International Journal of
Operations Research

International Journal of Operations Research Vol. 13, No. 2, 058–076 (2016)

# On Two Priority Multi-Server Queues with Impatient Customers

**Hung-Wei Li***

Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan

*Received June 2016; Revised June 2016; Accepted July 2016*

***Abstract*** **—** This study considers a system of multi-server queues with two classes of impatient customers: high-priority and low-priority. Customers join the system according to a Poisson process and customers may abandon service after entering the queue for an exponentially distributed duration with distinct rates. In this paper, we consider last come-first served (LCFS) and first come-first served (FCFS), and service time is assumed to be distributed exponentially among all customers. Deriving the Laplace transforms of the defined random variables and applying the matrix geometric method with direct truncation makes it possible to obtain an approximation of the stationary distribution in order to calculate the expected waiting time for both classes of customers. For each class of customer, we derive performance measures related to stationary probability distributions and conditional waiting times.

***Keywords*** **—** multi-server queues, impatient customers, non-preemptive service policy

## 1. INTRODUCTION

In this paper, we consider a system of multi-server queueing models with two classes of impatient customers, in which one class is given higher priority than the other. Our focus is on the non-preemptive priority policy related to the start of service, wherein service cannot be interrupted by other customers and impatient customers are prone to abandoning the system. Priority queues and the issue of abandonment are encountered in many applications, such as telecommunication networks, customer contact centers, and healthcare systems. Customers join the system according to a Poisson process, and customers may abandon service after waiting in a queue for an exponentially distributed duration. This study deals with two common service disciplines: last come-first served (LCFS) and first come-first served (FCFS). High priority customers are dealt with by focusing on performance measures related to queueing times and conditional waiting times in cases where service is provided and in cases of abandonment. The focus is on expected waiting times when dealing with low-priority customers. An approximation of stationary probability distribution is obtained using a direct truncation method. The proposed method also uses Laplace-Stieltjes transforms (LST) to measure the waiting time distribution.

Baccelli and Hebuterne [1] presented analysis of a queue system in which impatient customers are prone to abandonment. Garnett et al. [4] presented the sim- plest abandonment model, in which patience is exponentially distributed among all customers and the waiting capacity of the system is unlimited. Brandt and Brandt [2] presented a multi-server queueing system wherein customers may leave due to impatience. Choi et al. [3] introduce a simple approach to the analysis of M/M/c queues using a single class of customers and constant impatience time. This study included two classes of customer: class-1 (customers with impatience of constant duration rates) and class-2 (customers with patience and lower priority than class-1 customers).

Other researchers have dealt with non-preemptive priority queue systems. Kella and Yechiali [10] used probabilistic equivalence between the M/G/1 queue with multiple server vacations and the M/M/c system, in which the Laplace-Stieltjes transform is applied to waiting times. Sleptchenko [15] developed a multi-class, multi-server queueing system with non-preemptive priorities, in which steady-state probabilities are estimated. Zeltyn et al. [17] introduced a multi-server queue with K priority classes. The LST of waiting times is calculated explicitly and the LST of sojourn times is provided in an implicit form via a system of functional equations. Choi et al. [3] analyzes M/M/1 queues with impatient customers of higher priority. Kao and Wilson [9] analyzed non-preemptive priority queues with multiple servers and two priority classes. They developed a multi-server queueing system with two priority classes, in which high priority customers have non-preemptive priority over low priority customers.

Wang [16] considered a single-server with non-preemptive priority queuing for two classes of impatient customers. Iravani and Balcıoglu [5] analyzed three different problems in which one class of customer is given priority over another class. In the first problem, a single server receives two classes of customers with general service time requirements and follows a preemptive-resume policy. In the second model, the low-priority class is assumed to be patient and the single server chooses the next customer to serve according to a non-preemptive priority policy. The third problem involves a

* Corresponding author's email: 102751002@nccu.edu.tw

multi-server system that can be used to analyze a call center offering a call-back option to its impatient customers. Sarhangian and Balcioglu [14] analyzed three delay systems where different classes of impatient customers arrive according to independent Poisson processes. In all models, they obtain the LST of the virtual waiting time for each class by exploiting the level-crossing method. This enables us to compute the steady-state system performance measures. Jouini and Roubos [7] recently proposed multi-server queues with two classes of impatient customer: high-priority and low-priority. The two classes have different arrival rates but the same abandonment rates and service is performed according to LCFS and FCFS. They explicitly derive the LST of the defined random variables. They compare FCFS and LCFS and gain insights through numerical experiments. They have derived the expected waiting time of two classes of customers with the same impatience rate while Sarhangian and Balcioglu [14] considered the expected waiting time under only FCFS. The objective of paper is to investigate the expected waiting time of both classes customers with different impatient rates under LCFS and FCFS.

The method presented in this study can be summarized as follows. We assign an expression for the LST of the random variable related to busy periods in order to investigate high-priority customers in an LCFS queue. Our analysis of high-priority customers is based on the method proposed by Jouini et al. [8] which derives all moments of the probability distribution of conditional waiting time in cases where service is provided and in cases of abandonment. We also adopted the method of Jouini and Roubos [7] wherein we derive the stationary probability of class-1 customers in queue-1. For low-priority customers, we obtain the stationary probability of class-2 customers in queue-2 based on truncation and the matrix analytic method.

This paper is organized as follows. In Sections 2.1 and 2.2, we describe two classes of queueing model with impatient customers, and define the notation in this paper. In Section 2.3, we outline some of the basics pertaining to expected waiting times. In Sections 3.1 and 3.2, we outline a method by which to compute stationary probability for the number of customers in different queues. Section 3.3 proposes a method by which to compute the probability of all servers being idle at the same time. In Section 4.1, we analyze the expected waiting time in cases where service is provided and in cases of abandonment for high- and low-priority customers under conditions of FCFS. In Section 4.2, we analyze the expected waiting times in cases where service is provided and in cases of abandonment for high-priority customers, and analyze the expected waiting time for low-priority customers. In Section 5, we present numerical results. In Section 6, we draw conclusions and indicate directions for future research.

## 2. PRELIMINARIES

### 2.1 Modeling

Consider a queueing model with two types of customer: important (high-priority) customers denoted as class-1, and less important (low-priority) customers denoted as class-2.

In the following, we model a queueing system in which identical multi-servers attend to the two classes of impatient customer. We adhere to the non-preemptive rule, in which a server can attend to a class-2 customer only when there is no class-1 customer in the queue; and class-1 customers must wait for the completion of service for class-2 customers before being served.

The proposed model depicted in Figure 1 comprises a finite-buffer queue for class-1 customers (queue-1), an infinite buffer for class-2 customers (queue-2), as well as a set of s identical servers running in parallel. In cases where all servers are busy, newly arriving class-1 customers must wait in queue 1 and newly arriving class-2 customers must wait in queue-2. All of the servers provide identical service rates to both types of customer and the system is conservative (i.e., no server is forced to be idle when customers are waiting).

In the following we consider two cases of service: FCFS and LCFS. Class-i customers arrive at the queueing system according to an independent Poisson process at the following rate: $\lambda_i, i = 1, 2$ .The total arrival rate is denoted by $\lambda = \lambda_1 + \lambda_2$ .

Both classes of customer are assumed to be impatient, as follows. After entering a queue in which all servers are busy, each arriving customer waits for a random length of time. If the service time exceeds the waiting time, the customer abandons the queue. All customers are assumed to be lacking in memory of the previous queuing experience; therefore, when service begins, they must wait again until the service is completed. We assume that class-i customers have exponentially distributed time-to-abandon at rate $\gamma_i, i = 1, 2$ .

We denote by $EX^k$ the $k$ -th order moment of a given random variable $X$ , for $k \geq 1$ . We also denote by $f_X(\cdot)$ and $F_X(\cdot)$ the probability density function and the cumulative distribution function of $X$ . Furthermore, we know that the 1st order moment of a given random variable X is the expected value of $X$ .

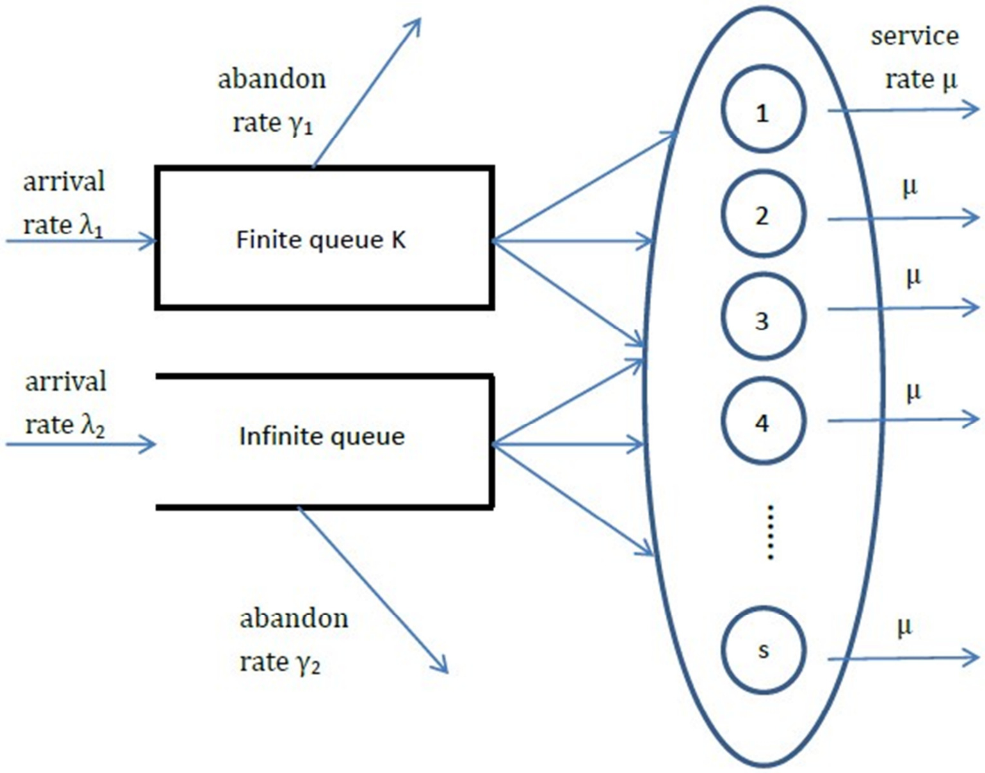


Figure 1: A queueing model with two impatient customers

### 2.2 Notation

Below are some important notations in this paper and they will be used in later sections.

$W$ : a random variable, the unconditional total waiting time of an arbitrary customer in the queue.

$W_i$ : a random variable, the unconditional waiting time of a class $i$ customer in the queue.

$W_{i,s}$ : a random variable, the conditional waiting time of a class $i$ customer, given that he will enter the service.

$p_{i,s}$ : the probability that a class $i$ customer in the service.

$W_{i,a}$ : the conditional waiting time of a class $i$ customer, given that he will abandon the service.

$p_{i,a}$ : the probability that that a class $i$ customer abandons the service.

$W_{i,w}$ : a random variable, the conditional waiting time of a class $i$ customer, given that he has to wait.

$p_w$ : the probability that a new arrival has to wait.

$W_{i,w,s}$ : a random variable, the conditional waiting time of a class $i$ customer, given that he has to wait, and he will enter service.

$p_{i,w,s}$ : the probability that a new arrival has to wait and he will enter service.

$Q_i$ : a random variable, the unconditional numbers of class i customers in the queue- $i$ .

A customer who does not abandon will necessarily enter the service, namely, we have $p_{i,s} + p_{i,a} = 1$ . A new customer who waits in the queue has two choices: He either enters the service or abandons in the queue, thereby implying that $p_{i,w,s} + p_{i,a} = p_w$ .

### 2.3 Expected waiting time

Because the arrival process is Poisson process, the probability of the new customer being class-i is $\lambda_i / \lambda$ , where $\lambda = \lambda_1 + \lambda_2$ . So, we have

$$EW = \frac{\lambda_1}{\lambda} EW_1 + \frac{\lambda_2}{\lambda} EW_2 .$$

Because of $p_{i,s} + p_{i,a} = 1$, we obtain

$$EW_i = p_{i,s} EW_{i,s} + p_{i,a} EW_{i,a} .$$

When the new customer arrivals, he may either enter the service immediately or wait in the queue. So

$$EW_i = p_w EW_{i,w} .$$

## 3. THE PROBABILITY OF NUMBER OF CUSTOMERS IN THE QUEUE

Denoted by $n_1(t)$ $n_2(t)$, nd $n(t)$ the number of customers in the qeueu-1, the number of customers in qeueu-2, and the total number of customers in both queues respectively, $n(t) = n_1(t) + n_2(t)$.

We assume that the maximum number of customers in the finite-buffer queue for class-1 is denoted by $K$.

In the long run, let $(n_1, n_2)$ represent the system state and their possible tran- sitions be given in the following table. Notice that $\gamma_1 \neq \gamma_2$.

Table 1: System state transition rates

| From | To | Rate | Condition |
|---|---|---|---|
| $(n_1,n_2)$ | $(n_1+1,n_2)$ | $\lambda_1$ | $0 \le n_1 \le K-1,\ \&\ n_2 \ge 0$ |
| $(n_1,n_2)$ | $(n_1,n_2+1)$ | $\lambda_2$ | $0 \le n_1 \le K,\ \&\ n_2 \ge 0$ |
| $(n_1,n_2)$ | $(n_1-1,n_2)$ | $s \cdot \mu + n_1 \cdot \gamma_1$ | $1 \le n_1 \le K,\ \&\ n_2 \ge 0$ |
| $(n_1,n_2)$ | $(n_1,n_2-1)$ | $s \cdot \mu + n_2 \cdot \gamma_2$ | $n_1 = 0,\ \&\ n_2 \in \mathbb{N}$ |
| $(n_1,n_2)$ | $(n_1,n_2-1)$ | $n_2 \cdot \gamma_2$ | $n_1 \neq 0,\ \&\ n_2 \in \mathbb{N}$ |

### 3.1 Analysis of high-priorty customers

Borrowing the terminology from Jouini and Roubos [7], we have the stationary probability of number of high-priority customers in this section. The stationary probability of number of customers in the service is denoted by $p_k, 1 \le k \le s-1$.

Thus $p_w$ is given by

$$p_w = 1 - \sum_{k=0}^{s-1} p_k .$$

The normalization condition gives

$$\sum_{k=0}^{s-1} p_k + \sum_{k=0}^{K} \sum_{j=0}^{\infty} \pi_{k,j} = 1$$

where $\pi_{k,j}$ is the stationary probability, $k$ class-1 customers are in queue-1, and $j$ class-2 customers are in queue-2, when all servers are busy.

Let $\pi_1(k)$ denote the stationary probability that all servers busy and there are $k$ class-1 customers in queue-1. Summing up all the states with respect to $k$, this system corresponding the M/M/s+M queue. The equation for class-1 leads to

$$\pi_1(k) = \frac{\lambda_1^k}{\prod_{j=1}^{k} (s\mu + j\gamma_1)} \pi_1(0),$$

And we know that

$$\pi_1(k) = \sum_{j=0}^{\infty} \pi_{k,j},$$

Applying $\pi_1(k)$ to normalization condition we have

$$\sum_{k=0}^{s-1} p_k + \sum_{k=0}^{K} \pi_1(k) = 1,$$

Also, it can be written as

$$\sum_{k=0}^{s-1} p_k + \sum_{k=0}^{K} \frac{\lambda_1^k}{\prod_{j=1}^{k}(s\mu + j\gamma_1)} \pi_1(0) = 1$$

By the above equation we obtain

$$\pi_1(0) = (1 - \sum_{k=0}^{s-1} p_k)(\sum_{k=0}^{K} \frac{\lambda_1^k}{\prod_{j=1}^{k}(s\mu + j\gamma_1)})^{-1},$$

Therefore we have the expected number of customers in queue-1 which is given by

$$EQ_1 = \sum_{k=1}^{K} k\pi_1(k).$$

Because of the queue-1 is finite, we have the blacking probability $p_b$ as follows:

$$p_b = 1 - \left(\sum_{k=0}^{s-1} p_k + \sum_{k=0}^{K-1} \pi_1(k)\right).$$

From Sarhangian and Balcioglu [14], $p_{i,s}$ is given by

$$p_{i,3} = 1 - p_{i,2} = p_0 + \int_0^{\infty} e^{-\gamma_i y} f_i(y)\,dy = p_0 + \tilde{f}_i(\gamma_i),$$

where $f_i(x)$ denoted the density function of the virtual waiting time for class-i customers and $\tilde{f}_i(s)$ denoted the LST's of $f_i(x)$.

By Eq. (13) in Sarhangian and Balcioglu [14], the waiting time distribution of class-i customer in the queue is

$$P(Wi \le x) = 1 - e^{-\gamma_i x} + e^{-\gamma_i x} F_i(x),$$

where $F_i(x)$ denoted the cumulative distribution function of the virtual waiting time for class-i customers. And it has the expected value

$$EW_i = \int_0^{\infty} P(W_i > x)\,dx = \int_0^{\infty} \mathrm{e}^{-\gamma_i x} \overline{F}_i(x)\,dx = \frac{1 - \tilde{f}_i(\gamma_i) - p_0}{\gamma_i} = \frac{p_{i,2}}{\gamma_i}.$$

So, we have

$$p_{i,2} = \frac{\gamma_i EQ_i}{\lambda_i},$$

$$p_{i,3} = 1 - p_{i,1},\ i = 1, 2.$$

### 3.2 Analysis of low-priority customers

Let $\pi_{k,j}$ denote the stationary distribution of $\{n_1(t), n_2(t); 0 \le n_1(t) \le K, n_2(t) \ge 0, t \ge 0\}$, i.e,

$$\pi_{k,j} = \lim_{t\to\infty} P[n_1(t) = k, n_2(t) = j].$$

Based on the classification of the states, we derive the infinitesimal generator matrix $\mathbf{Q}$ for the QBD process as follows:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \mathbf{Q}_2 \\ \mathbf{Q}_3 \mathbf{Q}_4 \end{bmatrix},$$

We also define $p_k, k = 0, 1, 2, \cdots, s-1$ be the probability that $k$ customers is served, and denote $p = (p_0, p_1, \cdots p_{s-1})$.
We then have

$$(p, \boldsymbol{\pi})\mathbf{Q} = 0,$$

$$(p, \boldsymbol{\pi})1^T = 1,$$

also it can be written as

$$\boldsymbol{\pi} 1^T = p_w,$$

where 0 and $1^T$ denote a row vector of zeros and column vector of ones. And $\mathbf{Q}_k, \mathrm{k} = 1, 2, 3, 4$, can be shown as follows:

$$\mathbf{Q}_1=\begin{bmatrix} -\lambda & \lambda & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & -\lambda-(s-2)\mu & \lambda \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & (s-1)\mu & -\lambda-(s-1)\mu \end{bmatrix}$$

where $\mathbf{Q}_1$ is $s\times s$ matrix.

$$\mathbf{Q}_2=\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \mathbf{E} & 0 & \cdots & 0 \end{bmatrix},$$

$$\mathbf{Q}_3=\begin{bmatrix} 0 & 0 & \cdots & \mathbf{D} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{Q}_4=\begin{bmatrix} \mathbf{B}_0 & \mathbf{A}_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \mathbf{C}_0 & \mathbf{B}_1 & \mathbf{A}_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \mathbf{C}_1 & \mathbf{B}_2 & \mathbf{A}_3 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \mathbf{C}_2 & \mathbf{B}_3 & \mathbf{A}_4 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

Each elements in $\mathbf{Q}_k$ , forcis defined as follows:

$$\mathbf{D}=\begin{bmatrix} s\mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and $\mathbf{D}$ is the $(K+1)\times 1$ column matrix.

$$\mathbf{E}=\begin{bmatrix}\lambda & 0 & \cdots & 0\end{bmatrix},$$

and $\mathbf{E}$ is the $1\times(K+1)$ row matrix.

$$\mathbf{A}_j=\begin{bmatrix} \lambda_2 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \lambda_2 \end{bmatrix},$$

where $\mathbf{A}_j$ is $(K+1)\times(K+1)$ matrices for all $j=1,2,\cdots$.

For brevity, we use the following simplified notation.

$$U_{j,k}=-\left(s\mu+k\gamma_1+j\gamma_2+\lambda_1+\lambda_2\right),\forall j=0,1,2,\cdots,\forall k=0,1,2,\cdots,K-1,$$
$$U_{j,K}=-\left(s\mu+K\gamma_1+j\gamma_2+\lambda_2\right),\forall j=0,1,2,\cdots,$$

And

$$T_k=s\mu+k\gamma_1,\forall j=0,1,2,\cdots,K, \tag{1}$$

Then we have $\mathbf{B}_j$ and $\mathbf{C}_j$ written as follows :

$$\mathbf{B}_j = \begin{bmatrix} U_{j,0} & \lambda_1 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ T_1 & U_{j,1} & \lambda_1 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & T_2 & U_{j,2} & \lambda_1 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & T_3 & 0 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & T_{K-1} & U_{j,K-1} & \lambda_1 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & T_K & U_{j,K} \end{bmatrix}$$

$$\mathbf{C}_j = \begin{bmatrix} s\mu+(j+1)\gamma_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & (j+1)\gamma_2 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & (j+1)\gamma_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & (j+1)\gamma_2 \end{bmatrix},$$

where $\mathbf{B}_j$ and $\mathbf{C}_j$ are $(K+1)\times(K+1)$ square matrices.

We have the state $(k, j)$ for $0 \le k \le K$ and $j \ge 0$, i.e., $(0,\ 0),(1,\ 0),\cdots,(\mathrm{K},\ 0),(0,\ 1),(1,\ 1),\cdots$.

Let $\boldsymbol{\pi}_j$ and $\boldsymbol{\pi}$ denote $\boldsymbol{\pi}_j = (\boldsymbol{\pi}_{0,j}, \boldsymbol{\pi}_{1,j}, \cdots, \boldsymbol{\pi}_{K,j})$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_j, \cdots)$. Moreover, according to the matrix analytic method, $\boldsymbol{\pi}_j$ has a matrix-product form solution given by

$$\boldsymbol{\pi}_j = \boldsymbol{\pi}_0 \prod_{k=1}^{j} R^{(k)},\ j \ge 1.$$

The vector $\boldsymbol{\pi}_0$ satisfies the following equations:

$$p_{s-2}\lambda + p_{s-1}\left[-\lambda + (s-1)\mu\right] + \boldsymbol{\pi}_0\mathbf{D} = 0,$$
$$p_{s-1}\mathbf{E} + \boldsymbol{\pi}_0\mathbf{B}_0 + \boldsymbol{\pi}_1\mathbf{C}_0 = 0.$$

Also, it can be written as

$$\boldsymbol{\pi}_0\mathbf{D} = -p_{s-2}\lambda - p_{s-1}\left[-\lambda + (s-1)\mu\right],$$
$$\boldsymbol{\pi}_0\left(\mathbf{B}_0 + R^{(1)}\mathbf{C}_0\right) = -p_{s-1}\mathbf{E},$$

given $p_{s-1}$, $p_{s-2}$ and $R^{(1)}$ we can obtain the unique solution of $\boldsymbol{\pi}_0$.
And $\{R^{(n)}; n \ge 1\}$ is the minimal nonnegative solution to the following system of

$$\mathbf{A}_n + R^{(n)}\mathbf{B}_n + R^{(n)}R^{(n+1)}\mathbf{C}_n = \mathbf{O}, n \ge 1,$$

where $\mathbf{O}$ is the matrix which each element is zeros.

So, if we obtain $R^{(n)}$ for all $n$, then we have $\boldsymbol{\pi}_j$ for $j \ge 1$.

### 3.2.1 Truncation point

In the previous section, the stationary distribution $\boldsymbol{\pi}_n$ of our QBD process can be expressed by a matrix-product rate $R^{(n)}$. But $R^{(n)}$ has no explicit form. So, we can only compute $\overline{\boldsymbol{\pi}}_n$ as the approximate of $\boldsymbol{\pi}_n$ for $0 \le n \le N$, where $N$ is the truncation point of $\boldsymbol{\pi}_n$.

We advance a method to compute $N$. For this truncation point $N$, we can obtain an approximate $\overline{\boldsymbol{\pi}}_n$. We need to choose a number $N$ by which the tail probability $\overline{\boldsymbol{\pi}}_n$ is small enough to be neglected.

Because of our model is similar to an M/M/s queue, we can use the stationary probability expression in M/M/s model to find $N$, such that the tail probability is small enough. Our model has arriving rate $\lambda_1$ and $\lambda_2$, abandon rate $\gamma_1$ and $\gamma_2$, and the maximum service rate $s\times\mu$.

Let

$$\overline{\lambda} = \lambda_1 + \lambda_2 - \gamma_1 - \gamma_2,$$

by using $\overline{\lambda}$, $s$, $\mu$ in an M/M/s queue.

We have

$$N = \inf\left\{ n \,\middle|\, 1 - \sum_{i=0}^{n} \overline{p}_i < \epsilon \right\},$$

where $\overline{p}_i$ is the stationary probability of M/M/s queue, and $\epsilon$ we can chosen small enough.

### 3.2.2 Matrix-product rate

When the truncation point $N$ is given, the stationary probability distribution can be obtained. Moreover, because of $N$ is chosen sufficiently large, so the probability after $N$ is small enough that we can disregard. In this paper, we use a method to compute $R^{(n)}$ in Proposition 1 and 2. Proposition 1 is borrowed from Proposition 1 in Phung-Duc et al. [13] and proposition 2 is borrowed from Proposition 2.4 in Phung-Duc et al. [12].

**Proposition 1.** *Let* $S$ *denote a set of real square matrices of* $(N+1)\times(N+1)$. *We define* $G_n : S \to S$ *as*

$$G_n(X) = \mathbf{A}_n(-\mathbf{B}_n - X\mathbf{C}_n)^{-1}, n \geq 1.$$

*Then, the matrices* $R^{(n)}$ *for* $n \geq 1$ *satisfies the following backward recursive equation.*

$$R^{(n)} = G_n\left(R^{(n+1)}\right) = G_n \circ G_{n+1} \circ \cdots \circ G_{n+k} \circ \cdots, n \geq 1,$$

*where* $f \circ g(\cdot) = f(g(\cdot))$ *is the composition function.*

This proposition tells that $R^{(n)}$ can be obtained for an infinite matrix composite. And the next proposition will provide a method that converges to $R^{(n)}$

**Proposition 2.** *If we define the matrix* $G_k^{(n)}$ *for* $n \geq 0$ *and* $k \geq 0$ *by*

$$G_0^{(n)} = O,$$

$$G_k^{(n)} = G_n\left(G_{k-1}^{(n+1)}\right) = \cdots = G_n \circ G_{n+1} \circ \cdots \circ G_{n+k-1}(O), k \geq 1, n \geq 1,$$

*then we have*

$$\lim_{k\to\infty} G_k^{(n)} = R^{(n)}, n \geq 1.$$

The proposition means that $G_k^{(n)}$ is the k-th order approximate of $R^{(n)}$. It also tells that we can obtain $R^{(n)}$ sufficiently close by a fairly large $k$.

## 4. METHOD TO COMPUTE THE PROBABILITY OF ALL SERVERS IDLE

As the last section, the probability of $p_k$ can be given by M/M/s queue :

$$p_k = \frac{\lambda^k}{k!\mu^k} p_0, 0 \leq k \leq s-1,$$

where $p_0$ is the probability of all servers idle.

In order to compute $p_k$, we need a method to calculate $p_0$. Moreover, we propose a proposition to compute $p_0$.

**Proposition 3.** *If we define the probability* $x_0$ *by*

$$x_0 = 1,$$

$$x_k = \frac{\lambda^k}{k!\mu^k} x_0, 0 \leq k \leq s-1,$$

$$\boldsymbol{x}_{s+k} = \prod_{j=1}^{k} R^{(j)} \boldsymbol{x}_s, 1 \leq k \leq N,$$

*and define*

$$(x_0, x_1, \cdots, x_{s-1}, \boldsymbol{x})\mathbf{Q} = 0,$$

*where* $\boldsymbol{x}$ *denote* $\boldsymbol{x} = (\boldsymbol{x}_s, \boldsymbol{x}_{s+1}, \cdots)$.

Then we have

$$p_0 = \left( \sum_{k=0}^{s-1} x_k + \sum_{k=0}^{N} \boldsymbol{x}_{s+k} \boldsymbol{e} \right)^{-1}. \tag{2}$$

*Proof.* First, we know that

$$x_0 = tp_0 = 1,$$

where $p_0$ is define before. Also, it can be written as

$$p_0 = t^{-1}.$$

Second, we need to prove that $x_k = tp_k, k = 0,1,2,\cdots,s-1$ and $\boldsymbol{x}_{s+k} = t\pi_k, k = 0,1,2,\cdots,K.$ Clearly, because of

$$x_k = \frac{\lambda^k}{k!\mu^k} x_0, 0 \le k \le s-1,$$

then we have

$$\begin{aligned} x_k &= \frac{\lambda^k}{k!\mu^k} tp_0 \\ &= tp_k, \end{aligned}$$

for all $k = 0,1,\cdots,s-1$.

The vector $\boldsymbol{x}_s$ is a solution from the equations

$$\begin{aligned} x_{s-2}\lambda + x_{s-1}\left[-\lambda + (s-1)\mu\right] + \boldsymbol{x}_s \mathbf{D} = 0, \\ x_{s-1}\mathbf{E} + \boldsymbol{x}_s \mathbf{B}_0 + \boldsymbol{x}_{s+1}\mathbf{C}_0 = 0. \end{aligned}$$

Also, it can be written as

$$\begin{aligned} x_{0,0} s\mu = t\{-p_{s-2}\lambda - p_{s-1}\left[-\lambda + (s-1)\mu\right]\}, \\ \boldsymbol{x}_s \left(\mathbf{B}_0 + R^{(1)}\mathbf{C}_0\right) = t\left(-p_{s-1}\mathbf{E}\right), \end{aligned}$$

therefore we have

$$\begin{aligned} x_{0,0} &= \frac{t\left\{-p_{s-2}\lambda - p_{s-1}\left[-\lambda + (s-1)\mu\right]\right\}}{s\mu} \\ &= t\pi_{0,0}, \\ \boldsymbol{x}_s &= t\left(-p_{s-1}\mathbf{E}\right)\left(\mathbf{B}_0 + R^{(1)}\mathbf{C}_0\right)^{-1} \\ &= t\boldsymbol{\pi}_0, \end{aligned}$$

it means that $\boldsymbol{x}_s = t\boldsymbol{\pi}_0$.

By the matrix-product forms

$$\boldsymbol{x}_{s+k} = \boldsymbol{x}_s \prod_{n=1}^{k} R^{(n)}, 1 \le k \le N,$$

then we have

$$\begin{aligned} \boldsymbol{x}_{s+k} &= t\boldsymbol{\pi}_0 \prod_{n=1}^{k} R^{(n)} \\ &= t\boldsymbol{\pi}_k, 1 \le k \le N. \end{aligned}$$

Finally, we obtain

$$\begin{aligned} &\sum_{k=0}^{s-1} x_k + \sum_{k=0}^{N} \boldsymbol{x}_{s+k}\boldsymbol{e} \\ &= t\left( \sum_{k=0}^{s-1} p_k + \sum_{k=0}^{N} \boldsymbol{\pi}_k \boldsymbol{e} \right) \\ &= t \times 1 \\ &= t. \end{aligned}$$

Moreover,

$$p_0 = t^{-1}$$
$$= \left( \sum_{k=0}^{s-1} x_k + \sum_{k=0}^{N} \boldsymbol{x}_{s+k} \boldsymbol{e} \right)^{-1}.$$

## 5. ANALYSIS OF QUEUEING DELAYS

For $n \geq 1$, an $n$-busy period is defined as the any point of time from the arriving customer to a busy M/M/s + M system with $n-1$ waiting customers in the queue until the time at which one server becomes idle.

Moreover, the 0-busy period is the classical busy-period definition defined to begin with the arriving customer to a system with $s-1$ busy servers and end with one server becomes idle again.

Suppose there is only one class of customers where the arrival rate is $\lambda$, the abandonment rate $\gamma$ and the service rate is $\mu$. We denoted the length of an $n$ - busy period by $B_{n,\lambda,\gamma}$, for $n \geq 1$. The Laplace-Stieltjes transform of the pdf of $B_{n,\lambda,\gamma}$ by $\tilde{F}_{\mathrm{B}_{n,\lambda,\gamma}}(x)$, is found from Eq. (1) of Iravani and Balcıoglu [5] by substituting $\tilde{b}_1(x) = s\mu / (x + s\mu)$. We obtain

$$\tilde{F}_{\mathrm{B}_{n,\lambda,\gamma}}(x) = \frac{\dfrac{s\mu}{x+s\mu} + \sum_{i=1}^{\infty} (-1)^i \left[ \prod_{j=0}^{i-1} \left( 1 - \dfrac{s\mu}{x+s\mu+j\gamma} \right) \right] \dfrac{s\mu}{x+s\mu+i\gamma} \Theta(n,i)}{1 + \sum_{i=1}^{\infty} \dfrac{\lambda^i}{i!\gamma^i} \left[ \prod_{j=0}^{i-1} \left( 1 - \dfrac{s\mu}{x+s\mu+j\gamma} \right) \right]},$$

with

$$\Theta(n,i) = \begin{cases} \sum_{j=0}^{i} \dfrac{(-1)^j \lambda^j}{j!\gamma^j} \dbinom{n}{i-j}, & 1 \leq i \leq n, \\ \sum_{j=i-n}^{i} \dfrac{(-1)^j \lambda^j}{j!\gamma^j} \dbinom{n}{i-j}, & i > n, \end{cases}$$

And the 0-busy period in finite queue is given by

$$\tilde{F}_{\mathrm{B}_{0,\lambda,\gamma}}(x) = \frac{\dfrac{s\mu}{x+s\mu} + \sum_{i=1}^{K} \dfrac{\lambda^i}{i!\gamma^i} \left[ \prod_{j=0}^{i-1} \left( 1 - \dfrac{s\mu}{x+s\mu+j\gamma} \right) \right] \dfrac{s\mu}{x+s\mu+i\gamma}}{1 + \sum_{i=1}^{K} \dfrac{\lambda^i}{i!\gamma^i} \left[ \prod_{j=0}^{i-1} \left( 1 - \dfrac{s\mu}{x+s\mu+j\gamma} \right) \right]},$$

In section 4.2, because our analyze the performance of class-1 that have $\gamma_1$ abandon rate and $\lambda_1$ arrival rate, we assume that $\gamma = \gamma_1$ and $\lambda = \lambda_1$.

In section 5.1 and 5.2, our method is based on Jouini et al. [8], Jouini and Roubos [7], but we can use the matrix geometric method to calculate the stationary probability with different abandonment rates $\gamma_1 \neq \gamma_2$. Notice that the matrix geometric approach can also compute the stationary probability for different service rates. Thus, it may compute the conditional expected waiting time in robust settings.

### 5.1 Analysis of Model$_{\mathrm{FCFS}}$

For high and low priority customers, we focus on the conditional probability of high-priority customers given served and abandon, and also compute the expected waiting time in the queue for low-priority customers.

#### High-Priority Customers

In the follows, we use the $k$ -th order moment of $W_{1,s}$ and $W_{1,a}$ to compute the expected waiting which it is 1-st moment of $W_{1,s}$ and $W_{1,a}$.

Consider the new class-1 customer arrival the system, and he find all servers busy, and $n_1$ waiting customers ahead in queue-1. And the contrary case (at least one server idle), he will immediately enter the service.

The class-2 customers, because of their low-priority, will not affect the class-1 customers sojourn time. Using Jouini et al. [8], Jouini and Roubos [7], we obtain

$$EW_{1,s}^k = \frac{1}{p_{1,s}} \sum_{n_1=0}^{K} p_1(n_1) \Psi_{n_1+1} EY_{n_1+1}^k, \tag{3}$$

with

$$\Psi_{n_1} = \prod_{i=1}^{n_1}\left(1-\frac{\gamma_1}{s\mu+i\gamma_1}\right)=\frac{s\mu}{s\mu+n_1\gamma_1}.$$

$Y_{n_1}$, a random variable, is the summation of $n_1$ independent exponential distributions with parameters $s\mu+\gamma_1$, $s\mu+2\gamma_1,\cdots,s\mu+n_1\gamma_1$. Its first moment is

$$EY_{n_1} = \sum_{j=1}^{n_1}\frac{1}{s\mu+j\gamma_1}.$$

Considering the $EW_{1,a}$, the new class-1 customer arrival who find at least one server idle, the expected conditional waiting time of class one customer given abandon. Let $Z_{n_1+1}$ denote the random variable measuring her sojourn time in the queue before abandonment. Removing the condition on $n_1$, we obtain

$$EW_{1,a} = \frac{1}{p_{1,a}}\sum_{n_1=0}^{K}p_1\left(n_1\right)EZ_{n_1+1}. \tag{4}$$

Seeing the probability to abandon at position $j$, for $1\le j\le n_1$, we obtain

$$\frac{\gamma_1}{s\mu+j\gamma_1}\prod_{l=j+1}^{n_1}\left(1-\frac{\gamma_1}{s\mu+l\gamma_1}\right)=\frac{\gamma_1}{s\mu+n_1\gamma_1}.$$

Averaging over all possibilities, we have

$$EZ_{n_1} = \frac{\gamma_1}{s\mu+n_1\gamma_1}\sum_{j=1}^{n_1}EZ_{n_1}(j).$$

The expected value of $Z_{n_1}$ can be written as

$$EZ_{n_1} = \frac{\gamma_1}{s\mu+n_1\gamma_1}\sum_{j=1}^{n_1}\frac{j\gamma_1}{s\mu+j\gamma_1}.$$

**Low-Priority Customers**

For a new type 2 customer, we only focus on the expected waiting time. In the last few section, we have the approximate of $\pi_n$.

Having $\pi_n$ on hand, we can compute the expected queueing length for class-2 customer. Therefore, we obtain

$$EQ_2 = \sum_{i=1}^{\infty} i\cdot\boldsymbol{\pi}_i\cdot\mathbf{1}^T,$$

where $\mathbf{1}^T$ is the column vector with element 1. Also, because of the class-2 has infinite queue, so we have

$$EW_2 = \frac{EQ_2}{\lambda_2}.$$

From the Eq.(11) in Sarhangian and Balcioglu [14], we have

$$EW_{2,s} = \frac{\int_0^\infty xe^{-\gamma_2 x}f_2(x)dx}{p_{i,s}}. \tag{5}$$

Having the LST of $f_i\left(x\right)$ on hand, Eq. (5) can be compute by taking the derivate with respect to $\gamma_2$.
Also, $EW_{2,a}$ can be obtained by

$$EW_2 = p_{2,s}EW_{2,s}+p_{2,a}EW_{2,a}.$$

### 5.2 Analysis of Model$_{\text{LCFS}}$

**High-Priority Customers**

Approaching to compute the expected waiting time is base on their virtual waiting time. The virtual waiting time is defined as the waiting time given that the customer has infinitely patient.

Let $V_i\left(t\right)$ be the virtual waiting time of a class-$i$ customer at time $t$ with $f_i\left(x\right)$ as its density function. We define the patience times by the random variable $T$.
We focus on the conditional waiting time of class-1 customer given served, and we obtain

$$F_{W_{1,s}}(t) = \frac{P(V_1 < t, V_1 < T)}{P(V_1 < T)}.$$

for $t \ge 0$.

Because of the discipline of service is LCFS, class-1 customer already in the queue is not affect the waiting time. So the conditional virtual waiting time given that the new customer has to wait is independent of the state, and denote it by $V_{1,w}$ We have

$$V_1 = p_w \cdot V_{1,w},$$

Also, we have

$$P(V_1 < T) = p_{1,s},$$

and

$$P(V_1 < t, V_1 < T) = \int_0^t e^{-\gamma_1 x} f_1(x)\,dx.$$

Therefore, a new class-1 customer finds at least one idle server will immediately enter the service with probability $1 - p_w$. Or, he finds all server busy.

Because of the conditional virtual waiting time is independent of the state. Thus we can write

$$P(V_1 < t, V_1 < T) = (1 - p_w) \cdot 1 + p_w \int_0^t e^{-\gamma_1 x} f_{V_{1,w}}(x)\,dx.$$

Taking the derivative with respect to t above, we obtain

$$f_{W_{1,s}}(t) = \frac{p_w}{p_{1,s}} e^{-\gamma_1 t} f_{B_{0,\lambda_1,\gamma_1}}(t),$$

Using the LST we gain

$$\tilde{F}_{W_{1,s}}(x) = \frac{p_w}{p_{1,s}} \tilde{F}_{B_{0,\lambda_1,\gamma_1}}(x + \gamma_1),$$

Applying the 1-st order moment

$$EW_{1,s} = -\frac{p_w}{p_{1,s}} \tilde{F}^{(1)}_{B_{0,\lambda_1,\gamma_1}}(\gamma_1),$$

where $g(k)(\cdot)$ is the $k$ -th derivative of $g(\cdot)$.

Now, we focus on the conditional waiting time given abandonment.
Having that

$$F_{W_{1,a}}(t) = \frac{P(T < t, V_1 > T)}{P(V_1 > T)},$$

We obtain

$$F_{W_{1,a}}(t) = \frac{1}{p_{1,a}} \left\{ 1 - e^{-\gamma_1 t} - \int_0^t \gamma_1 e^{-\gamma_1 x} (1 - p_w + p_w F_{B_{0,\lambda_1,\gamma_1}}(x))dx \right\},$$

Taking derivative with respect to t on both sides

$$f_{W_{1,a}}(t) = \frac{p_w \gamma_1}{p_{1,a}} (e^{-\gamma_1 t} - e^{-\gamma_1 t} F_{B_{0,\lambda_1,\gamma_1}}(t)),$$

Applying the LST, we obtain

$$\tilde{F}_{W_{1,a}}(x) = \frac{p_w \gamma_1}{p_{1,a}(x + \gamma_1)} \left(1 - \tilde{F}_{B_{0,\lambda_1,\gamma_1}}(x + \gamma_1)\right), \tag{6}$$

Finally, taking the $k$ -th order moment of $W_{1,a}$, we obtain the expected conditional waiting time given abandonment.

### Low-Priority Customers

Using Jouini and Roubos [7], we know that the expected waiting time is unchanged for all work-conserving policies. Clearly, because the Markov chain is unchanged, so does the matrix $\mathbf{Q}$. Then we know that $\boldsymbol{\pi}_n$ πn is same as before.

Having πn on hand, we can compute the expected queueing length for class-2 customer. Therefore, we obtain

$$EQ_2 = \sum_{i=1}^{\infty} i \cdot \boldsymbol{\pi}_i \cdot \mathbf{1}^T,$$

where $\mathbf{1}^T$ is the column vector with element 1. Also, because of the class-2 has infinite queue, so we have

$$EW_2 = \frac{EQ_2}{\lambda_2}.$$

## 6. NUMERICAL RESULTS

In this section, we use numerical solutions to prove that our results are reasonable. First, we use the same conditions to state that our solution is the same as the solution given in Eq. (2). Second, we use Eqs.(3), (4), (6) and (7) to gain the useful result.

### 6.1 Comparison the probability of all servers idle

In the Sarhangian and Balcioglu [14], it gives

$$P_0 = \left( \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i} + \frac{\lambda}{\lambda_2} \frac{\lambda^{s-1}}{(s-1)!\mu^{s-1}} \sum_{j=0}^{\infty} \prod_{k=0}^{j} \lambda_2 \tilde{g}_0(k\gamma_2) \right)^{-1},$$

where $\tilde{g}_0(x)$ is the LST of $B_{0,\lambda_1,\gamma_1}$ with the complementary distribution function of an 0-busy period. We make a comparison between this P0 and p0 in Eq. (2).

Assuming $\lambda_1 = 30$, $\lambda_2 = 40$, $s = 5$, $\mu = 15$, $K = 15$, $\epsilon = 10^{-10}$. Table 1 indicates that the difference between $P_0$ and $p_0$ is less than 10−4, thereby demonstrating the accuracy of the proposed method. Moreover, an increase in the rate of abandonment leads to a decrease in the number of customers in the system, which in turn increases the probability of all servers being idle. This is a clear demonstration that these numerical results are reasonable.

Next, we focus on differences in service rates using a graph. It is clear that an increase in the service rate leads to a reduction in the number of customers in the system, which increases the probability of all servers being idle. Figure 2 shows that when the service rate increases, the probability of all servers being idle also increases.

Table 2: Difference between $P_0$ and $p_0$

| $\gamma_1$ | $\gamma_2$ | $P_0$ | $p_0$ | $\gamma_1$ | $\gamma_2$ | $P_0$ | $p_0$ | $\gamma_1$ | $\gamma_2$ | $P_0$ | $p_0$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 0.0075 | 0.0075 | 6 | 5 | 0.0076 | 0.0076 | 7 | 5 | 0.0076 | 0.0076 |
| | 6 | 0.0077 | 0.0077 | | 6 | 0.0078 | 0.0078 | | 6 | 0.0079 | 0.0079 |
| | 7 | 0.0079 | 0.0079 | | 7 | 0.0080 | 0.0080 | | 7 | 0.0081 | 0.0081 |
| | 8 | 0.0081 | 0.0081 | | 8 | 0.0082 | 0.0082 | | 8 | 0.0082 | 0.0082 |
| | 9 | 0.0083 | 0.0083 | | 9 | 0.0083 | 0.0083 | | 9 | 0.0084 | 0.0084 |
| | 10 | 0.0084 | 0.0084 | | 10 | 0.0085 | 0.0085 | | 10 | 0.0085 | 0.0085 |
| | 11 | 0.0085 | 0.0085 | | 11 | 0.0086 | 0.0086 | | 11 | 0.0087 | 0.0087 |
| | 12 | 0.0087 | 0.0087 | | 12 | 0.0087 | 0.0087 | | 12 | 0.0088 | 0.0088 |
| | 13 | 0.0088 | 0.0088 | | 13 | 0.0088 | 0.0088 | | 13 | 0.0089 | 0.0089 |
| | 14 | 0.0089 | 0.0089 | | 14 | 0.0089 | 0.0089 | | 14 | 0.0090 | 0.0090 |
| | 15 | 0.0090 | 0.0090 | | 15 | 0.0090 | 0.0090 | | 15 | 0.0091 | 0.0091 |
| 8 | 5 | 0.0077 | 0.0077 | 10 | 5 | 0.0078 | 0.0078 | 15 | 5 | 0.0081 | 0.0081 |
| | 6 | 0.0079 | 0.0079 | | 6 | 0.0081 | 0.0081 | | 6 | 0.0083 | 0.0083 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 0.0081 | 0.0081 | 7 | 0.0082 | 0.0082 | 7 | 0.0085 | 0.0085 |
| 8 | 0.0083 | 0.0083 | 8 | 0.0084 | 0.0084 | 8 | 0.0086 | 0.0086 |
| 9 | 0.0085 | 0.0085 | 9 | 0.0086 | 0.0086 | 9 | 0.0088 | 0.0088 |
| 10 | 0.0086 | 0.0086 | 10 | 0.0087 | 0.0087 | 10 | 0.0089 | 0.0089 |
| 11 | 0.0087 | 0.0087 | 11 | 0.0088 | 0.0088 | 11 | 0.0090 | 0.0090 |
| 12 | 0.0088 | 0.0088 | 12 | 0.0089 | 0.0089 | 12 | 0.0091 | 0.0091 |
| 13 | 0.0089 | 0.0089 | 13 | 0.0090 | 0.0090 | 13 | 0.0092 | 0.0092 |
| 14 | 0.0090 | 0.0090 | 14 | 0.0091 | 0.0091 | 14 | 0.0093 | 0.0093 |
| 15 | 0.0091 | 0.0091 | 15 | 0.0092 | 0.0092 | 15 | 0.0094 | 0.0094 |

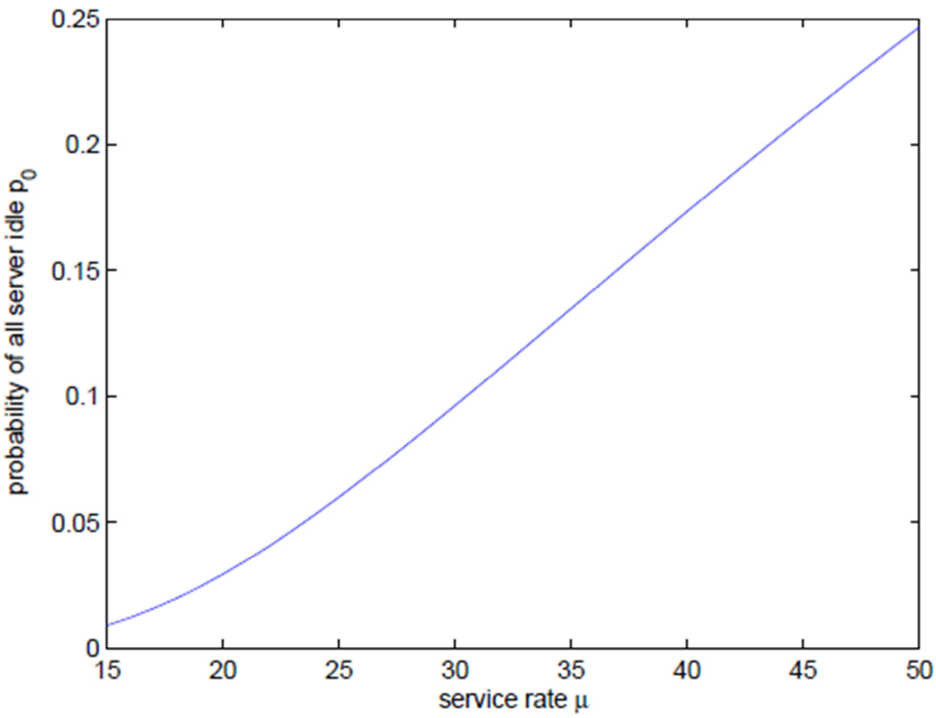


Figure 2: Probability of all server idle with different service rate.

### 6.2 Comparison between FCFS and LCFS

The number of servers s is used as a variable in the numerical tests to determine the expected waiting times associated with two service policies. It depicts the conditional expected waiting times of class-1 given service and abandonment of customers in the same class in Figures 3 - 8. We begin by assigning the same rate of abandonment for both classes of customer, and compare the expected waiting time computed by Jouini and Roubos [7]. The model presented by Jouini and Roubos [7] assumes an infinite buffer for class-1 customers; however, we can still perform a comparison with the proposed model if we permit a sufficiently large buffer size for class-1 customers. Second, we assume that the rate of abandonment differs between the two classes of customers. Comparing with Figures 4 and 6 (borrowed form Jouini and Roubos [7]), we take Eqs.(3), (4), (6) and (7) to compute the expected waiting time in order to verify our model that are shown in Figures 3 and 5. Figure 3 shows that our numerical results are close to those presented by Jouini and Roubos [7]; therefore, it would be reasonable to describe the proposed model and computer code as valid. The error that does exist can be attributed to the difference in the size of the buffer queue for class-1 customers. For the single- class M/M/s+M queue, the conditional expected waiting for those given service in FCFS is greater than that under LCFS policy that was demonstrated by Jouini et al. [8]. Also, the conditional waiting time for those who abandon the queue in LCFS is greater than FCFS. These results are easily extended to our models in Figure 7 and Figure 8, wherein a doubling in the arrival rate affects only the conditional expected waiting time and the property does not change.

We present the Central Processing Unit (CPU) time for the expected waiting time with different s in Figure 9. The computing times under FCFS and LCFS are significantly different. This is due to using Eqs.(3) and (4) for FCFS while using Eqs.(6) and (7) for LCFS. Clearly, because of the matrix dimensions, we waste a lot of time in matrix calculation. Therefore, the future research can be shorten the calculation time.

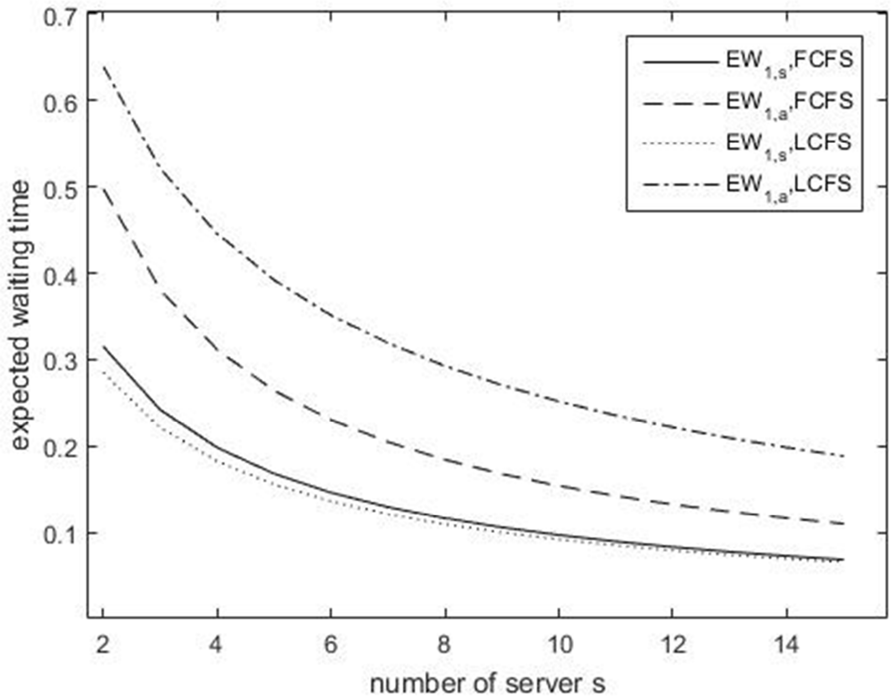


Figure 3: Expected waiting time given service and abandonment. ( $\gamma_1 = \gamma_2 = 0.5, \lambda_1 = \lambda_2 = s/2, \mu = 1$ )

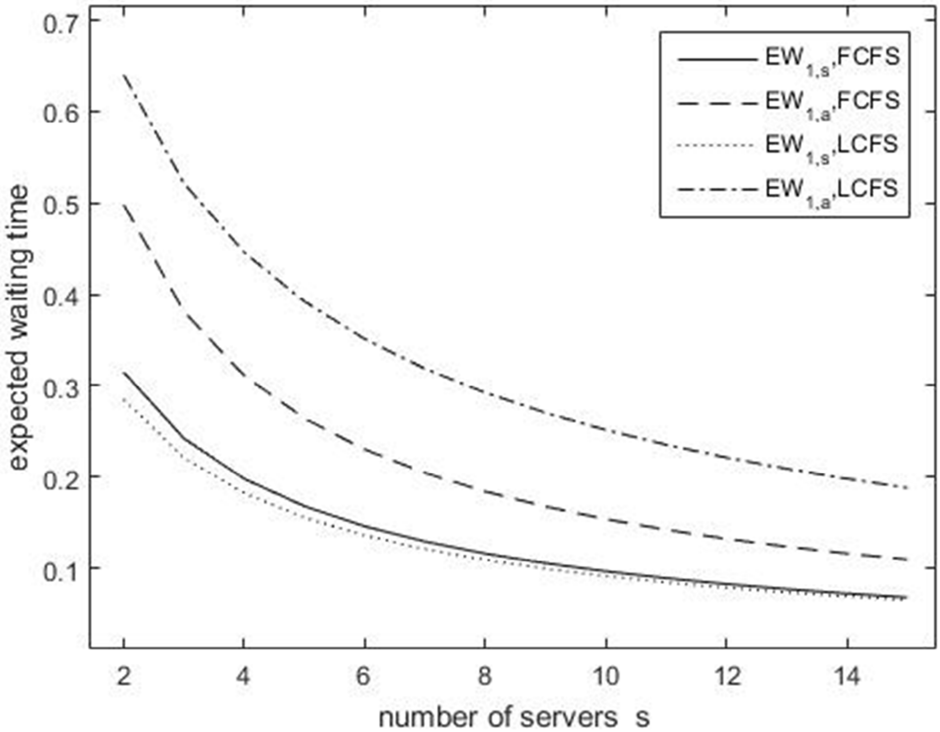


Figure 4: Expected waiting time given service and abandonment in Jouini and Roubos [7].
($\gamma_1 = \gamma_2 = 0.5, \lambda_1 = \lambda_2 = s/2, \mu = 1$)

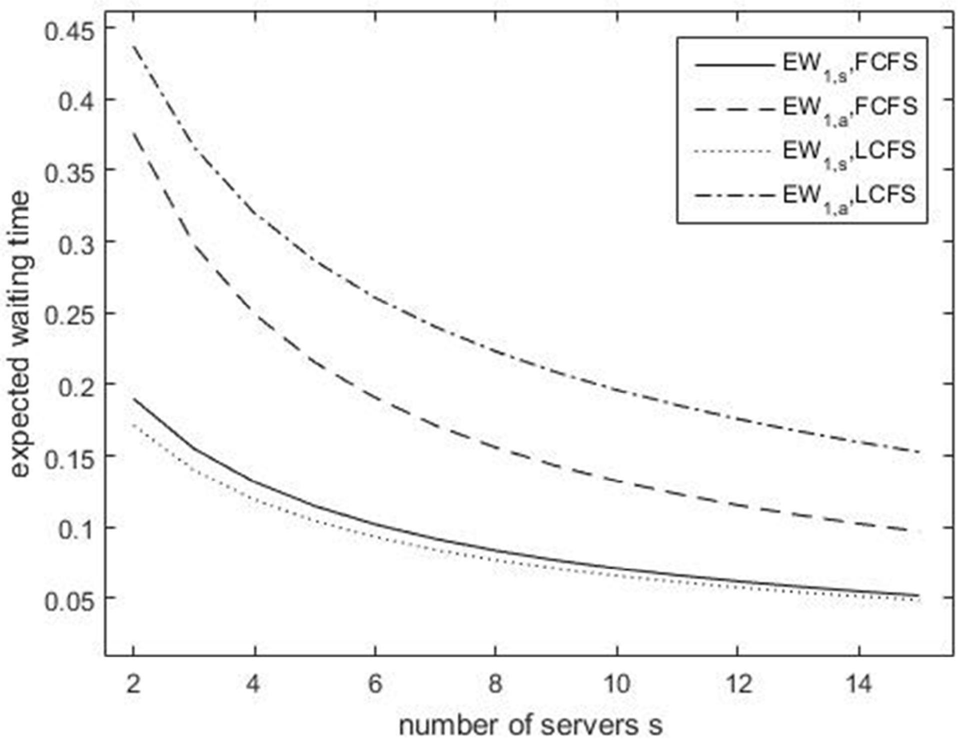


Figure 5: Expected waiting time given service and abandonment. ($\gamma_1 = \gamma_2 = 1, \lambda_1 = \lambda_2 = s/2, \mu = 1$)

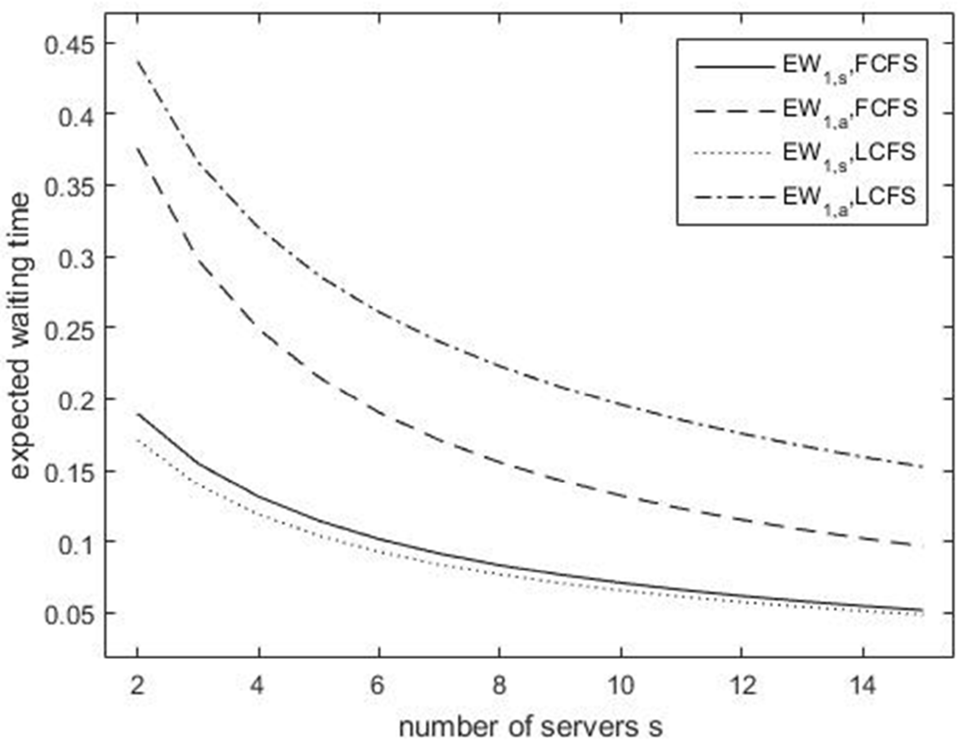


Figure 6: Expected waiting time given service and abandonment in Jouini and Roubos [7].
$\left(\gamma_1 = \gamma_2 = 1, \lambda_1 = \lambda_2 = s/2, \mu = 1\right)$

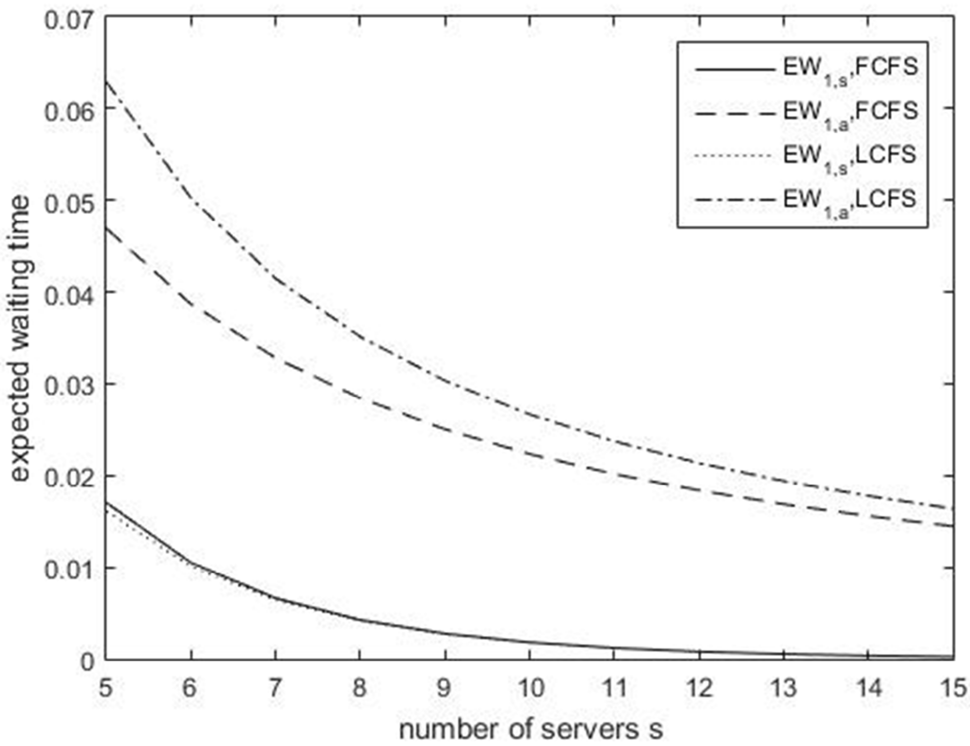


Figure 7: Expected waiting time given service and abandonment. ( $K = 15, \epsilon = 10^{-10}, \mu = 5, \gamma_1 = 3, \gamma_2 = 5, \lambda_1 = 10, \lambda_2 = 2s$ )

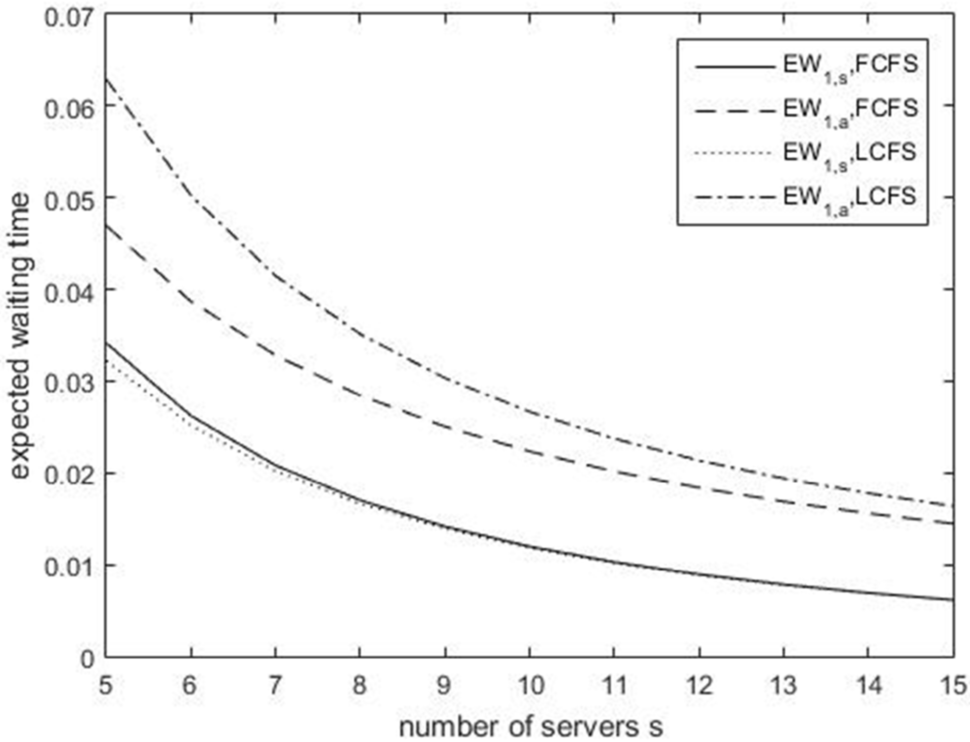


Figure 8: Expected waiting time given service and abandonment. ( $K=15, \epsilon=10^{-10}, \mu=5, \gamma_1=3, \gamma_2=5, \lambda_1=10, \lambda_2=4s$ )

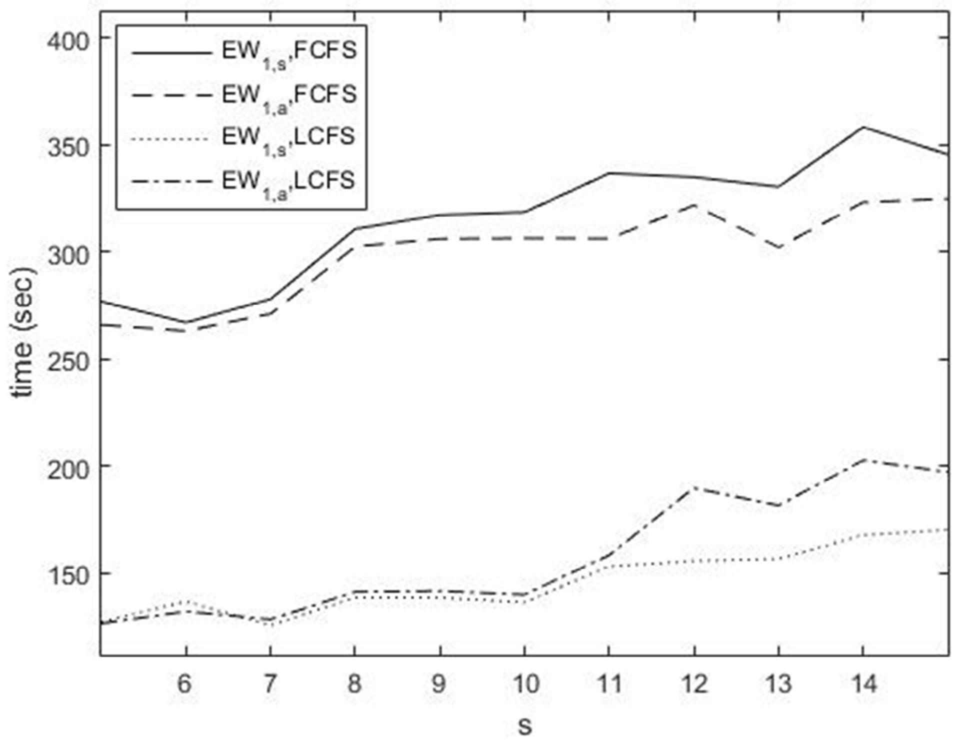


Figure 9: CPU time to compute the conditional expected waiting time in Figure 8.
( $K=15, \epsilon=10^{-10}, \mu=5, \gamma_1=3, \gamma_2=5, \lambda_1=10, \lambda_2=4s$ )

## 7. CONCLUSION

This paper considers a system of multi-server queues with two classes of impatient customer: high-priority and low-priority. Customers join the system according to a Poisson process and customers may abandon service after entering the queue for an exponentially distributed duration with distinct rates. First, we developed a method by which to compute the probability of all servers being idle. Next, we continue to compute the stationary probability of number of customers in both classes in order to derive the expected waiting time for both classes of customers. This paper investigate the expected waiting time of both classes customers with different impatient rates under LCFS and FCFS. We also propose the method to compute the stationary probability of number of customers in both classes in order to derive the expected waiting time for both classes of customers. For high-priority customers, we developed performance measures related to queueing times and conditional waiting times in cases where service is provided and cases of abandonment.

Because it is complicated to analyze the waiting behavior of low-priority customrers which is beyond the scope of the paper and is not discussed in the paper. Future research will be aimed at determining conditional waiting times for low-priority customers in cases where service is provided and in cases of abandonment. Researchers could also examine this model in cases where the two classes of customer differ with regard to service rates.

## REFERENCES

1. F. Baccelli and G. Hebuterne.(1981). On queues with impatient customers. https://hal.archives-ouvertes.fr/docs/00/07/64/67/PDF/RR-0094.pdf
2. A. Brandt and M. Brandt.(1999). On the m (n)/m (n)/s queue with impatient calls. *Performance Evaluation,* 35(1):1–18.
3. B. D. Choi, B. Kim, and J. Chung. (2001). M/m/1 queue with impatient customers of higher priority. *Queueing Systems*, 38(1):49–66.
4. O. Garnett, A. Mandelbaum, and M. Reiman. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4 (3):208–227.
5. F. Iravani and B. Balcıoglu. (2008). On priority queues with impatient customers. *Queueing Systems*, 58(4):239–260.
6. D. L. Jagerman. (2000). Difference equations with applications to queues. *CRC Press.*
7. O. Jouini and A. Roubos. (2013). On multiple priority multi-server queues with impatience. *Journal of the Operational Research Society*, 65(5):616–632.
8. O. Jouini, Z. Aksin, and Y. Dallery. (2011). Call centers with delay information: Models and insights. *Manufacturing & Service Operations Management*, 13 (4):534–548.
9. E. P. Kao and S. D. Wilson. (1999). Analysis of nonpreemptive priority queues with multiple servers and two priority classes. *European Journal of Operational Research*, 118(1):181–193.
10. O. Kella and U. Yechiali. (1985). Waiting times in the non-preemptive priority m/m/c queue. *Stochastic Models*, 1(2):257–262.
11. T. Phung-Duc and K. Kawanishi. (2011). Multiserver retrial queues with after-call work. *Numerical Algebra, Control and Optimization*, 1(4):639–656.
12. T. Phung-Duc, H. Masuyama, S. Kasahara, and Y. (2010). Takahashi. A simple algo- rithm for the rate matrices of level-dependent qbd processes. *In Proceedings of the 5th international conference on queueing theory and network applications*, pages 46–52. ACM.
13. T. Phung-Duc, H. Masuyama, S. Kasahara, and Y. Takahashi. (2013). A matrix continued fraction approach to multiserver retrial queues. *Annals of Operations Research*, 202(1):161–183.
14. V. Sarhangian and B. Balcioglu. (2013). Waiting time analysis of multi-class queues with impatient customers. *Probability in the Engineering and Informational Sciences*, 27(3):333–352.
15. A. Sleptchenko. (2003). Multi-class, multi-server queues with non-preemptive priorities. *Eurandom.*
16. Q. Wang. (2004). Modeling and analysis of high risk patient queues. *European Journal of Operational Research,* 155(2):502–515.
17. S. Zeltyn, Z. Feldman, and S. Wasserkrug. (2009). Waiting and sojourn times in a multi-server queue with mixed priorities. *Queueing Systems*, 61(4):305–328.