

Performance Analysis of Two Priority Queuing Systems in Tandem

Faouzi Kamoun

College of Technological Innovation, Zayed University, Dubai, UAE

Email: faouzi.kamoun@zu.ac.ae

Received September 19, 2012; revised October 22, 2012; accepted November 2, 2012

ABSTRACT

In this paper, we consider a tandem of two head-of-line (HOL) non-preemptive priority queuing systems, each with a single server and a deterministic service-time. Two classes of traffic are considered, namely high priority and low priority traffic. By means of a generating function approach, we present a technique to derive closed-form expressions for the mean buffer occupancy at each node and mean delay. Finally, we illustrate our solution technique with some numerical examples, whereby we illustrate the starvation impact of the HOL priority scheduling discipline on the performance of the low-priority traffic stream. Our research highlights the important fact that the unfairness of the HOL priority scheduling becomes even more noticeable at the network level. Thus this priority mechanism should be used with caution.

Keywords: Priority Queuing System; Tandem Queues; Performance Analysis; Discrete-Time Queues

1. Introduction

In this contribution, we consider two discrete-time priority queuing systems (hereafter referred to as PQS1 and PQS2) in tandem. Each PQS implements a non-preemptive head-of-line (HOL) priority scheduling with a single server. We consider two types of priority customers, namely high-priority and low priority customers (hereafter referred to as class-1 and class-2, respectively). The number of class-1 and class-2 arrivals within a slot can be correlated random variables.

The main thrust behind our interest to explore the performance of tandem priority queues is that previous studies have focused on an isolated priority queuing system. However, it is well known that in a typical (store and forward) network environment, customers go through a number of switching nodes before reaching their final destinations. For instance, customers belonging to a virtual connection are stored and forwarded through a number of queues in tandem. Similarly, in many manufacturing and service applications (including health-care and banking), customers (or jobs) need to go through a number of prioritized service facilities before completing a single transaction. At the same time, none of the previous studies on tandem queues have considered multi-class priority scheduling schemes. It is well known however that next generation telecommunications networks are being built around the inspiration of having a single cost-effective packet-based network that is capable of sup-

porting diverse classes of services, each with its own Quality of Service (QoS) requirement. For instance, real-time traffic is delay and delay-jitter sensitive, while non-real-time applications are mainly loss sensitive. To achieve this goal, various packet service disciplines that determine the order by which customers are served have been proposed. Among the simplest time-priority scheduling schemes, the non-preemptive head-of-line (HOL) priority scheduling discipline has been proposed to provide differentiated services to the high-priority traffic class. Under this scheme, the server always schedules the delay-sensitive traffic first (if present). Many previous studies (see for example [1-4]) have analyzed the performance of discrete-time HOL-priority queuing systems (in terms of system content and/or customer delay) under various assumptions regarding the arrival and service time distributions. In particular, Walraevens *et al.* [5] analyzed a single server discrete-time HOL priority queuing system in steady-state. The corresponding transient results were reported in [6], while the case of multi-server HOL priority queue is treated in [7]. In [8], Walraevens *et al.* analyzed the busy periods and the output characteristics of a HOL priority queue. Kamoun [9] analyzed a HOL priority queuing system which is subjected to a Markovian interruption process. However, to our best knowledge, none of the previous contributions addressed the case where customers have to go through more than one priority queuing stage.

The contribution of this paper consists of the practical

queuing model being considered, the solution technique being developed to extract explicit expressions for the various performance measures, and the new insights provided on the performance of HOL priority queuing systems in a small “network” environment:

First, we add to the existing literature on non-preemptive priority models by considering two priority-based queuing systems in tandem. This tandem configuration of two PQS stages makes the resulting model more practical and flexible in studying more realistic priority-based queuing systems.

Second, we demonstrate how a generating function can be used to derive closed-form expressions for the mean buffer occupancies and mean delays. The closed-form expressions of these performance measures make it very convenient to understand the effect of various system parameters on the performance of both traffic types. In particular, this research demonstrates how the starvation of the low-priority traffic, due to the HOL priority scheduling discipline, is particularly accentuated at the second stage. Our approach also bypasses the need to characterize or approximate the output process of the first PQS stage. Last, the exact expressions for some key performance measures can also be used as baselines to validate the accuracy of potential approximation techniques.

The remainder of this paper is organized as follows: In Sections 2 and 3, we describe the queuing model and derive an expression for the Joint Probability Generating Function (JPGF) of the system state vector. In Section 4, we derive explicit expressions for some boundary terms appearing in the JPGF. From this JPGF, the marginal PGFs of the system contents at three of the four nodes, as well as several other JPGFs are derived, as illustrated in Section 5. Closed-form expressions for the first moments of the system contents and customer delay are presented in Section 6. Numerical results that provide insights into the behavior of the queuing model are provided in Section 7. In particular, we wanted to further explore the impact of the HOL priority scheduling discipline in a tandem configuration on the performance of class-1 and class-2 customers. Finally, a summary of the main findings of the papers and some suggestions for further research are provided in Section 8.

2. Analytical Model Description

We consider two discrete-time priority queuing systems (PQS) in a tandem configuration. Each PQS has an infinite storage capacity and a single (FCFS) server. The time axis is divided into equal length slots and customer service completion is synchronized to occur at the slot boundaries. Here a slot is the time period required to transmit exactly one customer from the system. It is

assumed that a customer which arrives during a slot cannot be served before the beginning of the next slot. A “conceptual” diagram for our queuing model is depicted in **Figure 1**.

As shown in **Figure 1**, in each PQS, customers are served based on a HOL time-priority scheme, whereby type-1 customers have absolute non-preemptive priority over type-2 customers. Under this priority scheme, the server will always serve type-1 customers (if any). If there are no type-1 customers, then type-2 customers (if any) will be served. The model shown in **Figure 1** can be viewed as two HOL priority queuing systems cascaded in tandem (PQS1 and PQS 2), or as two parallel tandem queuing systems, which are coupled. In the latter case, the first tandem system consist of two high-priority queues (nodes 1 and 3), while the second system consists of the two low-priority queues (nodes 2 and 4). We also note that the input to the high/low priority queues (node 3/4 in PQS 2) consists of new type 1/2 customers joining the network (exogenous arrivals), as well as type 1/2 customers which completed service at the first PQS. It is also clear from the model description that the high-priority traffic stream, traversing nodes 1 and 3, is not interrupted by the HOL priority scheduling mechanism. The low-priority traffic stream, on the other hand, is subjected to two interruptions, at nodes 2 and 4. It is also assumed that inter-stage departures are not allowed.

In the sequel, we denote the number of arrivals of class- j customers to PQS 1 during slot k by $a_{j,k}$ ($j = 1, 2$). These arrivals constitute a series of *i.i.d.* random variables with the common joint probability generating function $A(z_1, z_2) = E[z_1^{a_{1,k}} z_2^{a_{2,k}}]$, independent of k . We define the marginal pgfs of the arrivals to PQS 1 from class-1 and class-2 customers during a slot by

$$A_1(z_1) = E[z_1^{a_{1,k}}] \text{ and } A_2(z_2) = E[z_2^{a_{2,k}}]$$

respectively.

The marginal arrival rates of class- j ($j = 1, 2$) to PQS 1 are denoted by $\lambda_j = A_j'(1)$. We further denote the total number of arriving customers to PQS 1 during slot k by $a_{1T,k} = a_{1,k} + a_{2,k}$, with corresponding pgf $A_T(z)$. The total arrival rate to PQS 1 is denoted

$\lambda_{1T} = A_T'(1) = \lambda_1 + \lambda_2$. Similarly, we denote the number

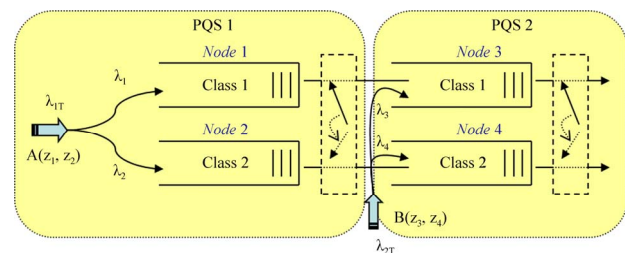


Figure 1. Conceptual diagram of the tandem queuing model with HOL priority.

of “exogenous” arrivals of class- j customers to PQS 2 during slot k by $b_{j,k} (j=1,2)$. These arrivals constitute a series of i.i.d. random variables with the common joint probability generating function $B(z_3, z_4) = E[z_3^{b_{1,k}} z_4^{b_{2,k}}]$, independent of k . We define the marginal pgfs of the exogenous arrivals to PQS 2 from class-1 and class-2 customers during a slot by $B_3(z_3) = E[z_3^{b_{1,k}}]$ and $B_4(z_4) = E[z_4^{b_{2,k}}]$ respectively.

The marginal exogenous arrival rates of class- j ($j = 1, 2$) to PQS 2 are denoted by $\lambda_j = B_j'(1) (j=3,4)$. We further denote the total number of exogenous arrivals to PQS 2 during slot k by $b_{2T,k} = b_{1,k} + b_{2,k}$, with corresponding pgf $B_T(z)$. The total arrival rate to PQS 2 is denoted by $\lambda_{2T} = B_T'(1) = \lambda_3 + \lambda_4$. The total arrival rate to the whole system is denoted by

$$\lambda_T = A_T'(1) + B_T'(1) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4.$$

In the sequel, we assume that the system is stable and that a stochastic equilibrium exists.

Our model allows for the number of class-1 and class-2 arrivals to each PQS, within a slot, to be correlated random variables. This type of correlation can occur for instance when an incoming job (customer) can be split into two parts, an urgent and a non-urgent part, where the urgent part is of class-1, while the non-urgent part is of class-2. Taking into account this correlation in the arrival process leads to more accurate performance results [10]. In the next sections, we will demonstrate that an analysis based on probability generating functions approach can be used to derive some relevant performance measures for the queuing system under consideration.

3. Mathematical Model

The queuing model, shown in **Figure 1**, can be formulated as a discrete-time multidimensional Markov chain. The state of the system is defined by $(u_{1,k}, u_{2,k}, u_{3,k}, u_{4,k})$ where $u_{i,k}$ is the system content of node i ($i = 1 - 4$) at the end of the k th slot. The evolution of the system contents is described by the following system equations:

$$\begin{aligned} u_{1,k+1} &= [u_{1,k} - 1]^+ + a_{1,k+1} \\ u_{2,k+1} &= \begin{cases} [u_{2,k} - 1]^+ + a_{2,k+1} & \text{if } u_{1,k} = 0 \\ u_{2,k} + a_{2,k+1} & \text{if } u_{1,k} > 0 \end{cases} \\ u_{3,k+1} &= [u_{3,k} - 1]^+ + b_{1,k+1} + [u_{1,k}]^+ \\ u_{4,k+1} &= \begin{cases} [u_{4,k} - 1]^+ + b_{2,k+1} + 1 \Big|_{u_{2,k} > 0 \wedge u_{1,k} = 0} & \text{if } u_{3,k} = 0 \\ u_{4,k} + b_{2,k+1} + b_{2,k+1} + 1 \Big|_{u_{2,k} > 0 \wedge u_{1,k} = 0} & \text{if } u_{3,k} > 0 \end{cases} \end{aligned} \tag{1}$$

where $[x]^+$ is a binary-valued random variable which takes the value 1 if $x > 0$ and 0 otherwise, while $1|_x$ is

the indicator function which takes the value 1 if x is true and 0 otherwise.

The first two expressions in (1) follow from the fact that high priority customers are not influenced by low-priority customers, while a low priority customer can only be served, if there are no high priority customers in PQS1 (which leads to the second equation). The last two expressions in (1) follow from similar arguments, in addition to taking into account the fact that the output of high/low priority queue in PQS1 (if any) will also constitute an input to the high/low priority queue in PQS2. We also took into account the fact that the low-priority node (2) can only forward a customer to the downstream low-priority node (3) if it is not empty and if there are no high priority customers in PQS1.

Next, define the JPGF of system contents at the end of the k th slot as follows:

$$\begin{aligned} Q_k(z_1, z_2, z_3, z_4) &= E[z_1^{u_{1,k}} z_2^{u_{2,k}} z_3^{u_{3,k}} z_4^{u_{4,k}}] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} z_1^i z_2^j z_3^k z_4^l p_k(i, j, k, l) \end{aligned} \tag{2}$$

where

$$p_k(i, j, k, l) = pr[u_{1,k} = i, u_{2,k} = j, u_{3,k} = k, u_{4,k} = l]$$

is the joint distribution of the system state vector at the end of the k th slot.

It follows that:

$$Q_{k+1}(z_1, z_2, z_3, z_4) = E[z_1^{u_{1,k+1}} z_2^{u_{2,k+1}} z_3^{u_{3,k+1}} z_4^{u_{4,k+1}}].$$

Taking the z -transform of the system Equations (1), and by considering the nine cases depicted in **Figure 2**, we derive, by means of some standard z -transform manipulations, the following relationship between

$$Q_{k+1}(z_1, z_2, z_3, z_4) \text{ and } Q_k(z_1, z_2, z_3, z_4),$$

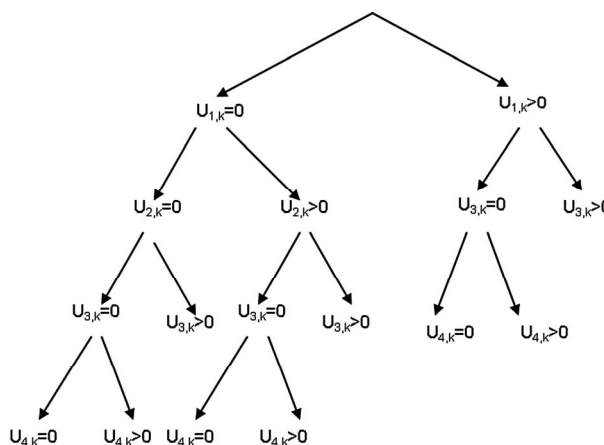


Figure 2. The nine cases used in the derivation of Equation (3).

as illustrated in Equation (3) (below).

Next, we define the steady-state JPGF of the system contents:

$$Q(z) = Q(z_1, z_2, z_3, z_4) \triangleq \lim_{k \rightarrow \infty} Q_k(z_1, z_2, z_3, z_4)$$

Applying the above limit to Equation (3) yields (see Equation (4) below):

Although it was not possible to determine all the seven boundary functions appearing in (4), as one of the 2 unknown boundary terms ($Q(0, 0, z_3, z_4)$ or $Q(0, z_2, 0, z_4)$) could not be determined, we were able to derive the relevant performance measures related to mean system contents at each node and mean delays. It should be noted that the strong coupling between node 4 and the remaining nodes made the direct application of standard pgf techniques to extract all the boundary terms, appearing in (4), an intricate task.

4. Determination of Relevant Boundary Terms

In this section, we show how to progressively determine some unknown boundary terms in order to extract relevant performance measures for the queuing model under consideration.

4.1. Determination of $Q(z_1, z_2, 0, 0)$, $Q(0, z_2, 0, 0)$ and $Q(0, 0, 0, 0)$

First, to determine $Q(z_1, z_2, 0, 0)$, let $k \rightarrow \infty, z_4 \rightarrow 0$

and then $z_3 \rightarrow 0$ in Equation (3). Then:

$$Q(z_1, z_2, 0, 0) = A(z_1, z_2)B(0, 0) \cdot K \tag{5}$$

where:

$$K = Q(0, 0, 0, 0) + \frac{\partial Q(0, 0, 0, z_4)}{\partial z_4} \Big|_{z_4=0} + \frac{\partial Q(0, 0, z_3, 0)}{\partial z_3} \Big|_{z_3=0} \tag{6}$$

Next, to determine the constant K , we set $z_1=z_2=1$ in Equation (5), which yields:

$$K = \frac{Q(1, 1, 0, 0)}{B(0, 0)} \tag{7}$$

The boundary constant $Q(1, 1, 0, 0)$ is readily obtained from the normalization condition $Q(1, 1, 1, 1) = 1$. In fact by setting $z_1 = z_2 = z_3 = z_4 = z$ in Equation (4), it is easy to derive the following expression for the PGF of the total system content:

$$Q_T(z) = \frac{A_T(z)B_T(z)(z-1)Q(z, z, 0, 0)}{z - A_T(z)B_T(z)} \tag{8}$$

The normalization condition $Q_T(1) = 1$ leads to $Q(1, 1, 0, 0) = 1 - \lambda_T$. By substituting this result back into Equation (7) and then into Equation (5), we immediately obtain the first unknown boundary function:

$$Q(z_1, z_2, 0, 0) = (1 - \lambda_T)A(z_1, z_2) \tag{9}$$

$$\begin{aligned} Q_{k+1}(z_1, z_2, z_3, z_4) &= E \left[z_1^{u_{1,k+1}} z_2^{u_{2,k+1}} z_3^{u_{3,k+1}} z_4^{u_{4,k+1}} \right] \\ &= A(z_1, z_2)B(z_3, z_4) \left\{ Q_k(0, 0, 0, 0) + \frac{1}{z_4} [Q_k(0, 0, 0, z_4) - Q_k(0, 0, 0, 0)] + \frac{1}{z_3} [Q_k(0, 0, z_3, z_4) - Q_k(0, 0, 0, z_4)] \right. \\ &\quad + \frac{z_4}{z_2} [Q_k(0, z_2, 0, 0) - Q_k(0, 0, 0, 0)] + \frac{1}{z_2} [Q_k(0, z_2, 0, z_4) - Q_k(0, 0, 0, z_4) - Q_k(0, z_2, 0, 0) + Q_k(0, 0, 0, 0)] \\ &\quad + \frac{z_4}{z_2 z_3} [Q_k(0, z_2, z_3, z_4) - Q_k(0, 0, z_3, z_4) - Q_k(0, z_2, 0, z_4) + Q_k(0, 0, 0, z_4)] \\ &\quad + \frac{1}{z_1} [Q_k(z_1, z_2, z_3, z_4) - Q_k(0, z_2, z_3, z_4) - Q_k(z_1, z_2, 0, z_4) + Q_k(0, z_2, 0, z_4)] \\ &\quad \left. + \frac{z_3}{z_1} [Q_k(z_1, z_2, 0, 0) - Q_k(0, z_2, 0, 0)] + \frac{z_3}{z_1 z_4} [Q_k(z_1, z_2, 0, z_4) - Q_k(0, z_2, 0, z_4) - Q_k(z_1, z_2, 0, 0) + Q_k(0, z_2, 0, 0)] \right\} \tag{3} \end{aligned}$$

$$\begin{aligned} Q(z) &= A(z_1, z_2)B(z_3, z_4) \cdot \frac{\{ z_4(z_1 z_4 - z_2 z_3)Q(0, z_2, z_3, z_4) + z_2 z_3(z_3 - z_4)Q(z_1, z_2, 0, z_4) \\ &\quad + (z_3 - z_4)(z_1 z_4 - z_2 z_3)Q(0, z_2, 0, z_4) + z_2 z_3^2(z_4 - 1)Q(z_1, z_2, 0, 0) \\ &\quad + z_1(z_3 - z_4)(z_2 - z_4)Q(0, 0, 0, z_4) + z_1 z_4(z_2 - z_4)Q(0, 0, z_3, z_4) \\ &\quad + z_3(z_4 - 1)(z_1 z_4 - z_2 z_3)Q(0, z_2, 0, 0) + z_1 z_3(z_4 - 1)(z_2 - z_4)Q(0, 0, 0, 0) \}}{z_2 z_3 z_4 [z_1 - A(z_1, z_2)B(z_3, z_4)]} \tag{4} \end{aligned}$$

The other two related boundary terms are readily derived from Equation (9):

$$Q(0, z_2, 0, 0) = (1 - \lambda_T) A(0, z_2) \tag{10}$$

$$Q(0, 0, 0, 0) = (1 - \lambda_T) A_T(0) \tag{11}$$

and therefore for the system to be stable, we require that $Q(0, 0, 0, 0) > 0$ or equivalently, $\lambda_T < 1$.

4.2. Relationship between $Q(z_1, z_2, 0, z_4)$ and $Q(0, z_2, 0, z_4)$

In this subsection, we determine $Q(z_1, z_2, 0, z_4)$ up to the boundary term $Q(0, z_2, 0, z_4)$. For this purpose, let $k \rightarrow \infty$ and $z_3 \rightarrow 0$ in Equation (3). Then:

$$Q(z_1, z_2, 0, z_4) = A(z_1, z_2) B(0, z_4) \Phi(z_2, z_4) \tag{12}$$

where:

$$\begin{aligned} &\Phi(z_2, z_4) \\ &= Q(0, 0, 0, 0) + \frac{1}{z_4} [Q(0, 0, 0, z_4) - Q(0, 0, 0, 0)] \\ &+ \frac{\partial Q(0, 0, z_3, z_4)}{\partial z_3} \Big|_{z_3=0} + \frac{z_4}{z_2} [Q(0, z_2, 0, 0) - Q(0, 0, 0, 0)] \\ &+ \frac{z_4}{z_2} \left[\frac{\partial Q(0, z_2, z_3, z_4)}{\partial z_3} \Big|_{z_3=0} - \frac{\partial Q(0, 0, z_3, z_4)}{\partial z_3} \Big|_{z_3=0} \right] \\ &+ \frac{1}{z_2} [Q(0, z_2, 0, z_4) - Q(0, 0, 0, z_4) - Q(0, z_2, 0, 0) \\ &+ Q(0, 0, 0, 0)] \end{aligned}$$

is independent of z_1 . Substituting $z_1 = 0$ in Equation (12), we get:

$$\Phi(z_2, z_4) = \frac{Q(0, z_2, 0, z_4)}{A(0, z_2) B(0, z_4)}$$

and therefore:

$$Q(z_1, z_2, 0, z_4) = \frac{A(z_1, z_2)}{A(0, z_2)} Q(0, z_2, 0, z_4) \tag{13}$$

Thus $Q(z_1, z_2, 0, z_4)$ is determined up to the boundary term $Q(0, z_2, 0, z_4)$.

4.3. Determination of $Q(0, 0, z, z)$ and $Q(0, z, 0, z)$

To determine the boundary term $Q(0, 0, z, z)$, we proceed as follows:

First by setting $z_1 = z_2 = x$ and $z_3 = z_4 = z$ in Equation (4), we obtain (see Equation (14) below):

From (9), we note that $Q(x, x, 0, 0) = (1 - \lambda_T) A_T(x)$. Also since $Q(x, x, z, z)$ is analytic inside the polydisk ($(|x| < 1, |z| < 1)$), then the numerator must be zero whenever the denominator is zero. It follows that:

$$Q(0, 0, z, z) = (1 - \lambda_T)(z - 1) \left[\frac{z A_T(Y_z) + (Y_z - z) A_T(0)}{z - Y_z} \right] \tag{15}$$

where Y_z is the unique solution for x in the unit disk ($|z| < 1$) of the equation $x - A_T(x) B_T(z) = 0$.

Next, we determine $Q(0, z, 0, z)$ by proceedings as follows:

First by setting $z_1 = z_3 = x$ and $z_2 = z_4 = z$ in Equation (4), we obtain (see Equation (16) below):

From (9), we note that $Q(x, x, 0, 0) = (1 - \lambda_T) A_T(x)$. Also from (13):

$$Q(x, z, 0, z) = \frac{A(x, z)}{A(0, z)} Q(0, z, 0, z) \tag{17}$$

Substituting for $Q(x, x, 0, 0)$ and $Q(x, z, 0, z)$ in (16) and since $Q(x, z, x, z)$ is analytic inside the polydisk ($(|x| < 1, |z| < 1)$), then from Rouché's theorem, we get:

$$Q(0, z, 0, z) = (1 - \lambda_T)(z - 1) A(0, z) \frac{R_z}{z - R_z} \tag{18}$$

where R_z is the unique solution for x of the equation $x = A(x, z) B(x, z)$ in the unit circle $|x| < 1$.

5. Joint and Marginal PGFs

Having determined some relevant boundary terms, we can now easily derive the marginal PGFs of the system contents at nodes 1, 2, and 3, as well as several other JPGFs. This is illustrated below.

5.1. Joint and Marginal PGFs of High-Priority (Class-1) System Contents

Let $P_{13}(z_1, z_3) = Q(z_1, 1, z_3, 1)$ denote the JPGF of the

$$Q(x, x, z, z) = A_T(x) B_T(z) \cdot \left[\frac{(x - z) Q(0, 0, z, z) + z(z - 1) Q(x, x, 0, 0) + (z - 1)(x - z) Q(0, 0, 0, 0)}{z[x - A_T(x) B_T(z)]} \right] \tag{14}$$

$$Q(x, z, x, z) = A(x, z) B(x, z) \cdot \left[\frac{(x - z) Q(x, z, 0, z) + x(z - 1) Q(x, z, 0, 0)}{z[x - A(x, z) B(x, z)]} \right] \tag{16}$$

system contents of the high-priority queues at nodes 1 and 3. By setting $z_2 = z_4 = 1$ in Equation (4), we get (see Equation (19) below):

To determine the boundary terms $P_{13}(z_1, 0)$ and $P_{13}(0, 0)$, appearing in (19), we first set $z_2 = z_4 = 1$ in (13):

$$P_{13}(z_1, 0) = \frac{A_1(z_1)}{A_1(0)} P_{13}(0, 0) \tag{20}$$

Next, from (19), the normalization condition $P_{13}(1, 1) = 1$ yields $P_{13}(1, 0) = 1 - \lambda_1 - \lambda_3$. Therefore setting $z_1 = 1$ in (20), we get:

$$\begin{aligned} P_{13}(0, 0) &= (1 - \lambda_1 - \lambda_3) A_1(0) \\ P_{13}(z_1, 0) &= (1 - \lambda_1 - \lambda_3) A_1(z_1) \end{aligned} \tag{21}$$

The remaining boundary term $P_{13}(0, z_3)$ is readily obtained by substituting (20) and (21) into (19) and by invoking the analytical property of $P_{13}(0, z_3)$ inside the polydisk ($|z_1| < 1; |z_3| < 1$). It follows from Rouché's theorem that:

$$\begin{aligned} P_{13}(0, z_3) &= (1 - \lambda_1 - \lambda_3)(z_3 - 1) \cdot \frac{z_3 A_1(X_{z_3}) + (X_{z_3} - z_3) A_1(0)}{z_3 - X_{z_3}} \tag{22} \end{aligned}$$

where for a given z_3 ($|z_3| < 1$), X_{z_3} is the unique solution for z_1 of the equation $z_1 = A_1(z_1) B_3(z_3)$ in the unit circle $|z_3| < 1$.

Substituting (21) and (22) back into (19), we obtain after few algebraic manipulations:

$$\begin{aligned} P_{13}(z_1, z_3) &= (1 - \lambda_1 - \lambda_3) A_1(z_1) B_3(z_3) \frac{(z_3 - 1)}{z_3 - X_{z_3}} \\ &\cdot \frac{[(z_3 - X_{z_3}) A_1(z_1) + (z_1 - z_3) A_1(X_{z_3})]}{z_1 - A_1(z_1) B_3(z_3)} \end{aligned} \tag{23}$$

From the above, we can obtain expressions for the following:

• **Marginal PGF of the number of class-1 customers in queue 1:**

$$P_1(z_1) = P_{13}(z_1, 1) = (1 - \lambda_1) \frac{(z_1 - 1) A_1(z_1)}{z_1 - A_1(z_1)} \tag{24}$$

• **Marginal PGF of the number of class-1 customers in queue 3:**

$$\begin{aligned} P_3(z_3) &= P_{13}(1, z_3) \\ &= (1 - \lambda_1 - \lambda_3) B_3(z_3) \\ &\cdot \frac{(z_3 - 1) [z_3 - X_{z_3} + (1 - z_3) A_1(X_{z_3})]}{z_3 - X_{z_3} \quad 1 - B_3(z_3)} \end{aligned} \tag{25}$$

• **PGF of the total number of class-1 customers in the system:**

$$P_{13,T}(z) = P_{13}(z, z) = \frac{(1 - \lambda_1 - \lambda_3)(z - 1) A_1(z)^2 B_3(z)}{z - A_1(z) B_3(z)} \tag{26}$$

It should be noted that since the high-priority traffic (traversing nodes 1 and 3) is not influenced by the low-priority traffic (traversing nodes 2 and 4), then the above results for the JPGF $P_{1,3}(z_1, z_3)$ could have been derived by analyzing the performance of the two tandem queues 1 and 3, independently from the rest of the system. We carried this analysis and confirmed that the resulting JPGF of the system contents matches the one derived in Equation (23).

5.2. Joint PGFs of the System Contents in PQS1

Let $P_{12}(z_1, z_2) = Q(z_1, z_2, 1, 1)$ denote the JPGF of the system contents of the first priority queuing system(PQS 1). Then from Equation (4):

$$\begin{aligned} P_{12}(z_1, z_2) &= A(z_1, z_2) \cdot \frac{(z_1 - z_2) P_{12}(0, z_2) + z_1 (z_2 - 1) P_{12}(0, 0)}{z_2 [z_1 - A(z_1, z_2)]} \end{aligned} \tag{27}$$

The boundary constant $P_{12}(0, 0)$ is readily obtained from the normalization condition $P_{12}(1, 1) = 1$, yielding $P_{12}(0, 0) = 1 - \lambda_1 - \lambda_2$. The boundary term $P_{12}(0, z_2)$ is readily obtained from (27) by applying Rouché's theorem:

$$P_{12}(0, z_2) = (1 - \lambda_1 - \lambda_2) W_{z_2} \frac{z_2 - 1}{z_2 - W_{z_2}} \tag{28}$$

where for a given z_2 ($|z_2| < 1$), W_{z_2} is the unique solution for z_1 of the equation $z_1 = A(z_1, z_2)$ in the unit circle $|z_2| < 1$.

Next, by substituting for $P_{12}(0, z_2)$ and $P_{12}(0, 0)$ back into (27), we get:

$$\begin{aligned} P_{12}(z_1, z_2) &= (1 - \lambda_1 - \lambda_2) A(z_1, z_2) \cdot \frac{z_2 - 1}{z_2 - W_{z_2}} \frac{z_1 - W_{z_2}}{z_1 - A(z_1, z_2)} \end{aligned} \tag{29}$$

$$P_{13}(z_1, z_3) = A_1(z_1) B_3(z_3) \cdot \left[\frac{z_3(z_3 - 1) P_{13}(z_1, 0) + (z_1 - z_3) P_{13}(0, z_3) + (z_3 - 1)(z_1 - z_3) P_{13}(0, 0)}{z_3 [z_1 - A_1(z_1) B_3(z_3)]} \right] \tag{19}$$

From the above, we can obtain expressions for the following:

• **Marginal PGF of the number of class-1 customers in queue 1:**

$$P_1(z_1) = P_{12}(z_1, 1) = (1 - \lambda_1) \frac{(z_1 - 1)A_1(z_1)}{z_1 - A_1(z_1)} \quad (30)$$

• **Marginal PGF of the number of class-2 customers in queue 2:**

$$P_2(z_2) = P_{1,2}(1, z_2) = (1 - \lambda_1 - \lambda_2) A_2(z_2) \frac{z_2 - 1}{z_2 - W_{z_2}} \frac{1 - W_{z_2}}{1 - A_2(z_2)} \quad (31)$$

• **PGF of the total number of class-1 customers in PQS1**

$$P_{12,T}(z) = P_{12}(z, z) = \frac{(1 - \lambda_1 - \lambda_2)(z - 1)A_T(z)}{z - A_T(z)} \quad (32)$$

We also observe that the JPGF of the system contents of PQS 1 could have also been derived by considering PQS 1 as an isolated priority queuing system. Such an analysis of a single stage PQS has been presented in [5] and our expressions (31-32) match those derived therein.

5.3. PGFs of Total Number of Customers in PQS2

Let $P_{34,T}(z) = P_{34}(z, z) = Q(1, 1, z, z)$ denote the JPGF of the total number if customers in the second priority queuing system (PQS 2). Then from (4):

$$P_{34,T}(z) = (z - 1)B_T(z) \frac{zQ(1, 1, 0, 0) - Q(0, 0, z, z) - (z - 1)Q(0, 0, 0, 0)}{z[1 - B_T(z)]} \quad (33)$$

Using (9), (11) and (15), we can determine the unknown boundary terms appearing in the above equation, yielding:

$$P_{34,T}(z) = \frac{(1 - \lambda_T)(z - 1)B_T(z)[z - Y_z - (z - 1)A_T(Y_z)]}{(z - Y_z)(1 - B_T(z))} \quad (34)$$

where, as defined earlier, for a given $z (|z| < 1)$, Y_z is the unique solution for x of the equation $x - A_T(x)B_T(z) = 0$ in the unit circle $|x| < 1$.

5.4. PGF of the Total Number of Class-2 Customers in the System

Let $P_{24,T}(z) = P_{24}(z, z) = Q(1, z, 1, z)$ denote the PGF of the total number of class-2 customers in the system. Then from Equation (4):

$$P_{24,T}(z) = (z - 1)A_2(z)B_4(z) \frac{(1 - \lambda_T)A_2(z) - Q(1, z, 0, z)}{z[1 - A_2(z)B_4(z)]} \quad (35)$$

The unknown boundary term $Q(1, z, 0, z)$ can be determined from (13) and (18), which yields:

$$P_{24,T}(z) = \frac{(1 - \lambda_T)(z - 1)A_2(z)^2 B_4(z)}{1 - A_2(z)B_4(z)} \cdot \frac{1 - R_z}{z - R_z} \quad (36)$$

where R_z is the unique solution for x of the equation $x = A(x)B(x, z)$ inside the unit circle $|x| < 1$.

5.5. PGF of the Total System Contents

Let $P_T(z) = Q(z, z, z, z)$ denote the PGF of the total system contents. Then using (4) and (9) we get:

$$P_T(z) = (1 - \lambda_T)(z - 1) \frac{A_T(z)^2 B_T(z)}{z - A_T(z)B_T(z)} \quad (37)$$

6. Calculation of the Moments

To compute the various moments of system contents, we introduce and recall the following notations:

$$\lambda_{ij} = \begin{cases} \left. \frac{\partial^2 A(z_1, z_2)}{\partial z_1 \partial z_j} \right|_{z_1=z_2=1} & (i, j) \in \{1, 2\} \times \{1, 2\} \\ \left. \frac{\partial^2 B(z_3, z_4)}{\partial z_3 \partial z_j} \right|_{z_3=z_4=1} & (i, j) \in \{3, 4\} \times \{3, 4\} \end{cases}$$

$$\lambda_{1T} = A'_T(1) = \left. \frac{dA_T(z)}{dz} \right|_{z=1} = \lambda_1 + \lambda_2$$

$$\lambda_{2T} = B'_T(1) = \left. \frac{dB_T(z)}{dz} \right|_{z=1} = \lambda_3 + \lambda_4$$

$$A''_T(1) = \left. \frac{d^2 A_T(z)}{dz^2} \right|_{z=1} = \lambda_{11} + \lambda_{22} + 2\lambda_{12} \quad (38)$$

$$B''_T(1) = \left. \frac{d^2 B_T(z)}{dz^2} \right|_{z=1} = \lambda_{33} + \lambda_{44} + 2\lambda_{34}$$

Next, we observe that the functions $X_{z_3}, W_{z_2}, Y_z, R_z$, appearing in the various pgfs of the system contents, can only be explicitly defined under simple arrival processes. Fortunately, the derivation of the first and higher moments of the system contents involves evaluating these functions and their first and higher order derivatives at $z_i = 1$. These can be readily computed in closed-form.

From the corresponding pgfs (24 - 26, 30 - 32, and 34, 36 - 37), the mean system content of class- j customers ($j = 1, 2$) and the mean total system contents are readily obtained:

• **Mean system content of class-1 customers at node 1:**

$$\bar{N}_1 = P_1(1) = \lambda_1 + \frac{\lambda_{11}}{2(1-\lambda_1)} \tag{39}$$

• **Mean system content of class-1 customers at node 3:**

$$\bar{N}_3 = \lambda_1 - \frac{\lambda_{11}}{2(1-\lambda_1)} + \frac{\lambda_{11} + \lambda_{33} + 2\lambda_3(1-\lambda_3)}{2(1-\lambda_1-\lambda_3)} \tag{40}$$

• **Total mean system contents of class-1 customers:**

$$\bar{N}_{13,T} = P_{13,T}'(1) = 2\lambda_1 + \lambda_3 + \frac{\lambda_{11} + 2\lambda_1\lambda_3 + \lambda_{33}}{2(1-\lambda_1-\lambda_3)} \tag{41}$$

• **Mean system content of class-2 customers at node 2:**

$$\bar{N}_2 = P_2'(1) = \lambda_2 + \frac{2\lambda_{12} + \lambda_{22}}{2(1-\lambda_{1T})} + \frac{\lambda_2\lambda_{11}}{2(1-\lambda_{1T})(1-\lambda_1)} \tag{42}$$

• **Total mean system contents of class-2 customers:**

$$\begin{aligned} \bar{N}_{24,T} \\ = P_{24,T}'(1) = 2\lambda_2 + \frac{A_T''(1) + B_T''(1) + 2\lambda_{2T}(1-\lambda_{2T})}{2(1-\lambda_T)} \\ - \frac{2\lambda_3(1-\lambda_3) + \lambda_{11} + \lambda_{33}}{2(1-\lambda_1-\lambda_3)} \end{aligned} \tag{43}$$

• **Mean system content of class-2 customers at node 4:**

The mean number of class-2 customers at node 4 can be derived by subtracting the mean system content of class-2 customers at node 2 from the total mean system contents of class-2 customer. Hence, using (42) and (43), we get:

$$\begin{aligned} \bar{N}_4 = \bar{N}_{24,T} - \bar{N}_2 \\ = \lambda_2 + \frac{A_T''(1) + B_T''(1) + 2\lambda_{2T}(1-\lambda_{2T}) - 2\lambda_{12} - \lambda_{22}}{2(1-\lambda_{1T})} \\ - \frac{2\lambda_3(1-\lambda_3) + \lambda_{11} + \lambda_{33}}{2(1-\lambda_1-\lambda_3)} - \frac{\lambda_1\lambda_2}{2(1-\lambda_{1T})(1-\lambda_1)} \end{aligned} \tag{44}$$

• **Total mean system contents at PQS1:**

$$\bar{N}_{12,T} = P_{12,T}'(1) = \lambda_{1T} + \frac{A_T''(1)}{2[1-\lambda_{1T}]} \tag{45}$$

• **Total mean system content at PQS2:**

$$\begin{aligned} \bar{N}_{34,T} = P_{34,T}'(1) \\ = \lambda_T + \frac{A_T''(1) + B_T''(1) + 2\lambda_{2T}\lambda_{1T}}{2(1-\lambda_T)} - \frac{A_T''(1)}{2(1-\lambda_{1T})} \end{aligned} \tag{46}$$

• **Total mean system content**

$$\bar{N}_T = P_T'(1) = \lambda_T + \lambda_{1T} + \frac{A_T''(1) + B_T''(1) + 2\lambda_{1T}\lambda_{2T}}{2(1-\lambda_T)} \tag{47}$$

It can be easily verified that Equations (39-47) satisfy the expected results:

$$\begin{aligned} \bar{N}_{12,T} &= \bar{N}_1 + \bar{N}_2 \\ \bar{N}_{34,T} &= \bar{N}_3 + \bar{N}_4 \\ \bar{N}_{13,T} &= \bar{N}_1 + \bar{N}_3 \\ \bar{N}_T &= \bar{N}_1 + \bar{N}_2 + \bar{N}_3 + \bar{N}_4 \end{aligned} \tag{48}$$

By using Little's law, the following mean customer delays are readily obtained:

• **Mean delay at each node:**

$$\bar{d}_i = \frac{\bar{N}_i}{\lambda_i} \tag{49}$$

• **Mean delay \bar{d}_{cj} of an arbitrary class- j customer ($j = 1, 2$):**

$$\begin{aligned} \bar{d}_{c1} &= \frac{\lambda_1}{\lambda_1 + \lambda_3} \bar{d}_1 + \frac{\lambda_3}{\lambda_1 + \lambda_3} \bar{d}_3 = \frac{\bar{N}_{13,T}}{\lambda_1 + \lambda_3} \\ \bar{d}_{c2} &= \frac{\lambda_2}{\lambda_2 + \lambda_3} \bar{d}_2 + \frac{\lambda_4}{\lambda_2 + \lambda_4} \bar{d}_4 = \frac{\bar{N}_{24,T}}{\lambda_2 + \lambda_4} \end{aligned} \tag{50}$$

• **Total average delay \bar{d}_T in the system:**

$$d_T = \frac{\sum_{i=1}^4 \bar{N}_i}{\lambda_T} = \frac{\bar{N}_T}{\lambda_T} \tag{51}$$

7. Numerical Examples and Discussions

In this section, we illustrate our approach through some numerical examples. In particular, we will further probe into the influence of HOL priority on the performance of low-priority traffic stream. For this purpose, we consider two-dimensional Binomial exogenous arrival processes with JPGFs:

$$\begin{aligned} A(z_1, z_2) &= \left(1 - \frac{\lambda_1}{N}(1-z_1) - \frac{\lambda_2}{N}(1-z_2) \right)^N \\ B(z_3, z_4) &= \left(1 - \frac{\lambda_3}{N}(1-z_3) - \frac{\lambda_4}{N}(1-z_4) \right)^N \end{aligned}$$

Further, we define $v_i = \lambda_i/\lambda_T$ ($i = 1, 2, 3$) as the fraction of arrivals to node i among the total arrivals. For the numerical calculations, we have set $N = 16$ and $(v_1, v_2, v_3) = (0.45, 0.25, 0.20)$.

In **Figure 3**, the mean customer delay at each node as well as the total average delay, are plotted as a function of the total arrival rate λ_T . The mean delay \bar{d}_i at each node and the total average delay \bar{d}_T were computed by substituting for \bar{N}_i as given in Equations (39), (40), (42), and (44) into (49) and (51), respectively.

Figure 4 shows the mean delay of an arbitrary class- j

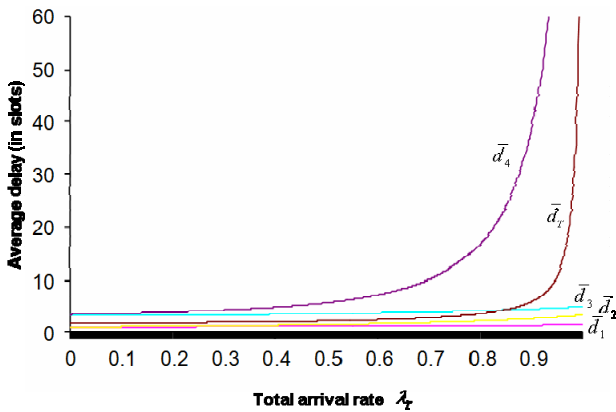


Figure 3. Mean delays versus total arrival rate λ_T .

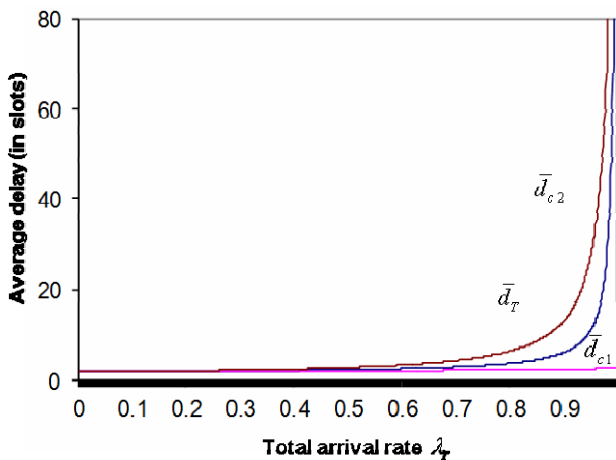


Figure 4. Mean delay of a class- j customer ($j = 1, 2$).

customer ($j = 1, 2$) and the total average delay as a function of the total arrival rate λ_T . In this case, the mean delay λ_T for each class j customer was computed by substituting for $\bar{N}_{13,T}$ and $\bar{N}_{24,T}$, as given in Equations (41) and (43), into (50).

As may be seen from the above figures, the HOL priority scheduling has a significant influence on the mean delay of the low-priority (class-2) customers. This starvation effect is even more significant at node 4. To this regards, recall that class-2 customers arriving at the access node (2) are subjected to two consecutive HOL interruption processes due to the potential presence of higher priority customers at nodes 1 and 3. The above results not only echo but also amplify previous concerns related to the unfairness of the static HOL priority scheduling mechanism for low-priority traffic. Therefore HOL scheduling should be used with caution.

Our research also suggests that more research is needed to come-up with efficient priority mechanisms that exhibit better fairness and that can be easily implemented in real-time. Earlier contributions in this regards focused on a class of dynamic priority schemes that are

based on the queue-length-threshold scheduling discipline. Under this scheme, the priority scheduling can dynamically change, based on the queue lengths (see for example [11-15]) [16]. Some More recent contributions include the work of Maertens *et al.* [17-19] on priority jumping mechanisms and the contribution of De Vuyst *et al.* [20,21] on reservation-based priority disciplines.

8. Conclusions and Future Research

In this paper, we presented an exact analysis of a two HOL priority queuing systems in a tandem configuration. For each PQS, our model allows for possible correlation between the numbers of arrivals of the two classes of customers during a slot. We showed how a generating function approach can be used to derive closed-form expressions for several performance measures. Finally we have demonstrated, via numerical examples, the negative impact the HOL priority scheduling on performance of the low-priority traffic.

This work can be further explored in many directions. For instance, it would be interesting to derive expressions for the pgf of customer delay and characterize the end-to-end delay of a tagged customer. The characterization of the asymptotic behavior of the tail distributions of the various system contents is still an open research issue.

Finally, our analysis can also be extended to handle the case where customers leaving the first PQS stage are allowed to leave the system with a pre-assigned probability. These are left for future research.

REFERENCES

- [1] A. Khamisy and M. Sidi, "Discrete-Time Priority Queues with Two-State Markov Modulated Arrivals," *Stochastic Models*, Vol. 8, No. 2, 1992, pp. 337-357. [doi:10.1080/15326349208807228](https://doi.org/10.1080/15326349208807228)
- [2] M. Sidi and A. Segall, "Structured Priority Queuing Systems with Applications to Packet-Radio Networks," *Performance Evaluation*, Vol. 3, No. 4, 1983, pp. 265-275. [doi:10.1016/0166-5316\(83\)90036-6](https://doi.org/10.1016/0166-5316(83)90036-6)
- [3] T. Takine, B. Sengupta and T. Hasegawa, "An Analysis of a Discrete-Time Queue for Broadband ISDN with Priorities among Traffic Classes," *IEEE Transactions on Communications*, Vol. 42, No. 2-4, 1994, pp. 1837-1845. [doi:10.1109/TCOMM.1994.582893](https://doi.org/10.1109/TCOMM.1994.582893)
- [4] T. Takine, "A Nonpreemptive Priority MAP/G/1 Queue with Two Classes of Customers," *Journal of Operations Research Society of Japan*, Vol. 39, No. 2, 1996, pp. 266-290.
- [5] J. Walraevens, B. Steyaert and H. Bruneel, "Performance Analysis of a Single-Server ATM Queue with a Priority Scheduling," *Computers and Operations Research*, Vol. 30, No. 12, 2003, pp. 1807-1829. [doi:10.1016/S0305-0548\(02\)00108-9](https://doi.org/10.1016/S0305-0548(02)00108-9)
- [6] J. Walraevens, D. Fiems and H. Bruneel, "Transient

- Analysis of a Discrete-Time Priority Queue,” *Proceedings of the 12th International Conference on Analytical and Stochastic Modelling Techniques and Applications (ASMTA 2005)*, Riga, June 2005, pp. 17-24.
- [7] K. Laevens and H. Bruneel, “Discrete-Time Multiserver Queues with Priorities,” *Performance Evaluation*, Vol. 33, No. 4, 1998, pp. 249-275.
[doi:10.1016/S0166-5316\(98\)00024-8](https://doi.org/10.1016/S0166-5316(98)00024-8)
- [8] J. Walraevens, D. Fiems, S. Wittevrongel and H. Bruneel, “Calculation of Output Characteristics of a Priority Queue through a Busy Period Analysis,” *European Journal of Operational Research*, Vol. 198, No. 3, 2009, pp. 891-898.
[doi:10.1016/j.ejor.2008.11.018](https://doi.org/10.1016/j.ejor.2008.11.018)
- [9] F. Kamoun, “Performance Analysis of a Non-Preemptive Priority Queuing System Subjected to a Correlated Markovian Interruption Process,” *Computers and Operations Research*, Vol. 35, No. 12, 2008, pp. 3969-3988.
[doi:10.1016/j.cor.2007.06.001](https://doi.org/10.1016/j.cor.2007.06.001)
- [10] J. Walraevens, D. Fiems and H. Bruneel, “The Discrete-Time Preemptive Repeat Identical Priority Queue,” *Queueing Systems*, Vol. 53, No. 4, 2006, pp. 231-243.
[doi:10.1007/s11134-006-7770-x](https://doi.org/10.1007/s11134-006-7770-x)
- [11] R. Chipalkatti, J. F. Kurose and D. Towsley, “Scheduling Policies for Real-Time and Nonreal-Time Traffic Packet Switching Node,” *Proceedings of the IEEE INFOCOM '89*, Ottawa, April 1998, pp. 774-783.
- [12] B. I. Choi and Y. Lee, “Performance Analysis of a Dynamic Priority Queue for Traffic Control of Bursty Traffic in ATM Networks,” *IEEE Proceedings of Communications*, Vol. 148, No. 3, 2001, pp. 181-187.
[doi:10.1049/ip-com:20010115](https://doi.org/10.1049/ip-com:20010115)
- [13] D. I. Choi, B. D. Choi and D. Sung, “Performance Analysis of Priority Leaky Bucket Scheme with Queue-Length-Threshold Scheduling Policy,” *IEEE Proceedings of Communications*, Vol. 145, No. 6, 1998, pp. 395-401.
[doi:10.1049/ip-com:19982287](https://doi.org/10.1049/ip-com:19982287)
- [14] C. Knessl, D. I. Choi and C. Trier, “A Dynamic Priority Queue Model for Simultaneous Service of Voice and Data Traffic,” *SIAM Journal on Applied Mathematics*, Vol. 63, No. 2, 2002, pp. 398-422.
[doi:10.1137/S0036139901390842](https://doi.org/10.1137/S0036139901390842)
- [15] J. T. Lee and Y. H. Kim, “Performance Analysis of a Hybrid Priority Control Scheme for Input and Output Queuing ATM Switches,” *Proceedings of INFOCOM '98*, San Francisco, 29 March-2 April 1998, pp.1470-1477.
[doi:10.1109/INFCOM.1998.662965](https://doi.org/10.1109/INFCOM.1998.662965)
- [16] T. Maertens, J. Walraevens and H. Bruneel, “Performance Analysis of a Single-Server Queue with HOL-PJ Priority Scheduling Discipline,” *Proceedings of the Second International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs '04)*, Ilkley, 26-28 July 2004, pp. 1-10.
- [17] T. Maertens, J. Walraevens and H. Bruneel, “On Priority Queues with Priority Jumps,” *Performance Evaluation*, Vol. 63, No. 12, 2006, pp. 1235-1252.
[doi:10.1016/j.peva.2005.12.003](https://doi.org/10.1016/j.peva.2005.12.003)
- [18] T. Maertens, J. Walraevens and H. Bruneel, “A Modified HOL Priority Scheduling Discipline: Performance Analysis,” *European Journal of Operational Research*, Vol. 180, No. 3, 2007, pp. 1168-1185.
[doi:10.1016/j.ejor.2006.05.004](https://doi.org/10.1016/j.ejor.2006.05.004)
- [19] T. Maertens, J. Walraevens, M. Moeneclaey and H. Bruneel, “Performance Analysis of a Discrete-Time Queuing System with Priority Jumps,” *International Journal of Electronics and Communications*, Vol. 63, No. 10, 2009, pp. 853-858. [doi:10.1016/j.aeue.2008.07.004](https://doi.org/10.1016/j.aeue.2008.07.004)
- [20] S. De Vuyst, S. Wittevrongel, D. Fiems and H. Bruneel, “Controlling the Delay Trade-off Between Packet Flows Using Multiple Reserved Places,” *Performance Evaluation*, Vol. 65, No. 6-7, 2008, pp. 484-511.
[doi:10.1016/j.peva.2007.12.008](https://doi.org/10.1016/j.peva.2007.12.008)
- [21] S. De Vuyst, D. Wittevrongel and H. Bruneel, “Place Reservation: Delay Analysis of a Novel Scheduling Mechanism,” *Computers and Operations Research*, Vol. 35, No. 8, 2008, pp. 2447-2462.
[doi:10.1016/j.cor.2006.12.003](https://doi.org/10.1016/j.cor.2006.12.003)