_1(z) + (z-1)U_1(0)]. \tag{4.2}$$

The fact that $u_{1,k}$ and $f_{1,k}$ are uncorrelated is used to obtain the above expression. The pgf of $f_{1,k}$ i.e., $F_1(z)$ can be calculated as in Bruneel and Kim (1993) and given by

$$F_1(z) = \frac{A_1(z)-1}{\lambda_1(z-1)}. \tag{4.3}$$

Using expressions (3.9) and (4.3) in (4.2), we finally find the pgf of the packet delay of class-1 packets is given by

$$D_1(z) = \frac{(1-\lambda_1)z(A_1(z)-1)}{\lambda_1 A_1(0)(z-A_1(z))}. \tag{4.4}$$

The analysis of the cell delay of a class-2 cell is more complicated. Consider a logically equivalent queueing system where all high priority cells are stored in front of the class-2 cells, and let us tag an arbitrary class-2 cell that arrives in the system. If k be the arrival slot of a tagged type 2 packet, $u_{T,k}$ be the total system contents at the beginning of the slot k, and $a_{1,k}$ and $f_{2,k}$ be the number of class 1 and class 2 packets which arrive during the slot k, but have to be served before the tagged packet then the total number of slots that a tagged type 2 packet spends in the system can give as

$$d_2 = [u_{T,k} - 1]^+ + a_{1,k} + f_{2,k} + p + 1, \tag{4.5}$$

where p being the number of type 1 packets which arrive during the slots following the tagged packet's arrival slot and due to the priority scheme these have to be served before the tagged one. We can calculate the packet delay of class 2 packets inspite of being complicated to obtain an explicit expression for its pgf. The expression for packet delay of the class 2 packets is obtained in the next section.

5. Calculation of Moments

For calculating the moments of the system contents and packet delays we will require the derivatives of the function $Y(z)$, defined in sections 3, for $z = 1$. These can be easily calculated in closed-form and the first derivative of $Y(z)$, at $z = 1$, i.e., $Y'(1)$ is given by

$$Y'(1) = \frac{\lambda_2 - A_1(0)}{1-\lambda_1 - A_1(0)}. \tag{5.1}$$

Let us define $\lambda_{i,j}$ and λ_{TT} as

$$\lambda_{i,j} = \left. \frac{\partial^2 A(z_1,z_2)}{\partial z_i \partial z_j} \right|_{z_1=z_2=1}$$

and

$$\lambda_{TT} = \frac{d^2 A_T(z)}{dz^2}\bigg|_{z=1},$$

with $i, j = 1, 2$. Now the mean value of the system contents and mean packet delay of class 1 can be calculated by taking the first derivative of the respective pgf's for $z = 1$. We find

$$E[u_1] = \lambda_1 + \frac{1}{A_1(0)}\left[\lambda_1\big(1 - A_1(0)\big) + \frac{\lambda_{11}}{2(1-\lambda_1)}\right], \qquad (5.2)$$

and

$$E[d_1] = 1 + \frac{1}{A_1(0)}\left[\big(1 - A_1(0)\big) + \frac{\lambda_{11}}{2\lambda_1(1-\lambda_1)}\right]. \qquad (5.3)$$

The mean value of the system contents of class 2, $E[u_T]$, can also be calculated by taking the first derivative of the respective pgf for $z = 1$ and also by using the following relation:

$$E[u_T] = E[u_1] + E[u_2], \qquad (5.4)$$

where $E[u_T]$ is the mean total system contents which can be obtained by taking first derivative of $U_T(z)$ for $z = 1$ and is given by

$$E[u_T] = \lambda_T + \frac{A_T''(1)}{2(1-\lambda_T)}. \qquad (5.5)$$

The mean packet delay of class 2 packets can be obtained by using the relation proposed in Maertens *et al.* (2008) and is given by

$$E[d_2] = \frac{E[u_T] - \lambda_1 E[d_1]}{\lambda_2}. \qquad (5.6)$$

As the values of $E[d_1]$ and $E[u_T]$ are already calculated, hence we can calculate the value of $E[d_2]$.

6. Numerical Example

In this section, the results obtained in the previous sections, are applied to an output-queueing switch having N inlets and N outlets, see in Walraevens *et al*. (2003). Two types of traffic are assumed: class 1traffic which is delay-sensitive (e.g., video, voice etc.) and class 2 traffic that is assumed to be delay-insensitive (e.g., data). The inlet packet arrivals are generated by i.i.d. Bernoulli processes with arrival rate λ_T. An arriving packet is assumed to be of class j ($j = 1,2$) with probability $\lambda_j/\lambda_T(\lambda_1 + \lambda_2 = \lambda_T)$. We assume the traffic of the two classes is arriving according to a two-dimensional binomial process. The joint pgf of the arriving traffic, $A(z_1, z_2)$, is given by:

$$A(z_1, z_2) = \left(1 - \frac{\lambda_1}{N}(1 - z_1) - \frac{\lambda_2}{N}(1 - z_2)\right)^N . \tag{6.1}$$

It is also noticed that if $N \to \infty$, the arrival process is supposed to be a superposition of two Poisson variates. The fraction of class-1 arrivals in the overall traffic mix is defined and denoted by α (i.e., $\alpha = \lambda_1/\lambda_T$). The remaining part of this section is analyzed by taking $N = 16$ and assuming that the service times for both classes are deterministic.

In Figure 1 the mean value of the system contents of class-1 and class-2 packets is shown as a function of the total arrival rate, when α =0.25, 0.50 and 0.75 respectively. One can easily see the influence of priority scheduling: the mean of the number of class-1 packets in the system is severely reduced by the HOL priority scheduling; the opposite holds for class-2 cells. In addition, it also becomes apparent that increasing the fraction of high priority cells in the overall mix increases the amount of high priority traffic while decreasing the amount of low priority traffic in the buffer. Finally, it is also clear that the impact of priority scheduling is more important if the load is high.

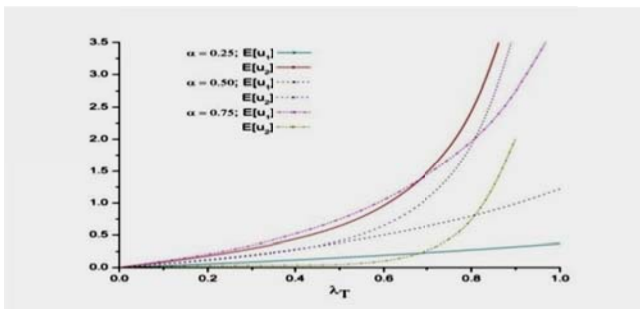

Figure 1: Mean value of system contents versus the total arrival rate.

Figure 2 shows the mean value of the packet delay as a function of the total load for α = 0.25, 0.50 and 0.75. To compare with FIFO scheduling, we have also shown the mean value of the packet delay in that case. The packet delay in this case for class-1 and class-2 packets is of course equal and can be easily calculated as if there is only one class arriving according to an arrival process with pgf $A(z, z)$ which is a special case of the arrival process (Equation 6.1). This has already been analyzed, e.g., in Bruneel *et al.* (1992) for the special case of a multiserver output-queueing switch. It can be observed that the influence of HOL non-preemptive priority scheduling with priority jumps is quite large. Mean delay and delay-jitter of class-2 packets reduces considerably compared to FIFO scheduling. The price to pay is of course a bigger mean delay and delay-jitter for class-1 packets. In the case of delay insensitive traffic, this is not too big a problem. It can also be observed from these figures that the delay of high and low priority packets increases with increasing the fraction of high priority packets in the overall traffic mix.

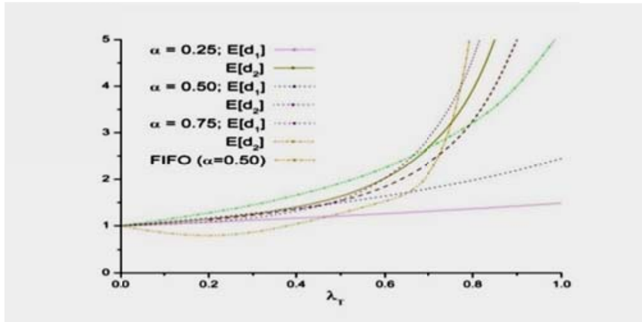

Figure 2: Mean value of packet delays versus the total arrival rate.

The mean value of the system contents of both classes are plotted versus α for $\lambda_T = 0.3$ and 0.6 which are shown in Figure 3. The mean value of the class-1 system contents increase with the fraction of class-1 packets in the traffic mix, while the opposite case is seen for the mean value of the class-2 system contents. It is clear from the figure that the difference between class-1 and class-2 system contents for different values of α is large when the load is high. The mean of class-1 system contents can be larger than the mean of the class-2 system contents for the high value of α. This is due to the fact that most of the arriving packets are of class-1 for high α and also because of jumping scheme which results in the building of the class-1 queue than the class-2 queue (although class-1 cells are served with priority). There are only class-1 packets in the most extreme case is when $\alpha = 1$, which means that the class-2 buffer stays empty.

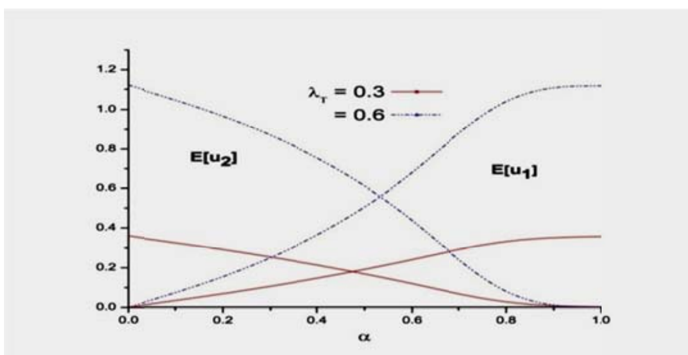

Figure 3: Mean value of system contents versus the fraction of class-1 arrivals.

The graphs for mean system contents of class-1 and class-2 packets versus the total arrival rate are plotted in Figure 4 for $\alpha = 0.25$ and $N = 2$, 4, 16 and ∞. It is clear from this figure that the output queueing switch plays a considerable role in the mean system contents. Especially the mean class-2 system contents increases considerably when N increases. Also in Figure 5 the mean class-1 and class-2 packet delays are plotted versus the total arrival rate for $\alpha = 0.25$ and $N = 2$, 4, 16 and ∞. Similar conclusions as for the mean system contents can be drawn.

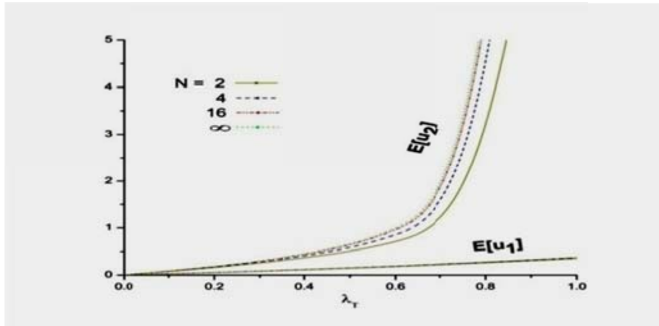

Figure 4: Influence of output queueing switch of arrival process on the mean system contents for $\alpha = 0.25$.

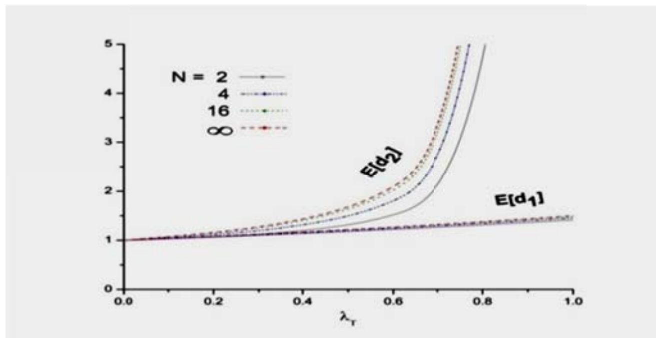

Figure 5: Influence of output queueing switch of arrival process on the mean packet delay for $\alpha = 0.25$.

7. Summary and Conclusions

In this paper, we analyzed the system contents in a queueing system with non-preemptive HOL priority scheduling. A generating-functions-approach was adopted, which led to closed-form expressions of performance measures, such as mean of system contents and packet delay of both classes, which are easy to evaluate. The model included possible correlation between the number of arrivals of the two classes during a slot and general service times for packets of both classes. Therefore, the results could be used to analyze performance of an output-queueing switch having Bernoulli arrivals and dynamically prioritized. Finally, the effect of jumping mechanism is analyzed which clearly shows that the queueing system provides better results when the fraction of class-1 arrivals in the overall traffic mix is small.

Acknowledgment

The present research work is carried out under the INSPIRE Fellowship funded by Department of Science and Technology, Government of India.

REFERENCES

Bruneel, H. (1993). Performance of discrete-time queueing systems, Computers & Operations Research*,* 20:3, 303–320.

Bruneel, H. and Kim, B. G. (1993). Discrete-time models for communication systems including ATM, Kluwer Academic Publishers, Boston.

Bruneel, H., Steyaert, B., Desmet, E. and Petit, G. H. (1992). An analytical technique for the derivation of the delay performance of ATM switches with multiserver output queues, International Journal of Digital and Analog Communication Systems, 5, 193-201.

Choi, B. I. and Lee, Y. (2001). Performance analysis of a dynamic priority queue for traffic control of bursty traffics in ATM networks, IEE Proceedings-Communications, 148(3), 181-187.

Chaudhry, M. and Gupta, U. (2001). Modeling and analysis of discrete-time multiserver queues with batch arrivals: $GI^X / Geo / m$, *INFORMS,* Journal on Computing, 13:3, 172–180.

Cidon, I. and Sidi, M. (1997). Dispersed messages in discrete-time queues: Delay, Jitter and Threshold Crossing, Performance Evaluation, 29, 85–104.

Fratini, S. (1990). Analysis of a dynamic priority queue. Commun. Statist.-Stochastic Models, 6(3), 415-444.

Haghighi, Aliakbar Montazer, Chukova, Stefanka S. and Mishev, Dimitar P. (2011). Single-server Poisson queueing system with splitting and Delayed-feedback: Part I. *International Journal of Mathematics in Operational Research (IJMOR)*, Volume 3 - Issue 1 – 2011, pp 1 – 21.

Ishizaki, F. and Takine, T. (1995). Analysis of a discrete-time queue with gated priority, Performance Evaluation, 23, 121–143.

Kleinrock, L. (1975). Queueing Systems. Wiley, New York.

Knessl, C., Choi, D. I. and Tier, C. (2002). A dynamic priority queue model for simultaneous service of voice and data traffic, SIAM Journal on Applied Mathematics, 63(2), 398-422.

Lee, Y. and Choi, B. D. (2001). Queueing system with multiple delay and loss priorities for ATM networks, Information Sciences, An International Journal, 138:1-4, 7–29.

Lim, Y. and Kobza, J. E. (1990). Analysis of a delay-dependent priority discipline in an integrated multiclass traffic fast packet switch, IEEE Transactions on Communications, 38 (5), 659-685.

Maertens, T, Walraevens, J. and Bruneel, H. (2006a). On priority queues with priority jumps, Performance Evaluation, 63(12), 1235-1252.

Maertens, T., Walraevens, J. and Bruneel, H. (2007). A modified HOL priority scheduling discipline: performance analysis, European Journal of Operational Research, 180(3), 1168-1185.

Maertens, T., Walraevens, J. and Bruneel, H. (2008). Performance comparison of several priority schemes with priority jumps, *A*nn Oper Res, 162, 109-125.

Maertens, T., Walraevens, J., Moeneclaey, M. and Bruneel, H. (2006b). A new dynamic priority scheme: performance analysis, In Proceedings of the 13th international conference on analytical and stochastic modelling techniques and applications (ASMTA2006), 74-84.

Nassar, H. and AL Mahdy, H. (2003). Queueing analysis of an ATM multimedia multiplexer with nonpreemptive priority, IEE Proceedings on Communications, 150:3, 189–196.

Rubin, I. and Tsai, Z. H. (1989). Message delay analysis of multiclass priority TDMA, FDMA, and discrete time queueing systems, IEEE Transactions on Information Theory, 35(3), 637-647.

Steyaert, B. and Xiong, Y. (1996). Analysis of a discrete-time queue with general three-state Markovian traffic sources, Performance Evaluation, 27, pp. 277–294.

Sugahara, A., Takine, T., Takahashi, Y. and Hasegawa, T. (1995). Analysis of a nonpreemptive priority queue with SPP arrivals of high class, Performance Evaluation 1(3), 215-238.

Takahashi, M., Osawa, H. and Fujisawa, T. (1999). $Geo^X/G/1$ retrial queue with nonpreemptive priority, Asia Pacific Journal of Operational Research, 16:2, 215–234.

Takine, T., Sengupta, B. and Hasegawa, T. (1994). An analysis of a discrete-time queue for broadband ISDN with priorities among traffic classes, IEEE Transactions on Communication, 42(2-4), 1837- 1845.

Walraevens, J., Steyaert, B. and Bruneel, H. (2003). Performance analysis of a single-server ATM queue with a priority scheduling, Computers and Operations Research, 30(12), 1807-1829.