WAITING TIMES OF THE MULTIPLE PRIORITY DUAL QUEUE WITH FINITE BUFFER AND NON-PREEMPTIVE PRIORITY SCHEDULING

P. ZEEPHONGSEKUL and A. BEDFORD

School of Mathematical and Geospatial Sciences
RMIT University
GPO Box 2476V
Melbourne, Victoria 3001
Australia
e-mail: panlopz@rmit.edu.au

Abstract

This paper considers a novel queueing model called the Multi-Priority Dual Queue (MPDQ), which was introduced and analyzed in Bedford and Zeephongsekul [1, 2]. The MPDQ consists of two queues with a finite buffer, the primary and the secondary queue with the additional feature of a priority scheme to assist with congestion control and meeting Quality of Service (QoS) objectives for differentiated classes of customers. In this paper, it is assumed that the service discipline is non-preemptive whereas the model in the cited paper assumes a preemptive service discipline. Using some key ideas in Zeephongsekul and Bedford [13], we give a detailed analysis of the waiting time for each class of customers with numerical examples to illustrate the results.

1. Introduction

An essential prerequisite in the design of any new communication system is that it should attain a low loss probability, i.e., the probability
that a packet is dropped due to a full buffer, and it should minimize the expected waiting time of clients. The shorter the waiting time, the more demands can be met, and the lower the loss, the higher is the throughput. The Non-Preemptive Multi-Priority Dual Queue (MPDQ) discussed in this paper is such a system, aimed to meet high demands and enhance Quality of Service (QoS) to high class customers without compromising the waiting times of low class ones.

Various dynamic scheduling algorithms have been introduced with the aim of improving QoS to customers. Some new algorithms, such as Dual Queue (DQ) scheme proposed in Hayes et al. [5], are designed to give better QoS to most customers at the expense of a few, rather than give poor QoS uniformly but fairly to all customers. In the cited paper, the authors analyzed and compared the delay characteristics of the DQ against First In First Out (FIFO) and a modified Deficit Round Robin (DRR) scheduling technique (Shreedhar and Varghese [10]) and demonstrated distinct advantages using the DQ scheme. This work was later extended in Ranasinghe et al. [9] to a wireless local area network where minor modifications were made to the DQ and it was shown to perform better than standard Round Robin scheduling.

The Multiple Priority Dual Queue (MPDQ), introduced in Bedford and Zeephongsekul [1], was derived from the DQ scheme and incorporated the idea of differentiated classes of customer proposed in Odlyzko [8] in order to further reduce traffic congestion. The MPDQ introduces different classes into the DQ with the aim of providing better service to high class customers without completely disadvantaging low class customers. This is possible not only because of the prioritization of customers, but also due to the MPDQ’s queue structure which partitions a single \( \frac{M_1 + M_2}{M} / \frac{1}{c} \) queue into two queues with a finite waiting room so some lower class customers will have an opportunity to move through the combined system even in the presence of some high class customers. The MPDQ also differs from the DQ which utilized large finite buffers whereas our buffer size can be fairly small. Also, loss in the MPDQ is determined by the available buffer space and not by a time-discard policy as was used in Hayes et al. [5] where packets are ejected from the system after a period of time has elapsed.
Introducing different classes into the DQ adds additional complexity but it achieves the objective of producing a class-based QoS. Class-based system is prevalent in modern communication systems such as the internet, where traffic can be categorized into several broad areas including real time (voice, video, data), interactive (web) and best effort traffic (ftp, http) (Dimopoulos et al. [4]). High priority traffic, such as real time and interactive traffic, often have precedence of service over low priority traffic. The objective is to reduce expected waiting times as well as improve other performance characteristics.

In the model proposed, two classes of customers arrive to the dual queueing system as two independent Poisson processes. The class of customers labeled 1 (the high class) has a mean arrival rate \( \lambda_1 \) and the other class, labeled 2 (the low class), has mean arrival rate \( \lambda_2 \). The dual queue is partitioned into the primary queue and secondary queue. There is a single server at the primary queue dispensing an exponentially distributed service time with rate \( \mu_i \) to class \( i \) customers, \( i = 1, 2 \). Both queues are assumed to have finite buffers. The primary queue employs a non-preemptive priority service discipline where class 1 customers have priority over class 2 customers. An arriving customer who finds the primary queue full waits in the secondary queue (provided it is not full) which has no service facility and only serves as a holding queue for the overloaded primary queue. If both queues are full, an arriving customer is lost to the system. In the secondary queue, customers also arrange themselves in order of priority, with class 1 customers ahead of class 2 customers. Once a space is vacant in the primary queue, the head of the line in the secondary queue joins the primary queue. A schematic diagram of the MPDQ is displayed in Figure 1 below. There, \( N_q \) is the number in the primary queue (including one in service), and \( c_i \) is the buffer size of queue \( i, i = 1, 2 \).

The MPDQ is not equivalent to a single queue with a non-preemptive priority service discipline. The partitioning of the single queue into a dual queue provides lower class customers with the opportunity to move through the combined system even in the presence of some high class customers. For example, if the primary queue is full of class 2 customers
waiting to be served, then any arriving class 1 customer must wait in the secondary queue until a space becomes available in the primary queue. Due to non-preemption, this will allow the head-of-the-line class 2 customer to obtain service. In a single queue, this customer will not get serve until all class 1 customers in the combined system are served and there is no other class 1 customers in the system.

Figure 1. The Multiple Priority Dual Queue.

In previous works (Bedford and Zeephongsekul [2], Zeephongsekul and Bedford [13]), we analyzed the MPDQ with preemptive service discipline and found it reduces the waiting times for both classes of customers and was superior to a single preemptive finite buffer model. In an extensive simulation study (Bedford and Zeephongsekul [1]), the non-preemptive MPDQ was shown to be superior to the preemptive MPDQ using several performance measures. The non-preemptive MPDQ poses the challenge of being more difficult to analyze since, unlike in the preemptive case, the customer in service will not be ejected from service with an arrival into the system of a higher class customer.

This paper is organized in the following manner. In Section 2, we describe the states of the non-preemptive MPDQ and exhibit the
 generator of the system in Appendix. To give the reader an overview of the general movements in the system, some examples of the state transitions generated by the arrival and departure patterns of the MPDQ will be given. We then provide an analysis of the waiting time for each class of customers in Section 3. Finally, the paper concludes in Section 4 with numerical examples where some comparisons are made between the non-preemptive and preemptive dual queues using waiting times and other performance measures.

In the sequel, all vectors will be regarded as row vectors. The symbol \( e_i(d) \) represents the \( i \)th unit vector of dimension \( d \). \( \mathbb{R}^n \) is the real Euclidean space of dimension \( n \) and \( \mathcal{N}_0 \) refers to the set of non-negative integers. The symbol \( 1_{\{ \} } \) denotes the indicator function of the set \( \{ \} \) and \( \sum_B \) represents summation over all subscripts in the set \( B \). Finally, we let \( \lambda = \lambda_1 + \lambda_2 \).

2. System State Space and Generator Matrix

Let \( c_i = \) buffer size of queue \( i, i = 1, 2 \) (\( c_1 \) does not include the customer being served). The homogeneous ergodic Markov process describing the non-preemptive MPDQ takes values in a state space \( S \) to be described below. With a dual queue, \( S \) can be partitioned in two mutually exclusive sets, corresponding to the case when the secondary queue is empty and when it is not empty, respectively. In the first case, define \( S_1 = \{0\} \cup \{(i, j, p) : 0 \leq i + j \leq c_1; p = 1, 2\} \), where \( i = \) number of class 1, \( j = \) number of class 2 customers waiting in the primary queue, respectively and \( p \) denotes the class of customer in service (the state \( 0 \) represents the empty system). In the second case, define \( S_2 = \{(i, i', j', p) : 0 < i' + j' \leq c_2, i = 0, 1, \ldots, c_1; p = 1, 2\} \) where \( i' = \) number of customers of class 1 and \( j' = \) the number customers of class 2 in the secondary queue. We note that as the primary queue is full, there is no need to state the number of class 2 customers waiting in this queue as it is simply given by \( c_1 - i \). The state space is \( S = S_1 \cup S_2 \).
The steady-state distribution will be represented as a vector of probabilities dependent upon the states of the MPDQ. The labeling of the members of these states is done according to the following ordering:

\[ i, p = \{(i, 0, p), (i, 1, p), \ldots, (i, c_1 - i, p)\}, \quad (1) \]

\[ i, 0, p = \{(i, 0, 1, p), (i, 0, 2, p), \ldots, (i, 0, c_2, p)\} \quad (2) \]

for \( i = 0, 1, 2, \ldots, c_1, p = 1, 2 \) and \( i' = 1, 2, \ldots, c_2, \)

\[ i, i', p = \{(i, i', 0, p), (i, i', 1, p), \ldots, (i, i', c_2 - i', p)\}. \quad (3) \]

Note that the symbols that appear in the above sets of states are the variables that are kept fixed among the tuples. Thus, for example, \( i, p \) refers to the set of states, where \( i \) and \( p \) are fixed but \( j \) is allowed to vary. Similarly, \( i, i', p \) refers to the set of states, where \( i, i' \) and \( p \) are fixed but \( j' \) is allowed to vary. Let \( \pi_{0, p} \) and \( \pi_{i, i', p} \) be the probability vectors for these sets. A typical component of \( \pi_{0, p} \) is the probability \( \pi_{i, j, p} \) and that of \( \pi_{i, i', p} \) is the probability \( \pi_{i, j', p} \). Thus, for the primary queue,

\[ \pi_{i, p} = (\pi_{i, 0, p}, \pi_{i, 1, p}, \ldots, \pi_{i, c_1 - i, p}), \quad i = 0, \ldots, c_1, \quad p = 1, 2 \]

and for the secondary queue when \( i' = 0, \)

\[ \pi_{i, 0, p} = (\pi_{i, 0, 1, p}, \pi_{i, 0, 2, p}, \ldots, \pi_{i, 0, c_2, p}), \quad i = 0, \ldots, c_1, \quad p = 1, 2 \]

and when \( i' > 0, \)

\[ \pi_{i, i', p} = (\pi_{i, i', 0, p}, \pi_{i, i', 1, p}, \ldots, \pi_{i, i', c_2 - i', p}), \quad i = 0, \ldots, c_1, \quad i' = 1, \ldots, c_2, \quad p = 1, 2. \]

Finally, \( \pi \), the state distribution, can be expressed as

\[ \pi = (\pi_0, \pi_{0, 1}, \pi_{0, 0, 1}, \ldots, \pi_{0, c_2, 1}, \pi_{1, 1}, \pi_{1, 0, 1}, \ldots, \pi_{1, c_2, 1}, \ldots, \pi_{c_1, 1}, \pi_{c_1, 0, 1}, \]

\[ \ldots, \pi_{c_1, c_2, 1}, \pi_{0, 2}, \pi_{0, 0, 2}, \ldots, \pi_{0, c_2, 2}, \pi_{1, 2}, \pi_{1, 0, 2}, \ldots, \pi_{1, c_2, 2}, \ldots, \pi_{c_1, 2}, \pi_{c_1, 0, 2}, \ldots, \pi_{c_1, c_2, 2}), \]

where \( \pi \in \mathbb{R}^{(c_1 + 1)(c_2^2 + 3c_2 + c_1 + 2) + 1} \).
Since the MPDQ is a finite ergodic Markov process, its steady-state distribution $\pi$ exists and is obtained by solving the system of equations $\pi A = 0$, where $A$ is the generator matrix of the process and $0$ is a vector of zeroes of appropriate dimension. The matrix $A$ has the general structure

$$A = \begin{bmatrix}
    \Lambda_{0,1} & \Pi_{0,1} & \Phi_{0,1} & \Sigma_{0,1} \\
    \Omega_{1,1} & \Lambda_{1,1} & \Pi_{1,1} & \Phi_{1,1} \\
    \vdots & \vdots & \vdots & \vdots \\
    \Lambda_{q-1,1} & \Pi_{q-1,1} & \Omega_{q-1,1} & \Lambda_{q-1,1} \\
    \Omega_{q,1} & \Lambda_{q,1} & \Pi_{q,1} & \Phi_{q,1} \\
    \Omega_{0,2} & \Sigma_{0,2} & \Lambda_{0,2} & \Sigma_{0,2} \\
    \vdots & \vdots & \vdots & \vdots \\
    \Omega_{q-1,2} & \Sigma_{q-1,2} & \Lambda_{q-1,2} & \Sigma_{q-1,2} \\
    \Omega_{q-1,2} & \Sigma_{q-1,2} & \Lambda_{q-1,2} & \Sigma_{q-1,2} \\
    \vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}$$

Submatrices in $A$ denoted by the same symbol have the same structure and their sub-scripts correspond to the states set out in (1), (2) and (3). The entries in each submatrix are either zeroes or simple linear combinations of $\lambda_p$ and $\mu_p$ and they represent transitions between the various states. To save space, detailed structure of these submatrices will not be given here and interested reader may refer to Bedford and Zeephongsekul [2] for an analogous description in the preemptive MPDQ case. As is apparent from comparison with the aforementioned paper, the structure of the generator for the non-preemptive case is more complicated than the preemptive case. The system $\pi' A = 0$ can be solved, as in the preemptive case, using a combination of matrix-analytic method (Neuts [7], Latouche and Ramaswami [6]) and computational recursive algorithms (Bruell and Balbo [3]). The computational algorithm and a worked example using MAPLE can be obtained from the second author.
3. Waiting Time Analysis

We break the analysis into two parts. The first part deals with the waiting time of an arbitrary tagged Class 1 customer $P_1$ and the second with an arbitrary tagged Class 2 customer $P_2$ joining the system. In each case, we have to differentiate between the situations when the primary queue is full and when it is not, respectively. We will derive the actual distribution of the waiting time of $P_1$ and, through a more involved analysis, the Laplace-Stieltjes transforms of the waiting time of $P_2$. The approach utilizes some key ideas in Zeephongsekul and Bedford [13], which generalize the method used by Wagner [11] and Wagner and Kreiger [12] from single to dual queues.

3.1. Waiting times of class 1 customers

3.1.1. Primary queue is not full

This case corresponds essentially to the single finite buffer, non-preemptive queue considered by Wagner and Kreiger [12] and we include it here for the sake of completeness. $P_1$ either goes to the head of the line or joins a queue where the customer in service is either a class 1 or a class 2 customer. Let

$$
\Pi Q_1 = \pi_0 + \sum_{s=1}^{2} \sum_{0 \leq i+j \leq q_1} \pi_{i,j,s},
$$

i.e., the probability that the primary queue is not full, then, using the PASTA property (Poisson Arrivals See Time Averages), the conditional density of the waiting time of $P_1$, given that the primary queue is not full, is

$$
W_1^{(1)}(t) = \frac{1}{\Pi Q_1} \pi_0 \delta(t) + \sum_{0 \leq i+j \leq q_1} \pi_{i,j,1} f_{i+1,\mu_1}(t)
$$

$$
+ \sum_{0 \leq i+j \leq q_1} \pi_{i,j,2} f_{i,\mu_2} * f_{i,\mu_1}(t),
$$

(4)
where \( \ast \) refers to the convolution operator, \( \delta(t) \) is the Dirac delta function and \( f_{n, \mu_i}(t) \) refers to the Gamma distribution with density function

\[
f_{n, \mu_i}(t) = n^\mu_i \frac{(\mu_i t)^{n-1} e^{-\mu_i t}}{(n-1)!}
\]

\( t > 0, \ i = 1, 2 \) if \( \hat{W}_1^{(1)}(s) \) denote the Laplace-Stieltjes transform of \( W_1^{(1)}(t) \), i.e.,

\[
\hat{W}_1^{(1)}(s) = \int_0^\infty e^{-st}W_1^{(1)}(t)dt,
\]

\( \text{Re}(s) > 0 \), then it follows from (4) that

\[
\hat{W}_1^{(1)}(s) = \frac{1}{\prod_{Q_1}} (\pi_0 + \sum_{0 \leq i + j < c_1} \pi_{i, j, 1} (\frac{\mu_1}{s + \mu_1})^{i+1} + \sum_{0 \leq i + j < c_1} \pi_{i, j, 2} (\frac{\mu_2}{s + \mu_2})(\frac{\mu_1}{s + \mu_1})^i).
\]

\[ (5) \]

3.1.2. Primary queue is full

Assume that the secondary queue is not full so \( P_1 \) joins the secondary queue and either sees no Class 1 customer or at least one Class 1 customers in the primary queue. In the first case, \( P_1 \)'s waiting time is equal to the sum of the service time of the Class 2 customer at the head of the primary queue and all the service times of Class 1 customers ahead of \( P_1 \) in the secondary queue. In the second case, \( P_1 \)'s waiting time is equal to the sum of all the service times of all Class 1 customers in front of \( P_1 \) in the combined queue. Thus, the conditional density of the waiting time of \( P_1 \) given that queue 1 is full is

\[
W_2^{(1)}(t) = \frac{1}{\prod_{Q_2}} \sum_{i=0}^{c_1} (\pi_{i, c_1-i, 1} f_{i+1, \mu_1}(t) + \pi_{i, c_1-i, 2} f_{i, \mu_1} \ast f_{1, \mu_2}(t))
\]
\[ + \sum_{i=0}^{C_2} \sum_{0<i'+j'<c_2} \left\{ \pi_{i,i',j',2} f_{i} \mu_2 f_{i+i',1} \mu_1 (t) + \pi_{i,i',j',1} f_{i+i',1} \mu_1 (t) \right\}, \] (6)

where

\[ \Pi Q_2 = \sum_{s=1}^{2} \sum_{i=0}^{C_2} (\pi_{i,c_1-i,s} + \sum_{0<i'+j'<c_2} (\pi_{i,i',j',s})]. \]

The Laplace-Stieltjes transform of \( W^{(1)}_2(t) \) is therefore

\[ \tilde{W}^{(1)}_2(s) = \frac{1}{\Pi Q_2} \left\{ \sum_{i=0}^{C_2} (\pi_{i,c_1-i,1} \left( \frac{\mu_1}{\mu_1 + s} \right)^{i+1} + \pi_{i,c_1-i,2} \left( \frac{\mu_1}{\mu_1 + s} \right)^i \left( \frac{\mu_2}{\mu_2 + s} \right)) \right\} + \sum_{i=0}^{C_2} \sum_{0<i'+j'<c_2} (\pi_{i,i',j',2} \left( \frac{\mu_2}{\mu_2 + s} \right) \left( \frac{\mu_1}{\mu_1 + s} \right)^{i+i'} + \pi_{i,i',j',1} \left( \frac{\mu_1}{\mu_1 + s} \right)^{i+i'+1}). \] (7)

Finally, the density of the waiting time of \( P_1 \) provided she can join the queue, is given by

\[ W^{(1)}_1(t) = \frac{1}{\Pi Q_1 + \Pi Q_2} \left\{ \Pi Q_1 W^{(1)}_1(t) + \Pi Q_2 W^{(1)}_2(t) \right\}. \] (8)

### 3.2. Waiting times of class 2 customers

The derivation of the waiting time of a Class 2 customer is more involved than that of a Class 1 customer. This is attributed to the fact that \( P_2 \) will be delayed by subsequent arrivals of Class 1 customers while waiting for service. However, any subsequent Class 2 arrivals will reduce the amount of waiting rooms available thus improving \( P_2 \)'s chance of moving to the head of the line. We will first obtain the Laplace-Stieltjes transform of the waiting time of \( P_2 \) through an analysis of the first passage time of an appropriate stochastic process related to queue length. An algorithm will, then be obtained which allows us to find the \( k \)th moment, \( k = 1, 2, \ldots \) of the waiting time of a Class 2 customer. As for Class 1 customers, we break our discussion into two cases, when the primary queue is full and when it is not, respectively.
3.2.1 Primary queue is not full

In this case, it is possible for $P_2$ to enter, at $t = 0$, into the dual queue through the primary queue. Assume that the system is non-empty and define the stochastic process $Z(t), t \geq 0$ by

$$Z(t) = (p(t), l_{11}(t), L_{21}(t), F_1(t)),$$

where $p(t) =$ class of customer being served,

$l_{11}(t) =$ number of Class 1 customers waiting in primary queue;

$L_{21}(t) =$ number of Class 2 customers in front and including tagged customer waiting in primary queue,

$F_1(t) =$ amount of free spaces in primary queue.

Due to the assumptions underlying the MPDQ, $Z(t)$ is an ergodic homogeneous continuous time Markov chain taking values in the set

$$\mathcal{R}_1 = \{(p, l_{11}, L_{21}, F_1) \in \{1, 2\} \times \mathbb{N}_0^3 : 0 \leq l_{11} \leq c_1, 0 \leq L_{21} \leq c_1, 0 \leq F_1 \leq c_1\}.$$

In the sequel, we will consider the stationary behaviour of $Z(t)$, i.e., by letting $t \to \infty$. Let

$$C_1 = L_{21} + l_{11} + F_1,$$

$C_1$ measures the intrinsic capacity of the queue and provides an indication of when the queue is likely to be saturated. An arrival of a Class 2 customer will reduce $C_1$ by 1 whereas an arrival of either class of customers will reduce $F_1$ by 1 but does not cause a reduction in $C_1$. No arrival can enter the system when $F_1 = 0$. We will assume that, immediately after $P_2$ joined queue 1, $C_1 = c_1$; thus, $P_2$ was the last person to join the system.

Define the set of states $\mathcal{A}_1 \subset \mathcal{R}_1$ by

$$\mathcal{A}_1 = \{(2, l_{11}, 0, F_1) : 0 \leq l_{11} + F_1 \leq c_1\}.$$
Notice that \( \mathcal{A}_1 \) is the set of states where the tagged customer is in service. Let \( T_1(p, l_{11}, L_{21}, F_1) \) be the first passage time for the process \( Z \) to enter into \( \mathcal{A}_1 \) starting initially in state \( (p, l_{11}, L_{21}, F_1) \), where \( L_{21} \geq 1 \). This is precisely the waiting time of \( P_2 \) in the primary queue.

Denote the Laplace transform of \( T_1(p, l_{11}, L_{21}, F_1) \) by \( \hat{W}_{p, l_{11}, L_{21}, F_1}(s) \), i.e.,

\[
\hat{W}_{p, l_{11}, L_{21}, F_1}(s) = E(e^{-sT_1(p, l_{11}, L_{21}, F_1)}),
\]

where \( \text{Re}(s) > 0 \).

Using the PASTA property, the Laplace transform of the time to absorption of \( Z(t) \) into \( \mathcal{A}_1 \) given that a Class 2 customer can join the primary queue is

\[
\hat{W}_1^{(2)}(s) = \frac{1}{\prod q_i} \{ \pi_0 + \sum \pi_{i,j_1} \hat{W}_{1,i,j_1+1,c_1-i-j-1}(s) + \sum \pi_{i,j_2} \hat{W}_{2,i,j_1+1,c_1-i-j-1}(s) \}.
\]

**Relationships between \( k \)th moments of waiting times for class 2 customers in primary queue**

We develop several relationships between the \( k \)th moments of waiting times for Class 2 customers in the primary queue for various combinations of \( l_{11}, L_{21} \) and \( F_1 \). But first, we exhibit the following relationships (for \( p = 1, 2 \)) which follows using a conditional argument based on the first transition of the process \( Z(t) \) from an initial state to a state in \( \mathcal{R}_1 \) before absorption into \( \mathcal{A}_1 \) and the exponentially distributed time it takes to accomplish this first transition:

1. \( 1 \leq L_{21} \leq c_1 \),

\[
\hat{W}_{p, L_{21}, 0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{2, 0, L_{21}-1, 1}(s);
\]
\[(ii) \ 1 \leq l_{11} < c_1, \ 1 \leq L_{21} \leq c_1, \]
\[\hat{W}_{p, l_{11}, L_{21}, 0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{l_{11}-1, L_{21}, 1}(s);\]

\[(iii) \ 1 \leq L_{21} \leq c_1, \ 1 \leq F_1 < c_1, \]
\[\hat{W}_{p, 0, L_{21}, F_1}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{2, 0, L_{21}-1, F_1+1}(s) + \frac{\lambda_1}{s + \mu_p + \lambda} \hat{W}_{p, 0, L_{21}, F_1-1}(s);\]

\[(iv) \ 1 \leq l_{11} \leq c_1, \ 1 \leq L_{21} < c_1, \ 1 \leq F_1 < c_1, \]
\[\hat{W}_{p, l_{11}, L_{21}, F_1}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{l_{11}-1, L_{21}, F_1+1}(s) - \frac{\lambda_1}{s + \mu_p + \lambda} \hat{W}_{p, l_{11}, L_{21}, F_1-1}(s).\]

Let \(W_{p, l_{11}, L_{21}, F_1}^{(k)},\) denote the \(k\)th moment of the waiting time, \(k = 1, 2, \ldots,\) i.e.,
\[W_{p, l_{11}, L_{21}, F_1}^{(k)} = (-1)^k \frac{d^k \hat{W}_{p, l_{11}, L_{21}, F_1}(s)}{ds^k}|_{s = 0}. \tag{12}\]

The next set of relationships between these moments are obtained from (i) - (iv) above.

\[(i') \ 1 \leq L_{21} \leq c_1, \]
\[\mu_p W_{p, 0, L_{21}, 0}^{(k)} - \mu_p W_{2, 0, L_{21}-1, 1}^{(k)} = kW_{p, 0, L_{21}, 0}^{(k-1)}; \tag{13}\]

\[(ii') \ 1 \leq l_{11} \leq c_1, \ 1 \leq L_{21} \leq c_1, \]
\[\mu_p W_{p, l_{11}, L_{21}, 0}^{(k)} - \mu_p W_{l_{11}-1, L_{21}, 1}^{(k)} = kW_{p, l_{11}, L_{21}, 0}^{(k-1)}; \tag{14}\]

\[(iii') \ 1 \leq L_{21} \leq c_1, \ 1 \leq F_1 < c_1,\]
\[ \begin{align*}
(\mu_p + \lambda)W_p^{(k)}(0, L_{21}, F_1) - \lambda_1 W_p^{(k)}(1, L_{21}, F_1 - 1) &= kW_p^{(k-1)}(0, L_{21}, F_1) \\
&+ \mu_p W_p^{(k)}(2, L_{21} - 1, F_1 + 1),
\end{align*} \]

(15)

\[ \begin{align*}
(\mu_p + \lambda)W_p^{(k)}(l_{11}, L_{21}, F_1) - \lambda_1 W_p^{(k)}(l_{11} + 1, L_{21}, F_1 - 1) &= kW_p^{(k-1)}(l_{11}, L_{21}, F_1) \\
&+ \mu_p W_p^{(k)}(l_{11} + 1, L_{21}, F_1 - 1).
\end{align*} \]

(16)

A recursive algorithm for calculating the $k$th moment of waiting time for a class 2 customer in primary queue

In this section, we will utilize (13)-(16) to obtain a recursive algorithm to calculate the $k$th moment of waiting time. For fixed $L_{21} \geq 1$, $C_1$ and $p = 1, 2$, we define the following $(C_1 - L_{21} + 1)$ dimensional vector

\[ W_p^{(k)}(L_{21}, C_1) = \begin{pmatrix} W_p^{(k)}(0, L_{21}, C_1 - L_{21}) \\
W_p^{(k)}(1, L_{21}, C_1 - L_{21} - 1) \\
\vdots \\
W_p^{(k)}(C_1 - L_{21}, L_{21}, 0) \end{pmatrix}, \]

In the following, we obtain a recursive relationship for $W_p^{(k)}(L_{21}, C_1)$.

We will first assume that $C_1 > L_{21}$. Using (15) in case $l_{11} = 0$, (16) in cases $l_{11} = 1, 2, \ldots, C_1 - L_{21} - 1$ and (14) in case $l_{11} = C_1 - L_{21}$, we obtain the following vector-matrix equation:

\[ A_p(L_{21}, C_1) W_p^{(k)}(L_{21}, C_1) = kW_p^{(k-1)}(L_{21}, C_1) + \Gamma_p(L_{21}, C_1) W_p^{(k)}(L_{21}, C_1) + \Lambda_p(L_{21}, C_1) W_p^{(k)}(L_{21}, C_1) \\
+ \mu_p C_1 (C_1 - L_{21} + 1) C_1 (C_1 - L_{21} + 2) W_p^{(k)}(L_{21}, C_1), \]

(17)
where

$$A_{p,L_21,c_1} = \begin{pmatrix}
(\mu_p + \lambda) & -\lambda_1 & 0 & \cdots & 0 \\
0 & (\mu_p + \lambda) & -\lambda_1 & \cdots & \\
0 & 0 & (\mu_p + \lambda) & -\lambda_1 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & 0 \\
(\mu_p + \lambda) & -\lambda_1 & 0 & \cdots & \mu_p
\end{pmatrix} \in \mathbb{R}^{(c_1-L_21+1)\times(c_1-L_21+1)},
$$

$$\Gamma_{p,L_21,c_1} = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
\mu_p & 0 & \cdots & 0 & \vdots \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \mu_p \\
0 & \cdots & \cdots & \cdots & \mu_p
\end{pmatrix} \in \mathbb{R}^{(c_1-L_21+1)\times(c_1-L_21+1)},
$$

$$\Lambda_{p,L_21,c_1} = \begin{pmatrix}
\lambda_2 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & \lambda_2
\end{pmatrix} \in \mathbb{R}^{(c_1-L_21+1)\times(c_1-L_21+1)}.
$$

Finally, using (13), we observe that the recursive equation (17) still holds when $L_21 = C$ provided we interpret $A_{p,L_21,L_21} = (\mu_p)$, $\Gamma_{p,L_21,L_21} = 0$ and make $\Lambda_{p,L_21,L_21}$ void for that case.

### 3.2.2 Primary queue is full

We assume that the secondary queue is not full so it will be possible for the tagged customer $P_2$ to enter into the dual queueing system. Similar to $Z(t)$, let us define the stochastic process $Y(t)$, $t \geq 0$, by

$$Y(t) = (p(t), l_{11}(t), l_{12}(t), L_{22}(t), F_2(t))$$

where $p(t)$ = class of customer being served

- $l_{11}(t)$ = number of Class 1 customers waiting in primary queue;
- $l_{12}(t)$ = number of Class 1 customers in secondary queue;
$L_{22}(t) =$ number of Class 2 customers in front and including tagged customer waiting in the dual system,

and $F_2(t) =$ amount of free space in secondary queue:

$Y(t)$ is an ergodic homogeneous continuous time Markov chain which takes values in the set

$$
\mathcal{R}_2 = \{(p, l_{11}, l_{12}, 0, 0) : 0 \leq l_{11} \leq c_1, 0 \leq l_{12} \leq c_2, 0 \leq 2 L_{22} \leq c_1 + c_2, 0 \leq F_2(t) \leq c_2 \}.
$$

We let $t \to \infty$ and consider the stationary behaviour of $Y$.

Next, we define

$$
C_2 = L_{22} + l_{11} + l_{12} + F_2.
$$

Similar to $C_1$, $C_2$ accounts for the reduction in free spaces due to arrival of Class 2 customers after the arrivals of the tagged customer. An arrival of a Class 2 customer will reduce $C_2$ by 1 whereas an arrival of either class of customers will reduce $F_2$ by 1. The system is blocked when $F_2 = 0$. We will assume that, immediately after the tagged customer joined queue 2, $C_2 = c_1 + c_2$.

Define the set of states $\mathcal{A}_2 \subset \mathcal{R}_2$ by

$$
\mathcal{A}_2 = \{(2, l_{11}, l_{12}, 0, F_2) : 0 \leq l_{11} + l_{12} + F_2 \leq c_1 + c_2 \}.
$$

Let $T_2(p, l_{11}, l_{12}, L_{22}, F_2)$ be the first passage time for the process $Y$ to enter $\mathcal{A}_2$ starting initially in state $(p, l_{11}, l_{12}, L_{22}, F_2)$, where $L_{22} \geq 1$. This is the waiting time of $P_2$ in the dual queue. Denote the Laplace transform of $T_2(p, l_{11}, l_{12}, L_{22}, F_2)$ by $\hat{W}_{p, l_{11}, l_{12}, L_{22}, F_2}(s)$.

Appealing to the PASTA property, the Laplace transform of the time to absorption of $Y$ given that a Class 2 customer finds the primary queue full and the secondary queue not full is
\[ \hat{W}_2^{(2)}(s) = \frac{1}{\Pi Q_2} \sum_{p=1}^{2} \left( \sum_{i=0}^{c_1} \pi_{i,c_1-i,p} \hat{W}_{p,i,0,c_1+1-i,c_2-1}(s) \right) \]

\[ + \sum_{0<i'+j'<c_2} \pi_{i',j',p} \hat{W}_{p,i',i',j'+c_1+1-i,c_2-i'-j'-1}(s). \]  \hfill (19)

Combining (11) and (19), the Laplace transform of the waiting time of \( P_2 \) given that the dual system is not full is

\[ \hat{W}^{(2)}(s) = \frac{1}{\Pi Q_1 + \Pi Q_2} \{ \Pi Q_1 \hat{W}_1^{(2)}(s) + \Pi Q_2 \hat{W}_2^{(2)}(s) \}. \]  \hfill (20)

**Relationships between the \( k \)th moments of waiting times for class 2 customers in secondary queue**

Similar to the case when the primary queue is not full, we have the following set of relationships \((p = 1, 2)\) between \( \hat{W}_{h_1}, h_2, L_{22}, F_2(s) \) for various combinations of \( p, l_{11}, l_{12}, L_{22} \) and \( F_2 \).

(i) \( 0 < l_{12} < c_2, 0 < L_{22} \leq c = c_1 + c_2 \),

\[ \hat{W}_{p,0,h_2,L_{22},0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{2,1,h_2-1,L_{22}-1,1}(s); \]

(ii) \( 0 < L_{22} \leq c \),

\[ \hat{W}_{p,0,0,L_{22},0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{2,0,0,L_{22}-1,1}(s); \]

(iii) \( 0 < l_{11} \leq c_1, 0 < L_{22} \leq c \),

\[ \hat{W}_{p,l_{11},0,L_{22},0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{1,l_{11}-1,0,L_{22},1}(s); \]

(iv) \( 0 < l_{11} \leq c_1, 0 < l_{12} < c_2, 0 < L_{22} \leq c \),

\[ \hat{W}_{p,l_{11},l_{12},L_{22},0}(s) = \frac{\mu_p}{s + \mu_p} \hat{W}_{1,l_{11},l_{12}-1,L_{22},1}(s); \]
(v) $0 < l_1 < c_2$, $0 < L_{22} \leq c$, $0 < F_2 < c_2$,

$$
\hat{W}_{p,0,l_{12},L_{22},F_2}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{1,0,l_{12}-1,L_{22}-1,F_2+1}(s) + \frac{\lambda_1}{s + \mu_p + \lambda} \times \hat{W}_{p,0,l_{12}+1,L_{22},F_2-1}(s) + \frac{\lambda_2}{s + \mu_p + \lambda} \hat{W}_{p,0,l_{12},L_{22},F_2-1}(s);
$$

(vi) $0 < L_{22} \leq c$, $0 < F_2 < c_2$,

$$
\hat{W}_{p,0,0,L_{22},F_2}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{2,0,0,L_{22}-1,F_2+1}(s) + \frac{\lambda_1}{s + \mu_p + \lambda} \times \hat{W}_{p,0,1,L_{22},F_2-1}(s) + \frac{\lambda_2}{s + \mu_p + \lambda} \hat{W}_{p,0,0,L_{22},F_2-1}(s);
$$

(vii) $0 < l_1 \leq c_1$, $0 < L_{22} \leq c$, $0 < F_2 < c_2$,

$$
\hat{W}_{p,l_{11},0,L_{22},F_2}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{1,l_{11}-1,0,L_{22},F_2+1}(s) + \frac{\lambda_1}{s + \mu_p + \lambda} \times \hat{W}_{p,l_{11},1,L_{22},F_2-1}(s) + \frac{\lambda_2}{s + \mu_p + \lambda} \hat{W}_{p,l_{11},0,L_{22},F_2-1}(s);
$$

(viii) $0 < l_1 \leq c_1$, $0 < l_{12} < c_2$, $0 < L_{22} \leq c$, $0 < F_2 < c_2$,

$$
\hat{W}_{p,l_{11},l_{12},L_{22},F_2}(s) = \frac{\mu_p}{s + \mu_p + \lambda} \hat{W}_{1,l_{11},l_{12}-1,L_{22},F_2+1}(s) + \frac{\lambda_1}{s + \mu_p + \lambda} \times \hat{W}_{p,l_{11},l_{12}+1,L_{22},F_2-1}(s) + \frac{\lambda_2}{s + \mu_p + \lambda} \hat{W}_{p,l_{11},l_{12},L_{22},F_2-1}(s).
$$

The next set of relationships, between various $k$th moments of the waiting time $W_{l_{11},l_{12},L_{22},F_2}^{(k)}$, can then be obtained using (12):

(i') $0 < l_{12} < c_2$, $0 < L_{22} \leq c$,

$$
\mu_p W_{p,0,l_{12},L_{22},0}^{(k)} - \mu_p W_{2,1,l_{12}-1,L_{22}-1,1}^{(k)} = k W_{p,0,l_{12},L_{22},0}^{(k-1)};
$$

(21)
(ii) \[ 0 < L_{22} \leq c, \]
\[
\mu_p W_p^{(k)}_{p,0,0,L_{22},0} - \mu_p W^{(k)}_{p,0,0,L_{22}-1,1} = kW^{(k-1)}_{p,0,0,L_{22},0};
\] (22)

(iii) \[ 0 < l_{11} \leq c_1, \ 0 < L_{22} \leq c, \]
\[
\mu_p W_p^{(k)}_{p,l_{11},0,L_{22},0} - \mu_p W^{(k)}_{p,l_{11}-1,0,L_{22},1} = kW^{(k-1)}_{p,l_{11},0,L_{22},0};
\] (23)

(iv) \[ 0 < l_{11} \leq c_1, \ 0 < l_{12} < c_2, \ 0 < L_{22} \leq c, \]
\[
\mu_p W_p^{(k)}_{p,l_{11},l_{12},L_{22},0} - \mu_p W^{(k)}_{p,l_{11},l_{12}-1,L_{22},1} = kW^{(k-1)}_{p,l_{11},l_{12},L_{22},0};
\] (24)

(v) \[ 0 < l_{12} < c_2, \ 0 < L_{22} \leq c, \ 0 < F_2 < c_2, \]
\[
(\mu_p + \lambda)W_p^{(k)}_{p,0,l_{12},L_{22},F_2} - \lambda_1 W^{(k)}_{p,0,l_{12}+1,L_{22},F_2-1} = kW^{(k-1)}_{p,0,l_{12},L_{22},F_2};
\] (25)

(vi) \[ 0 < l_{11} \leq c, \ 0 < F_2 < c_2, \]
\[
(\mu_p + \lambda)W_p^{(k)}_{p,0,0,L_{22},F_2} - \lambda_1 W^{(k)}_{p,0,1,L_{22},F_2-1} = kW^{(k-1)}_{p,0,0,L_{22},F_2};
\] (26)

(vii) \[ 0 < l_{11} \leq c_1, \ 0 < L_{22} \leq c, \ 0 < F_2 < c_2, \]
\[
(\mu_p + \lambda)W_p^{(k)}_{p,l_{11},0,L_{22},F_2} - \lambda_1 W^{(k)}_{p,l_{11}+1,L_{22},F_2-1} = kW^{(k-1)}_{p,l_{11},0,L_{22},F_2};
\] (27)

(viii) \[ 0 < l_{11} \leq c_1, \ 0 < l_{12} < c_2, \ 0 < L_{22} \leq c, \ 0 < F_2 < c_2, \]
\[
(\mu_p + \lambda)W_p^{(k)}_{p,l_{11},l_{12},L_{22},F_2} - \lambda_1 W^{(k)}_{p,l_{11},l_{12}+1,L_{22},F_2-1} = kW^{(k-1)}_{p,l_{11},l_{12},L_{22},F_2};
\] (28)
A recursive algorithm for calculating $k$th moment of waiting time for a class 2 customer in queue 2

For fixed $l_{11} \geq 0$, $L_{22} > 0$, $C_2$ (by necessity, $C_2 \geq L_{22} + l_{11}$) and $p = 1, 2$, we define the following $C_2 - L_{22} - l_{11} + 1$ dimensional vector

$$W^{(k)}_{p, l_{11}, L_{22}, C_2} = \begin{pmatrix}
W^{(k)}_{p, l_{11}, 0, L_{22}, C_2 - L_{22} - l_{11}} \\
W^{(k)}_{p, l_{11}, 1, L_{22}, C_2 - L_{22} - l_{11} - 1} \\
\vdots \\
W^{(k)}_{p, l_{11}, C_2 - L_{22} - l_{11}, L_{22}, 0}
\end{pmatrix}.$$

We will obtain a recursive relationship for $W^{(k)}_{p, l_{11}, L_{22}, C_2}$ for various mutually exclusive and exhaustive cases.

(Case 1 : $C_2 > L_{22} + l_{11}$ and $l_{11} = 1, 2, \ldots, c_1$.)

Applying (27) when $l_{12} = 0$, (28) when $l_{12} = 1, 2, \ldots, C_2 - L_{22} - l_{11} - 1$ and (24) when $l_{12} = C_2 - L_{22} - l_{11}$, we obtain the following vector-matrix equation:

$$A_{p, l_{11}, L_{22}, C_2} W^{(k)}_{p, l_{11}, L_{22}, C_2} = \lambda_1 W^{(k-1)}_{p, l_{11}, L_{22}, C_2} + \Gamma_{p, l_{11}, L_{22}, C_2} W^{(k)}_{1, l_{11}, L_{22}, C_2} + A_{p, l_{11}, L_{22}, C_2} \\
\times W^{(k)}_{p, l_{11}, L_{22}, C_2 - 1} + \mu_p e_1 (C_2 - L_{22} - l_{11} + 1) e_1^T (C_2 - L_{22} - l_{11} + 2) \\
\times W^{(k)}_{1, l_{11} - 1, L_{22}, C_2},$$

where

$$A_{p, l_{11}, L_{22}, C_2} = \begin{pmatrix}
(\mu_p + \lambda) & -\lambda_1 & 0 & \cdots & 0 \\
0 & (\mu_p + \lambda) & -\lambda_1 & \cdots & \vdots \\
0 & 0 & (\mu_p + \lambda) & -\lambda_1 & \cdots \\
\vdots & 0 & 0 & (\mu_p + \lambda) & -\lambda_1 \\
0 & \cdots & 0 & 0 & \mu_p
\end{pmatrix}.$$
\[
\Gamma_{p, l_1, L_{22}, c_2} = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
\mu_p & 0 & \cdots & 0 & : \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots \\
0 & \cdots & \cdots & \mu_p & 0 \\
\end{pmatrix} \in \mathbb{R}((c_2 - L_{22} - l_1 + 1) \times (c_2 - L_{22} - l_1 + 1)),
\]

and

\[
\Lambda_{p, l_1, L_{22}, c_2} = \begin{pmatrix}
\lambda_2 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & : \\
\vdots & & \ddots & \ddots \\
\vdots & & & 0 & \lambda_2 \\
0 & \cdots & \cdots & 0 \\
\end{pmatrix} \in \mathbb{R}((c_2 - L_{22} - l_1 + 1) \times (c_2 - L_{22} - l_1)).
\]

(Case 2 : \( c_2 = L_{22} + l_1 \) and \( l_1 = 1, 2, \ldots, c_1 \).)

In this case, note that

\[
W^{(k)}_{p, l_1, L_{22}, L_{22} + l_1} = (W^{(k)}_{p, l_1, 0, L_{22}, 0})
\]

and so, applying (23), we obtain the following equation

\[
A_{p, l_1, L_{22}, L_{22} + l_1} W^{(k)}_{p, l_1, L_{22}, L_{22} + l_1} = kW^{(k-1)}_{p, l_1, L_{22}, L_{22} + l_1} + \mu_p e_1(1)e_1(2)W^{(k)}_{1, l_1 - 1, L_{22}, L_{22} + l_1}
\]

where \( A_{p, l_1, L_{22}, L_{22} + l_1} = (\mu_p) \).

(Case 3 : \( c_2 > L_{22} \) and \( l_1 = 0 \).)

Applying (26) when \( l_1 = 0 \), (25) when \( l_1 = 1, 2, \ldots, c_2 - L_{22} - 1 \) and (21) when \( l_1 = c_2 - L_{22} - l_{11} \), we obtain the following equation:

\[
A_{p, 0, L_{22}, c_2} W^{(k)}_{p, 0, L_{22}, c_2} = kW^{(k-1)}_{p, 0, L_{22}, c_2} + \Gamma_{p, 0, L_{22}, c_2} W^{(k)}_{2, 1, L_{22} - 1, c_2} + A_{p, 0, L_{22}, c_2} W^{(k)}_{p, 0, L_{22} - c_2 - 1} + \mu_p e_1(1) e_1(2)(c_2 - L_{22} + 1)e_1(c_2 - L_{22} + 2)W^{(k)}_{2, 0, L_{22} - 1, c_2}.
\]

(Case 4 : \( c_2 = L_{22} \) and \( l_1 = 0 \).)
In this case,
\[ W^{(k)}_{p,0,L_{22},L_{22}} = \left( W^{(k)}_{p,0,0,L_{22},0} \right), \]
and applying (23), we obtain
\[ A_{p,0,L_{22},L_{22}} W^{(k)}_{p,0,L_{22},L_{22}} = k W^{(k-1)}_{p,0,L_{22},L_{22}} + \mu_p e_1(t) e_{1}^{T} (2) W^{(k)}_{2,0,L_{22}-1,L_{22}}, \]
where \( A_{p,0,L_{22},L_{22}} = (\mu_p) \).

4. Numerical Examples

In our first example, we illustrate the computation of expected waiting times for both classes of customers using the algorithm described in Section 3 and compare these with the results obtained in Zeephongsekul and Bedford [13]. To avoid repeating the same sequence of steps, we will consider a small dual queue with \( c_1 = 1, c_2 = 3, \lambda_1 = 2, \lambda_2 = 3, \mu_1 = 5 \) and \( \mu_2 = 4 \). The stationary probabilities of all states in \( S_1 \cup S_2 \) are obtained using a similar algorithm to that given in Bedford and Zeephongsekul [2].

Consider first the expected waiting time of a Class 1 customer. Differentiating (5) and (7) and evaluating at \( s = 0 \) give \( E(W_{1}^{(1)}) = 0.126 \) unit and \( E(W_{2}^{(1)}) = 0.356 \) unit after some routine calculations. Therefore, (8) gives the expected waiting time of a Class 1 customer in the dual queue as \( E(W_{1}^{(1)}) = 0.286 \) unit.

We next turn to Class 2 customers. The calculation of her expected waiting time in the primary and secondary queue utilizes the algorithms described in Section 3.2.1 and Section 3.2.2. If she is able to enter the primary queue, (11) gives the expected waiting time there as \( E(W_{1}^{(2)}) = 0.126 \) unit where we note that (17) gives \( W_{1,0,1,0}^{(1)} = \frac{1}{\mu_1} = 0.20 \) and
\[ W_{2,0,1,0}^{(1)} = \frac{1}{\mu_2} = 0.25 \]
Next, we consider the expected waiting time of a class 2 customer in the secondary queue. From (19) and for \( p = 1, 2 \), the following vectors \( W_{0,2,4}^{(1)}, W_{1,1,4}^{(1)}, W_{0,3,4}^{(1)}, W_{1,2,4}^{(1)}, W_{p,0,4,4}^{(1)} \) and \( W_{p,1,3,4}^{(1)} \) are required in order to calculate \( E(W_2^{(2)}) \). These vectors are obtained by recursively solving (29) through to (32) for various cases ending with the absorption of the process \( Y(t) \) into \( A_2 \) resulting in the following values:

\[
W_{1,0,2,4}^{(1)} = (0.587, 0.969, 1.128), \ W_{0,2,4}^{(1)} = (0.686, 1.028, 1.178), \ W_{1,1,1,4}^{(1)} = (0.586, 0.869, 1.069), \ W_{2,1,1,4}^{(1)} = (0.683, 0.928, 1.119), \ W_{1,0,3,4}^{(1)} = (0.986, 1.477), \ W_{2,0,3,4}^{(1)} = (1.047, 1.527), \ W_{1,1,2,4}^{(1)} = (0.827, 1.027),
\]

\[
W_{2,1,2,4}^{(1)} = (0.865, 1.077), \ W_{1,0,4,4}^{(1)} = (1.247), \ W_{2,0,4,4}^{(1)} = (1.295), \ W_{1,1,3,4}^{(1)} = (1.186) \text{ and } W_{2,1,3,4}^{(1)} = (1.236).
\]

Using (19) we obtain \( E(W_2^{(2)}) = 0.992 \) unit, and finally, from (20), \( E(W_2^{(2)}) = 0.727 \) unit.

In Table 1 below, we display the expected waiting times of both classes of customers in the primary, secondary and dual queues for the non-preemptive dual queue (npMPDQ) and preemptive dual queue (preMPDQ) with the same parameters (cf. Zeephongsekul and Bedford [13]). As would be expected, the expected waiting times for Class 1 customers are smaller for the preMPDQ than for npMPDQ. However, notice that class 2 customers generally have a longer expected waiting time than class 1 customers in the preMPDQ. This has also been confirmed in extensive simulations carried out with larger queues and reported in Bedford and Zeephongsekul [1].
Table 1. Comparison of waiting times between preMPDQ and npMPDQ

<table>
<thead>
<tr>
<th></th>
<th>npMPDQ</th>
<th>preMPDQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary (class 1)</td>
<td>0.126</td>
<td>0.021</td>
</tr>
<tr>
<td>Secondary (class 1)</td>
<td>0.356</td>
<td>0.201</td>
</tr>
<tr>
<td>Dual queue (class 1)</td>
<td>0.286</td>
<td>0.162</td>
</tr>
<tr>
<td>Primary (class 2)</td>
<td>0.126</td>
<td>0.131</td>
</tr>
<tr>
<td>Secondary (class 2)</td>
<td>0.992</td>
<td>0.737</td>
</tr>
<tr>
<td>Dual queue (class 2)</td>
<td>0.727</td>
<td>0.605</td>
</tr>
</tbody>
</table>

In the next example, for a slightly larger systems ($c_1 = c_2 = 4$), we use the following performance measures introduced in Bedford and Zeephongsekul [2] to compare between the npMPDQ and preMPDQ:

(i) Utilization, i.e., the probability of finding the server busy,

$$\pi_B = 1 - \pi_0.$$

(ii) Probability of finding only class 1 customers in the system,

$$\pi^1 = \sum_{i=0}^{c_1} \pi_{i,0,1} + \sum_{i=1}^{c_2} \pi_{c_1,i',0,1}.$$

(iii) Probability of finding only class 2 customers in the system,

$$\pi^2 = \sum_{j=0}^{c_1} \pi_{0,j,2} + \sum_{j'=1}^{c_2} \pi_{0,0,j',2}.$$

(iv) Probability of an empty secondary queue,

$$\pi(Q_2 = 0) = \pi_0 + \sum_{i=0}^{c_1} \sum_{j=0}^{c_1} (\pi_{i,j,1} + \pi_{i,j,2}).$$

(v) Probability that a potential customer is lost to the system,

$$\pi_{\text{loss}} = \sum_{s=1}^{2} \sum_{i+j=c_2} \pi_{i,i',j,s}.$$

(vi) $L_i^i$, $i = 1, 2$, expected number of class $i$ customers in the system.
The results are summarized in Table 2 for various values of $\lambda_1, \lambda_2$ and fixed $\mu_1 = \mu_2 = 10$. Here, the traffic intensity $\rho = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2}$.

Some conclusions can be drawn from Table 2:

- $\pi_{\text{loss}}$ improves for npMPDQ over preMPDQ with increasing traffic intensity indicating better flow through the system. This is partly attributed to the non-ejection of a class 2 customer from service with an arrival of a class 1 customer into the primary queue.

- npMPDQ exhibits a lower probability of an empty secondary queue indicating a better utilization of the secondary queue. Again, this is expected as the npMPDQ is not interrupted during service, so a longer queue is allowed to form.

- Utilization is always better for npMPDQ when compared to preMPDQ.

- Expected number of class 2 customers is higher in a npMPDQ than a preMPDQ and, with high traffic intensity, the expected number of class 1 customers is lower for a npMPDQ. However, refer to the next example with respect to the total number of customers.

Finally, we compare between npMPDQ, preMPDQ and single non-preemptive queue (npSQ) in terms of the total expected number of customers in the queue. The dual queue has $c_1 = c_2 = 10$ and Class 1 and 2 service rates were fixed at 1. We display results where the arrival rate of class 1 customers was fixed at 0.2 while class 2 arrival rate is varied from 0.05 to 20. From the figure, it can be seen that as the arrival rate of class 2 customers increases, the npMPDQ has a lower expected number in the queue in comparison to the preMPDQ and npMPDQ.
Table 2. Comparison of performance characteristics between preMPDQ and npMPDQ

<table>
<thead>
<tr>
<th>Model</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\rho$</th>
<th>$\pi^1$</th>
<th>$\pi^2$</th>
<th>$\pi_B$</th>
<th>$\pi(Q_2=0)$</th>
<th>$\pi_{loss}$</th>
<th>$L^1$</th>
<th>$L^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>npMPDQ</td>
<td>1</td>
<td>2</td>
<td>.3</td>
<td>.0769</td>
<td>.3016</td>
<td>.5567</td>
<td>.8370</td>
<td>.0056</td>
<td>.54</td>
<td>1.45</td>
</tr>
<tr>
<td>Pre MPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0539</td>
<td>.2895</td>
<td>.3979</td>
<td>.9042</td>
<td>.0127</td>
<td>.22</td>
<td>.85</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>2</td>
<td>1</td>
<td>.3</td>
<td>.1390</td>
<td>.2259</td>
<td>.6062</td>
<td>.8040</td>
<td>.0053</td>
<td>.72</td>
<td>1.46</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.1446</td>
<td>.1290</td>
<td>.3337</td>
<td>.9654</td>
<td>.0034</td>
<td>.46</td>
<td>.41</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>2</td>
<td>2</td>
<td>.4</td>
<td>.0923</td>
<td>.3251</td>
<td>.7598</td>
<td>.7116</td>
<td>.0108</td>
<td>.82</td>
<td>2.16</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0993</td>
<td>.2709</td>
<td>.4843</td>
<td>.8935</td>
<td>.0148</td>
<td>.47</td>
<td>.95</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>2</td>
<td>5</td>
<td>.7</td>
<td>.0323</td>
<td>.4450</td>
<td>.9115</td>
<td>.5486</td>
<td>.0374</td>
<td>.73</td>
<td>3.16</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0323</td>
<td>.5455</td>
<td>.7786</td>
<td>.6785</td>
<td>.0774</td>
<td>.48</td>
<td>2.33</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>5</td>
<td>2</td>
<td>.7</td>
<td>.0543</td>
<td>.2476</td>
<td>.9597</td>
<td>.4624</td>
<td>.0222</td>
<td>1.06</td>
<td>2.63</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.1680</td>
<td>.2115</td>
<td>.7912</td>
<td>.7242</td>
<td>.0328</td>
<td>1.44</td>
<td>1.41</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>.0177</td>
<td>.2697</td>
<td>.9851</td>
<td>.4624</td>
<td>.0533</td>
<td>.85</td>
<td>3.10</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0414</td>
<td>.3799</td>
<td>.9002</td>
<td>.5146</td>
<td>.1195</td>
<td>1.34</td>
<td>2.99</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>.0228</td>
<td>.2900</td>
<td>.9900</td>
<td>.0347</td>
<td>.0555</td>
<td>1.05</td>
<td>3.24</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0853</td>
<td>.2226</td>
<td>.8971</td>
<td>.5261</td>
<td>.1094</td>
<td>2.11</td>
<td>2.54</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>.0130</td>
<td>.3875</td>
<td>.9760</td>
<td>.3525</td>
<td>.0711</td>
<td>.65</td>
<td>3.56</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0196</td>
<td>.5595</td>
<td>.8994</td>
<td>.5173</td>
<td>.1427</td>
<td>.74</td>
<td>3.17</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>.0070</td>
<td>.7469</td>
<td>.9497</td>
<td>.4378</td>
<td>.0929</td>
<td>.33</td>
<td>4.16</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.0057</td>
<td>.7686</td>
<td>.8940</td>
<td>.5372</td>
<td>.1888</td>
<td>.24</td>
<td>3.22</td>
</tr>
<tr>
<td>npMPDQ</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>.0283</td>
<td>.1970</td>
<td>.9940</td>
<td>.2860</td>
<td>.0398</td>
<td>1.26</td>
<td>2.06</td>
</tr>
<tr>
<td>preMPDQ</td>
<td></td>
<td></td>
<td></td>
<td>.2411</td>
<td>.0792</td>
<td>.8916</td>
<td>.5460</td>
<td>.1013</td>
<td>3.41</td>
<td>1.41</td>
</tr>
</tbody>
</table>
Figure 2. Expected number in queue for three queueing systems.

5. Conclusion

In this paper, we presented an extension of the MPDQ with a preemptive priority service scheme, considered in our previous papers, to the situation where the priority scheme is non-preemptive. The states of the system were described in detail and the rudiments of the generator matrix, indicating transitions between these states, were also presented. A detailed analysis of waiting times for both classes of customers were provided, together with recursive algorithms for obtaining the $k$th moment of their waiting times in both the primary and secondary queues. Finally, through several numerical examples, we compare the performances between the non-preemptive and preemptive MPDQ using several performance measures. From these examples, it is apparent that the MPDQ does meet the objective of provide quality service to high class customers without unduly compromising that of low class customers.
References


