



On first-come, first-served queues with three classes of impatient customers

Vinay Kumar¹ · Neelesh S. Upadhye¹

Accepted: 31 December 2021 / Published online: 1 March 2022
© Indian Institute of Technology Madras 2022

Abstract In this article, we study queuing systems with three classes of impatient customers which differ across the classes in their distribution of service times and patience times. The customers are served on a first-come, first-served (FCFS) policy independent of their classes. Such systems are common in customer call centers, which often segment their arrivals into classes of callers whose requests differ in complexity and criticality. First of all, we consider an $M/G/1 + M$ queue and then analyze the $M/M/m + M$ system. Using the virtual waiting time process, we obtain performance measures such as the percentage of customers receiving service in each class, the expected waiting times of customers in each class, and the average number of customers waiting in the queue. We use our characterization to perform a numerical analysis of the $M/M/m + M$ system. Finally, we compare the performance of a system based on numerical solution with the steady-state performance measures of a comparable $M/M/m + M$ system.

Keywords Impatient customer · Virtual waiting time · $M/M/m + M$ queue · Steady state analysis

Mathematics Subject Classification 60K25 · 68M20 · 90B22

1 Introduction

In this article, we consider a queue system with three different classes of impatient customers which are independent in their arrival and service time distributions. As the customers have limited patience, customers may abandon the system whenever their waiting time exceeds the patience time. In a call center, such situations (abandonment without service) are frequent which motivates us to categorize the customers into the three different classes, as it may lead to improved performance of the system. For example, credit card call center service requests can be categorized into credit limit, pin change, and fraudulent activity on a credit card holder's account. Here customer representatives can respond to credit limit inquiries quickly, for pin change they might ask for verification details, effectively it may take a few minutes. But on the other hand, the representative who responds to a fraudulent activity call needs more time as compared to other categories. Customer's requests vary; therefore, call centers train their subset of employees to handle only a certain type of service requests. Based on the customer's request, the automatic call distributor will divert their call to a suitable representative. Depending on the type of service requests, each class of customers is independent of each other with respect to their distribution of patience times and service times. Therefore, a subset of employees may serve a queue that receives arrivals from different classes that vary from each other in their service requirements and their callers' patience levels. Here, note that call centers sometimes may assign tags to their customers like most valued, valued, or ordinary to priorities the customers. However, we have to consider the callers' request on a first-come, first-served (FCFS) basis independent of their class. In this paper, we analyzed the performance of the FCFS queue

✉ Neelesh S. Upadhye
neelesh@iitm.ac.in

¹ Department of Mathematics, Indian Institute of Technology, Madras, Chennai 600036, India

system with three classes of customers that may differ from each other in both their distribution of patience times and their distribution of service times.

Studies for describing the performance of queue systems with a single class of “impatient customers” have included analytical characterizations (Daley [10], Baccelli and Hebuterne [3], Stanford [16]) and approximations to performance (Garnett et al. [11], Zeltyn and Mandelbaum [18], Iravani and Balcioglu [13]). Literature has many studies of two-class systems, where most of the systems have prioritized one customer over the other. Choi et al. [9] study the underlying Markov process of an $M/M/1$ queue with impatient customers of higher priority. Brandt and Brandt [5] extend the approach in [9] for the generally distributed first-class customers in two-class $M/M/1$ system. Iravani and Balcioglu [14] use the level-crossing technique proposed by Brill and Posner [7, 8] to study two $M/G/1$ systems, where they consider a preemptive-resume discipline where customers’ classes have exponentially distributed patience times and then consider a non-preemptive discipline where customers in the first class have exponentially distributed patience times, but customers in the other class have patience. Iravani and Balcioglu [13] obtain the waiting time distributions for each class and the probability that customers in each class will abandon them. Adan et al. [1] design heuristics to determine the staffing levels required to meet target service levels in an overloaded FCFS multiclass system with impatient customers. Van Houdt [12] considers a MAP/PH/1 multiclass queue where customers in each of the classes have a general distribution of patience times. Van Houdt [12] derive a numerical method for analyzing the performance of the system by reducing the joint workload and arrival processes to a fluid queue and expresses the steady-state measures using matrix analytical methods. His method provides an exact characterization of the waiting time distribution and abandonment probability under a discrete distribution of patience times, and approximations of the same performance measures under a continuous distribution of patience times. Sakuma and Takine [15] study the $M/PH/1$ system and assume that customers in each class have the same deterministic patience time.

Our work is mainly focused on the system in which three classes of impatient customers are served on an FCFS basis, independent of their class. Ivo Adan et.al. [2] study a similar system with two classes of impatient customers, they analyze this process to obtain performance measures such as the percentage of customers receiving service in each class, the expected waiting times of customers in each

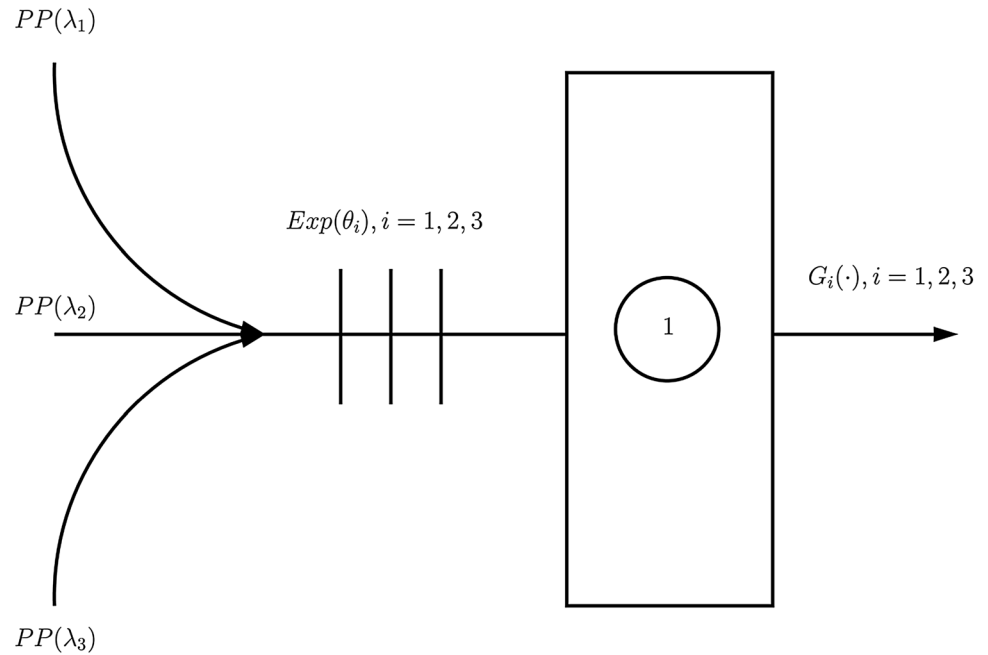
class, and the average number of customers waiting in the queue. We consider two systems ($M/M/1 + M, M/M/m + M$) with three classes of impatient customers, which are served according to an FCFS discipline. To analyze the performance of these systems, we used the virtual waiting time process, see Benes [4], Takács et al. [17] and Ivo Adan et al. [2]. In a virtual waiting time process, the service times of customers who will eventually abandon the system are not considered. By analyzing this process, we find performance characteristics such as the percentage of customers who receive service in each class, the expected waiting times of customers in each class, and the average number of customers waiting in the queue from each class. A related formula for the virtual waiting time in a single class $M/G/1 + PH$ queue is given by Brandt and Brandt [6], it is not suitable for direct computation as it consists of an exponentially growing number of terms. We next perform a numerical analysis of the $M/M/m + M$ system under many arrival rates, mean service times, and mean patience times. Our analysis explains that accounting for differences across classes in the distribution of customers’ service times and patience times is critical, as the performance of our system differs considerably from a system where only the service time distribution varies across classes. The results of our numerical analysis have several administrative implications including service level forecasting, revenue management, and the evaluation of server productivity. Finally, we compare the performance of a system based on numerical results of a comparable $M/M/m + M$ system.

This article is organized as follows: In Sect. 2, we study the $M/G/1 + M$ queue. In Sect. 3, we study the $M/M/m + M$ queue system, including a special case where the three classes have the same mean service time. In Sect. 4, we derive steady-state performance measures. In Sect. 5, we present our numerical analysis, and in Sect. 6, conclusion is given.

2 $M/G/1 + M$ system

Firstly we consider a single-server queueing system with three classes of impatient customers; see Figure 1. Assume that the arrival of Class i ($i = 1, 2, 3$) customers is according to the independent Poisson process with rate λ_i and needs independent and identically distributed (*iid*) service times with CDF $G_i(\cdot)$ and mean τ_i . The customers are impatient, and the patience time distribution for Class

Fig. 1 Single-server model



i ($i = 1, 2, 3$) customers is exponential with parameter θ_i and are independent. That is customers from Class i abandon the system after an exponential amount of time with parameter θ_i if their queuing time is longer than their patience time. For this system, our main aim is to determine performance characteristics such as the long-run fraction of customers entering service, server utilization, and the expected waiting time for the service. Since the customers are impatient in each class, the system will always be stable even if the total arrival rate exceeds the service rate.

It is natural to analyze this system through its queue length process. However, keeping the information of the number of customers from each class is not enough. We would also require the information of the class of each customer at each position in the queue since patience times depend on customer class. This provides the Markov process intractable. Therefore, we used the virtual queueing time process below as in [2].

Let $W(t)$ be the virtual queueing time at time t . $W(t)$ decreases with a rate 1 at all times, provided that it is positive. If an arrival from Class i customers occurs at time t and $W(t) = w$, the arrival leaves the system without service with probability $1 - e^{-\theta_i w}$, or enters for the service with probability $e^{-\theta_i w}$, and require a random amount of service with CDF $G_i(\cdot)$. Hence, the distribution of $S_i(t)$, the size of the upward jump at time t due to an arrival from Class i customers, given $W(t) = w$, is given by

$$P(S_i(t) \leq y | W(t) = w) = 1 - e^{-\theta_i w} + e^{-\theta_i w} G_i(y), \quad (1)$$

and thus

$$E(e^{-sS_i(t)} | W(t) = w) = 1 - e^{-\theta_i w} + e^{-\theta_i w} \tilde{G}_i(s),$$

where $\tilde{G}_i(\cdot)$ is the Laplace-Stieltjes transform (LST) of $G_i(\cdot)$. Let

$$\begin{aligned} \psi(s, t) &= E(e^{-sW(t)}), \\ p_0(t) &= P(W(t) = 0), \\ \phi(s, t) &= \psi(s, t) - p_0(t), \end{aligned}$$

and

$$\begin{aligned} \psi(s) &= \lim_{t \rightarrow \infty} \psi(s, t) = Ee^{-sW}, \\ p_0 &= \lim_{t \rightarrow \infty} p_0(t) = P(W = 0), \\ \phi(s) &= \psi(s) - p_0, \end{aligned}$$

where W is the limit (in distribution) of $W(t)$ as $t \rightarrow \infty$. In the interval $(t, t + h]$, an arrival of type i occurs with probability $\lambda_i h + o(h)$, and no event occurs with probability $1 - (\lambda_1 + \lambda_2 + \lambda_3)h + o(h)$ as $h \rightarrow 0$. Then, we obtain

$$\begin{aligned} \psi(s, t + h) &= (1 - \lambda_1 h - \lambda_2 h - \lambda_3 h)\phi(s, t)e^{sh} \\ &+ (1 - \lambda_1 h - \lambda_2 h - \lambda_3 h)p_0(t) \\ &+ \sum_{i=1}^3 \lambda_i h(\phi(s, t) - \phi(s + \theta_i, t)) \\ &+ \psi(s + \theta_i, t)\tilde{G}_i(s) + o(h). \end{aligned}$$

putting $e^{sh} = 1 + sh + o(h)$ and rearranging terms and dividing by h , and then letting $h \rightarrow 0$, we find

$$\frac{d}{dt}\psi(s, t) = s\phi(s, t) - \sum_{i=1}^3 \lambda_i \psi(s + \theta_i, t)(1 - \tilde{G}_i(s)).$$

Now let $t \rightarrow \infty$. Then, $\frac{d}{dt}\psi(s, t) \rightarrow 0, \psi(s, t) \rightarrow \psi(s)$ and $\phi(s, t) \rightarrow \phi(s)$, so

$$0 = s\phi(s) - \sum_{i=1}^3 \lambda_i \psi(s + \theta_i)(1 - \tilde{G}_i(s)).$$

Dividing by s and using

$$H_i(s) = \lambda_i \frac{1 - \tilde{G}_i(s)}{s},$$

we finally get

$$\psi(s) = p_0 + \sum_{i=1}^3 \psi(s + \theta_i)H_i(s). \tag{2}$$

Note that, the LST of the equilibrium distribution of the service times of customers from Class i is $H_i(s)/(\lambda_i \tau_i)$.

Repeated application of (2) shows that its solution can be written as

$$\psi(s) = p_0 c(s), \tag{3}$$

where

$$c(s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_{i,j,k}(s).$$

The terms $c_{i,j,k}(s)$ satisfy the recurrence relation

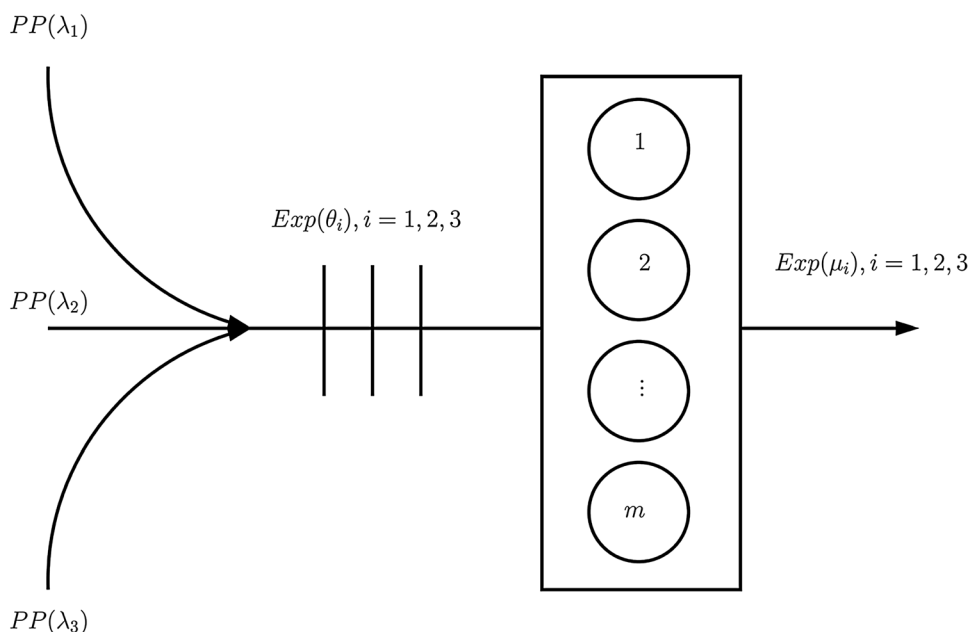
$$\begin{aligned} c_{i,j,k}(s) &= H_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)c_{i-1,j,k}(s) \\ &+ H_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)c_{i,j-1,k}(s) \\ &+ H_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)c_{i,j,k-1}(s), \end{aligned}$$

with $c_{0,0,0}(s) = 1$ and $c_{i,j,k}(s) = 0$ if $i < 0$ or $j < 0$ or $k < 0$. The recursive procedure to obtain equation (3) as well as convergence properties of the series $c(s)$ will be explained in detail in Sect. 3, where we analyze the multi-server system with three classes. Now, to find p_0 , we use $\psi(0) = 1$ in equation (2), we get

$$\begin{aligned} p_0 &= 1 - \sum_{i=1}^3 \psi(\theta_i)\lambda_i \tau_i = 1 - p_0 \sum_{i=1}^3 c(\theta_i)\lambda_i \tau_i, \\ \Rightarrow p_0 &= \left[1 + \sum_{i=1}^3 c(\theta_i)\lambda_i \tau_i \right]^{-1}. \end{aligned}$$

In Sect. 4, we shall see that many performance measures can be computed in terms of $\psi(\theta_i)$.

Fig. 2 m -server model



3 $M/M/m + M$ system

Now we consider a FCFS multi-server system with m servers serving 3 classes of impatient customers; see Figure 2.

3.1 Model

Unlike in Section 2, we assume that customers from Class i arrive according to a Poisson process with rate λ_i , need iid $\text{Exp}(\mu_i)$ service times, and have iid $\text{Exp}(\theta_i)$ patience times ($i = 1, 2, 3$). Whenever service does not begin before the patience time expires, the customer leaves without service.

3.2 Virtual queueing time process

We assume that $W(t)$ be the virtual queueing time at time t in this system. This is the queueing time that would be experienced by a virtual customer arriving at time t . Let $N_i(t)$ be the number of servers serving a Class i customer just after time $t + W(t)$ but before the next customer (if there is one) entering service at time $t + W(t)$. This tells that $N_i(t)$ is the number of servers busy with a Class i customer just before a customer arriving at time t enters for service. So we observe that $N_1(t) + N_2(t) + N_3(t)$ is always at most $m - 1$.

The above definition allows us to determine the size of the upward jump of the virtual queueing time if an arriving

customer at time t decides to join the queue (since his patience exceeds $W(t)$). The jump is the minimum of the service time of the arriving customer and the residual service times of the customers in service at the moment he enters service at time $t + W(t)$. As we will explain below, $\{(W(t), N_1(t), N_2(t), N_3(t)), t \geq 0\}$ is a Markov process with upward jumps, the size of which depends on $W(t)$, and a continuous downward deterministic drift of rate 1 between jumps.

Assume that $W(t) = 0$ and $(N_1(t), N_2(t), N_3(t)) = (i, j, k)$. Then, (i, j, k) is the number of busy servers of classes 1, 2 and 3 at time t . For $0 \leq i + j + k \leq m - 1$, the transition rates of services in state $(0, i, j, k)$ are given by

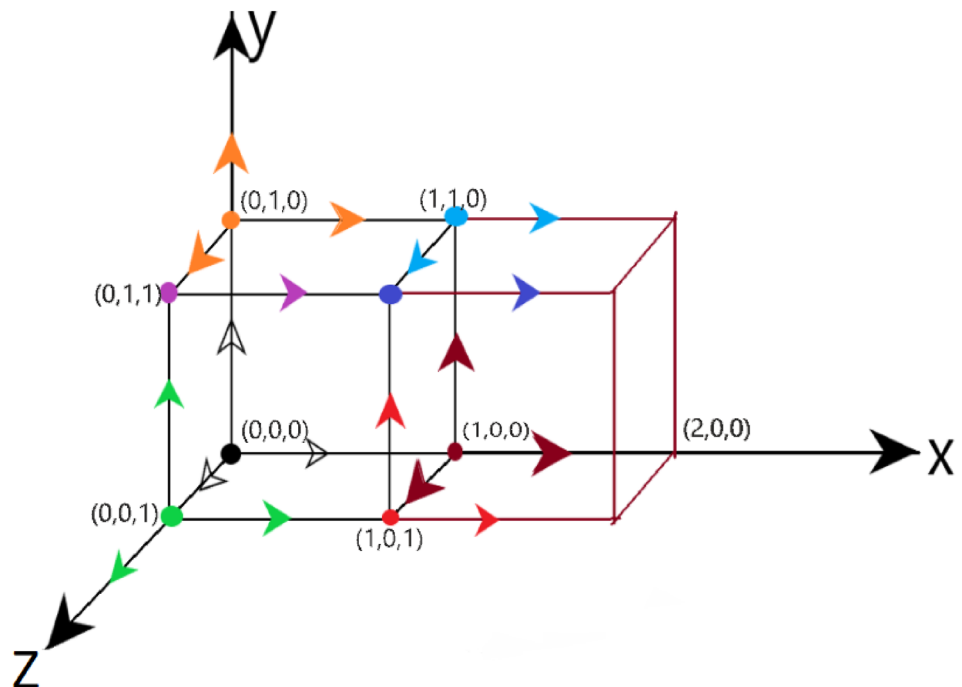
$$\begin{aligned} q_{(0,i,j,k),(0,i-1,j,k)} &= i\mu_1, \\ q_{(0,i,j,k),(0,i,j-1,k)} &= j\mu_2, \\ q_{(0,i,j,k),(0,i,j,k-1)} &= k\mu_3, \end{aligned}$$

and for $0 \leq i + j + k < m - 1$, the transitions rates of arrivals are

$$\begin{aligned} q_{(0,i,j,k),(0,i+1,j,k)} &= \lambda_1, \\ q_{(0,i,j,k),(0,i,j+1,k)} &= \lambda_2, \\ q_{(0,i,j,k),(0,i,j,k+1)} &= \lambda_3. \end{aligned}$$

Now we have all transitions from states $(0, i, j, k)$, $0 \leq i + j + k \leq m - 1$, except for the transition rates of arrivals in states with $i + j + k = m - 1$. These transition rates are explained below

Fig. 3 Path p



Now assume that the state of the quad-variate process at time t is (w, i, j, k) with $w \geq 0$ and $i + j + k = m - 1$. This explains that just after time $t + w$ we will have i busy servers of Class 1, j busy servers of Class 2 and k busy servers of Class 3. Consider an arrival from Class 1 at time t . This customer has to wait w amount of time for service to begin. He abandons the system before his service starts with probability $1 - e^{-\theta_1 w}$, in which case the state does not change, and he enters for service at time $t + w$ with probability $e^{-\theta_1 w}$. Then, the next departure occurs after an $\text{Exp}((i + 1)\mu_1 + j\mu_2 + k\mu_3)$ time X and the departure at time $t + w + X$ is from Class 1 with probability $(i + 1)\mu_1 / ((i + 1)\mu_1 + j\mu_2 + k\mu_3)$, from Class 2 with probability $j\mu_2 / ((i + 1)\mu_1 + j\mu_2 + k\mu_3)$ and from Class 3 with probability $k\mu_3 / ((i + 1)\mu_1 + j\mu_2 + k\mu_3)$. In the first case, the state jumps at time t from (w, i, j, k) to $(w + X, i, j, k)$. In the second and third cases, the state jumps to $(w + X, i + 1, j - 1, k)$ and $(w + X, i + 1, j, k - 1)$, respectively. In the case of a Class 2 and 3 arrival, the process is similar.

Hence, we get the following transition rates from states (w, i, j, k) with $w \geq 0$ and $i + j + k = m - 1$;

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i+1,j-1,k)} \\
 &= \lambda_1 e^{-\theta_1 w} ((i + 1)\mu_1 + j\mu_2 + k\mu_3) \\
 &\quad \times e^{-((i+1)\mu_1 + j\mu_2 + k\mu_3)x} dx \\
 &\quad \times \frac{j\mu_2}{(i + 1)\mu_1 + j\mu_2 + k\mu_3} \\
 &= \lambda_1 e^{-\theta_1 w} j\mu_2 \\
 &\quad \times e^{-((i+1)\mu_1 + j\mu_2 + k\mu_3)x} dx,
 \end{aligned}$$

similarly;

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i,j,k)} \\
 &= \lambda_1 e^{-\theta_1 w} (i + 1)\mu_1 e^{-((i+1)\mu_1 + j\mu_2 + k\mu_3)x} dx \\
 &\quad + \lambda_2 e^{-\theta_2 w} (j + 1)\mu_2 e^{-(i\mu_1 + (j+1)\mu_2 + k\mu_3)x} dx \\
 &\quad + \lambda_3 e^{-\theta_3 w} (k + 1)\mu_3 e^{-(i\mu_1 + j\mu_2 + (k+1)\mu_3)x} dx,
 \end{aligned}$$

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i+1,j,k-1)} \\
 &= \lambda_1 e^{-\theta_1 w} k\mu_3 e^{-((i+1)\mu_1 + j\mu_2 + k\mu_3)x} dx.
 \end{aligned}$$

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i-1,j+1,k)} \\
 &= \lambda_2 e^{-\theta_2 w} i\mu_1 e^{-(i\mu_1 + (j+1)\mu_2 + k\mu_3)x} dx.
 \end{aligned}$$

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i-1,j,k+1)} \\
 &= \lambda_3 e^{-\theta_3 w} i\mu_1 e^{-(i\mu_1 + j\mu_2 + (k+1)\mu_3)x} dx.
 \end{aligned}$$

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i,j+1,k-1)} \\
 &= \lambda_2 e^{-\theta_2 w} k\mu_3 e^{-(i\mu_1 + (j+1)\mu_2 + k\mu_3)x} dx.
 \end{aligned}$$

$$\begin{aligned}
 & Q_{(w,i,j,k),((w+x,w+x+dx),i,j-1,k+1)} \\
 &= \lambda_3 e^{-\theta_3 w} j\mu_2 e^{-(i\mu_1 + j\mu_2 + (k+1)\mu_3)x} dx.
 \end{aligned}$$

Between upward jumps, the virtual waiting time process W decreases continuously and deterministically at a rate 1, while it is positive. When W reaches 0 in state $(0, i, j, k)$, the process will stay in this state until either a service completion or arrival occurs.

3.3 Steady-state analysis

We first begin with the following notations. Let

$$\begin{aligned}
 \psi_{i,j}(s, t) &= E(e^{-sW(t)}; N_1(t) = i, N_2(t) = j, \\
 &\quad N_3(t) = m - 1 - i - j), \quad 0 \leq i + j \leq m - 1, \\
 p_{i,j,k}(t) &= P(W(t) = 0, N_1(t) = i, N_2(t) = j, \\
 &\quad N_3(t) = k), \quad 0 \leq i + j + k \leq m - 1, \\
 \phi_{i,j}(s, t) &= \psi_i(s, t) \\
 &\quad - p_{i,j,m-1-i-j}(t), \quad 0 \leq i + j \leq m - 1,
 \end{aligned}$$

and is the identity matrix where $(W(t), N_1(t), N_2(t), N_3(t))$ converges to (W, N_1, N_2, N_3) in distribution as $t \rightarrow \infty$. We begin with the balance equations for the steady-state probabilities $p_{i,j,k}$. Let

$$p_n = [p_{i,j,k}], \quad 0 \leq i, j, k \leq n, i + j + k = n, \text{ and } 0 \leq n < m.$$

where

$$\begin{aligned}
 p_0 &= [p_{0,0,0}], \\
 p_1 &= [p_{0,0,1} p_{0,1,0} p_{1,0,0}], \\
 p_2 &= [p_{0,0,2} p_{0,1,1} p_{0,2,0} p_{1,0,1} p_{1,1,0} p_{2,0,0}], \\
 p_3 &= [p_{0,0,3} p_{0,1,2} p_{0,2,1} \\
 &\quad p_{0,3,0} p_{1,0,2} \\
 &\quad p_{1,1,1} p_{1,2,0} p_{2,0,1} p_{2,1,0} p_{3,0,0}], \\
 &\vdots \\
 p_n &= [p_{0,0,n} p_{0,1,n-1} \cdots p_{0,n,0} p_{1,0,n-1} p_{1,1,n-2} \cdots \\
 &\quad p_{1,n-1,0} \cdots p_{n-1,0,1} p_{n-1,1,0} p_{n,0,0}]
 \end{aligned}$$

Then, the balance equations can be written in vector-matrix form as

$$\begin{aligned}
 & p_0(\lambda_1 + \lambda_2 + \lambda_3) \\
 &= p_1 M_1, \\
 & p_n(\lambda_1 + \lambda_2 + \lambda_3) \\
 &= p_n \Delta_n \\
 &= p_{n-1} \Lambda_{n-1} + p_{n+1} M_{n+1}, \\
 & 1 \leq n < m - 1,
 \end{aligned} \tag{4}$$

where the $\frac{(n+1)(n+2)}{2} \times \frac{(n+1)(n+2)}{2}$ matrix $\Delta_n = \text{diag}[(n - i - j)\mu_3 + j\mu_2 + i\mu_1]$, $0 \leq i + j \leq n$ (here i, j as in the $\phi(s)$), the matrix of the order $\frac{(n+1)(n+2)}{2} \times$

$\frac{(n+2)(n+3)}{2}$, $\Lambda_n = [\lambda_{n,u,v}]$ and the matrix $M_n = [\mu_{n,i,j,k}]$ of the order $\frac{(n+1)(n+2)}{2} \times \frac{n(n+1)}{2}$ is given by

$$\lambda_{n,u,v} = \begin{cases} \lambda_1, & v = u + m \\ \lambda_2, & v = u + 1, 1 \leq u \leq (m - 1), \\ & v = u + 2, m \leq u \leq (m - 1) + (m - 2), \\ & \vdots \\ & v = u + (m - 1), u = (m - 1) + (m - 2) + \dots + 2 + 1, \\ \lambda_3, & v = u, 1 \leq u \leq (m - 1), \\ & v = u + 1, m \leq u \leq (m - 1) + (m - 2), \\ & \vdots \\ & v = u + (m - 2), u = (m - 1) + (m - 2) + \dots + 2 + 1 \end{cases}$$

$$\begin{aligned} \mu_{n,(i,j,k),(i-1,j,k)} &= i\mu_1, & 1 \leq i \leq n, \\ \mu_{n,(i,j,k),(i,j-1,k)} &= j\mu_2, & 0 \leq j \leq n, \\ \mu_{n,(i,j,k),(i,j,k-1)} &= k\mu_3, & 0 \leq k \leq n. \end{aligned}$$

Further, we can write above balance equations as

$$p_n = p_{n+1}R_{n+1}, \quad 0 \leq n < m - 1. \tag{5}$$

The $\frac{(n+2)(n+1)}{2} \times \frac{(n+1)n}{2}$ matrices R_n , recursively written as

$$\begin{aligned} R_1 &= M_1(\lambda_1 \\ &+ \lambda_2 + \lambda_3)^{-1} \\ R_{n+1} &= M_{n+1}((\lambda_1 + \lambda_2 + \lambda_3)I \\ &+ \Delta_n - R_n\Lambda_{n-1})^{-1}, 1 \leq n < m - 1 \end{aligned}$$

where I is the identity matrix.

Now we shall derive differential equations for the time-dependent LSTs $\psi_{i,j}(s, t)$ and then take $t \rightarrow \infty$ to obtain the steady state equations. For small $h > 0$, we have

$$\begin{aligned} \psi_{i,j}(s, t+h) &= (1 - \lambda_1 h - \lambda_2 h - \lambda_3 h)\phi_{i,j}(s, t)e^{sh} \\ &+ (1 - \lambda_1 h - \lambda_2 h - \lambda_3 h - (i\mu_1 + j\mu_2 \\ &+ (m - 1 - i - j)\mu_3)h)p_{i,j,m-1-i-j}(t) \\ &+ \lambda_1 h(\phi_{i,j}(s, t) - \phi_{i,j}(s + \theta_1, t)) \\ &+ \lambda_2 h(\phi_{i,j}(s, t) - \phi_{i,j}(s + \theta_2, t)) \\ &+ \lambda_3 h(\phi_{i,j}(s, t) - \phi_{i,j}(s + \theta_3, t)) \\ &+ \lambda_1 h\psi_{i,j}(s + \theta_1, t) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1 h\psi_{i,j}(s + \theta_1, t) \frac{j\mu_2}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1 h\psi_{i-1,j}(s + \theta_1, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_2 h\psi_{i,j}(s + \theta_2, t) \frac{i\mu_1}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2 h\psi_{i,j}(s + \theta_2, t) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2 h\psi_{i,j-1}(s + \theta_2, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_3 h\psi_{i+1,j}(s + \theta_3, t) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3 h\psi_{i+1,j}(s + \theta_3, t) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3 h\psi_{i,j}(s + \theta_3, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_1 h p_{i-1,j,m-1-i-j}(t) \\ &+ \lambda_2 h p_{i,j-1,m-1-i-j}(t) \\ &+ \lambda_3 h p_{i,j,m-2-i-j}(t) + o(h), \end{aligned}$$

where $p_{i,j,k}(t) = 0$, if $i < 0$ or $j < 0$ or $k < 0$. Also substituting

$e^{sh} = 1 + sh + o(h)$, rearranging terms, dividing by h and then taking limit as $h \rightarrow 0$, we get

$$\begin{aligned} \frac{d}{dt}\psi_{i,j}(s, t) &= s\phi_{i,j}(s, t) \\ &- \lambda_1\psi_{i,j}(s + \theta_1, t) - \lambda_2\psi_{i,j}(s + \theta_2, t) \\ &- \lambda_3\psi_{i,j}(s + \theta_3, t) \\ &- (i\mu_1 + j\mu_2 + (m - 1 - i - j)\mu_3)h p_{i,j,m-1-i-j}(t) \\ &+ \lambda_1\psi_{i,j}(s + \theta_1, t) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1\psi_{i,j}(s + \theta_1, t) \frac{j\mu_2}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1\psi_{i-1,j}(s + \theta_1, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j}(s + \theta_2, t) \frac{i\mu_1}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j}(s + \theta_2, t) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j-1}(s + \theta_2, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_3\psi_{i+1,j}(s + \theta_3, t) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3\psi_{i+1,j}(s + \theta_3, t) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3\psi_{i,j}(s + \theta_3, t) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_1 p_{i-1,j,m-1-i-j}(t) \\ &+ \lambda_2 p_{i,j-1,m-1-i-j}(t) \\ &+ \lambda_3 p_{i,j,m-2-i-j}(t) \end{aligned}$$

For the steady state, as $t \rightarrow \infty$ the system attains steady state so that $\frac{d}{dt}\psi_{i,j}(s, t) \rightarrow 0$, $\psi_{i,j}(s, t) \rightarrow \psi_{i,j}(s)$, $\phi_{i,j}(s, t) \rightarrow \phi_{i,j}(s)$ and $p_{i,j,k}(t) \rightarrow p_{i,j,k}$. Therefore, we have

$$\begin{aligned} 0 &= s\phi_{i,j}(s) - \lambda_1\psi_{i,j}(s + \theta_1) - \lambda_2\psi_{i,j}(s + \theta_2) - \lambda_3\psi_{i,j}(s + \theta_3) \\ &- (i\mu_1 + j\mu_2 + (m - 1 - i - j)\mu_3)p_{i,j,m-1-i-j} \\ &+ \lambda_1\psi_{i,j}(s + \theta_1) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1\psi_{i,j}(s + \theta_1) \frac{j\mu_2}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_1\psi_{i-1,j}(s + \theta_1) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j}(s + \theta_2) \frac{i\mu_1}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j}(s + \theta_2) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_2\psi_{i,j-1}(s + \theta_2) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_3\psi_{i+1,j}(s + \theta_3) \frac{(i+1)\mu_1}{s + (i+1)\mu_1 + j\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3\psi_{i+1,j}(s + \theta_3) \frac{(j+1)\mu_2}{s + i\mu_1 + (j+1)\mu_2 + (m-1-i-j)\mu_3} \\ &+ \lambda_3\psi_{i,j}(s + \theta_3) \frac{(m-i-j)\mu_3}{s + i\mu_1 + j\mu_2 + (m-i-j)\mu_3} \\ &+ \lambda_1 p_{i-1,j,m-1-i-j} \\ &+ \lambda_2 p_{i,j-1,m-1-i-j} \\ &+ \lambda_3 p_{i,j,m-2-i-j} \end{aligned} \tag{6}$$

The above equations can be written in vector-matrix form. For this, let

$$a_{1,u,v} = \begin{array}{|l|l|} \hline A_{1,u,u} & \text{if } v = u, 1 \leq u \leq \frac{m(m+1)}{2}, \\ \hline A_{1,u,v} & \begin{array}{l} \text{if } v = u + m, 1 \leq u < m, \\ \text{if } v = u + m - 1, m < u < m + (m - 1), \\ \text{if } v = u + m - 2, (m - 1) < u < m + (m - 1) + (m - 2), \\ \vdots \\ \text{if } v = (u + m - (m - 1)), \\ m + (m - 1) + \dots + 3 < u < m + (m - 1) + \dots + 3 + 2 \end{array} \\ \hline \end{array},$$

$$A_{2,u,u} := \lambda_2 \frac{s + (m - 1 - i - j)\mu_3}{s + i\mu_1 + (j + 1)\mu_2 + (m - 1 - i - j)\mu_3}$$

$$A_{2,u,v} := -\lambda_2 \frac{(m - i - j)\mu_3}{s + i\mu_1 + j\mu_2 + (m - i - j)\mu_3}$$

$$a_{2,u,v} = \begin{array}{|l|l|} \hline A_{2,u,u} & \text{if } u = v, 1 \leq u \leq \frac{m(m+1)}{2} \\ \hline A_{2,u,v} & \begin{array}{l} \text{if } v = u + 1, u \neq m, u \neq m + (m - 1), \\ \dots, u \neq m + (m - 1) + (m - 2) + \dots + 3 + 2. \end{array} \\ \hline \end{array},$$

and

$$A_{3,u,u} := \lambda_3 \frac{s + i\mu_1 + j\mu_2}{s + i\mu_1 + j\mu_2 + (m - i - j)\mu_3}$$

$$A_{3,u+1,u} := -\lambda_3 \frac{(j + 1)\mu_2}{s + i\mu_1 + (j + 1)\mu_2 + (m - 1 - i - j)\mu_3}$$

$$A_{3,u,v} := -\lambda_3 \frac{(i + 1)\mu_1}{s + (i + 1)\mu_1 + j\mu_2 + (m - 1 - i - j)\mu_3}$$

$$a_{3,u,v} = \begin{array}{|l|l|} \hline A_{3,u,u} & \text{if } v = u, 1 \leq u \leq \frac{m(m+1)}{2}, \\ \hline A_{3,u+1,u} & \begin{array}{l} \text{if } u = u + 1, u \neq m, u \neq m + (m - 1), \\ \dots, u \neq m + (m - 1) + (m - 2) + \dots + 3 + 2. \end{array} \\ \hline A_{3,u,v} & \begin{array}{l} \text{if } u = v + m, 1 \leq u < m, \\ \text{if } u = v + m - 1, m < u < m + (m - 1), \\ \text{if } u = v + m - 2, (m - 1) < u < m + (m - 1) + (m - 2), \\ \vdots \\ \text{if } u = (v + m - (m - 1)), \\ m + (m - 1) + \dots + 3 < u < m + (m - 1) + \dots + 3 + 2 \end{array} \\ \hline \end{array}$$

$$\begin{aligned} \psi(s) &= [\psi_{0,0}, \psi_{0,1}, \psi_{0,2}, \dots, \\ &\quad \psi_{0,m-1}, \psi_{1,0}, \dots, \psi_{1,m-2}, \dots, \psi_{m-1,0}] \\ \phi(s) &= \psi(s) - \mathbf{p}_{m-1} \\ \phi(s) &= [\phi_{0,0}, \phi_{0,1}, \phi_{0,2}, \dots, \\ &\quad \phi_{0,m-1}, \phi_{1,0}, \dots, \phi_{1,m-2}, \dots, \phi_{m-1,0}] \end{aligned}$$

Here matrix $\psi(s)$ is of the order $1 \times \frac{m(m+1)}{2}$. Then for $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ matrices $A_l(s)$, ($l = 1, 2, 3$) we have

$$\begin{aligned} s\phi(s) &= \mathbf{p}_{m-1}\Lambda_{m-1} - \mathbf{p}_{m-2}\Lambda_{m-2} + \sum_{l=1}^3 \psi(s + \theta_l)A_l(s) \\ s\phi(s) &= \mathbf{p}_{m-1}\Lambda_{m-1} - \mathbf{p}_{m-1}R_{m-1}\Lambda_{m-2} + \sum_{l=1}^3 \psi(s + \theta_l)A_l(s) \end{aligned} \tag{7}$$

where nonzero entries of the matrix $A_l(s) = [a_{l,u,v}]$ for $l = 1, 2, 3$ are given by

$$\begin{aligned} A_{1,u,u} &:= \lambda_1 \frac{s + (m - 1 - i - j)\mu_3}{s + (i + 1)\mu_1 + j\mu_2 + (m - 1 - i - j)\mu_3}, \\ A_{1,u,v} &:= -\lambda_1 \frac{(m - i - j)\mu_3}{s + i\mu_1 + j\mu_2 + (m - i - j)\mu_3}, \end{aligned}$$

that is,

where i and j as in the v th entry of $\psi(s)$ or $\phi(s)$. Consider next

$$\begin{aligned} D(s) &:= I + \frac{\Lambda_{m-1} - R_{m-1}\Lambda_{m-2}}{s}, \\ H_l(s) &:= \frac{A_l(s)}{s}, \quad s > 0, l = 1, 2, 3 \end{aligned} \tag{8}$$

For $s > 0$, we can divide (7) by s , and using (8), we obtain

$$\psi(s) = \mathbf{p}_{m-1}D(s) + \sum_{l=1}^3 \psi(s + \theta_l)H_l(s). \tag{9}$$

Let

$$\begin{aligned} D_{i,j,k}(s) &= D(s + i\theta_1 + j\theta_2 + k\theta_3) \\ \psi_{i,j,k}(s) &= \psi(s + i\theta_1 + j\theta_2 + k\theta_3) \end{aligned}$$

Then (9) gives us for $i, j, k \geq 0$

$$\begin{aligned} \psi_{i,j,k}(s) &= \mathbf{p}_{m-1}D_{i,j,k}(s) + \psi_{i+1,j,k}(s)H_1(s + i\theta_1 + j\theta_2 + k\theta_3) \\ &\quad + \psi_{i,j+1,k}(s)H_2(s + i\theta_1 + j\theta_2 + k\theta_3) \\ &\quad + \psi_{i,j,k+1}(s)H_3(s + i\theta_1 + j\theta_2 + k\theta_3) \end{aligned} \tag{10}$$

To obtain $\psi(s) = \psi_{0,0,0}(s)$, we apply equation (10):

$$\begin{aligned} \psi_{0,0,0}(s) &= \mathbf{p}_{m-1}D_{0,0,0}(s) + \psi_{1,0,0}(s)H_1(s) \\ &\quad + \psi_{0,1,0}(s)H_2(s) + \psi_{0,0,1}(s)H_3(s) \\ &= \mathbf{p}_{m-1}(D_{0,0,0}(s) + D_{1,0,0}(s)H_1(s) \\ &\quad + D_{0,1,0}(s)H_2(s) + D_{0,0,1}(s)H_3(s)) \\ &\quad + \psi_{2,0,0}(s)H_1(s + \theta_1)H_1(s) + \psi_{0,2,0}(s)H_2(s + \theta_2)H_2(s) \\ &\quad + \psi_{0,0,2}(s)H_3(s + \theta_3)H_3(s) \\ &\quad + \psi_{1,1,0}(s)(H_2(s + \theta_1)H_1(s) + H_1(s + \theta_2)H_2(s)) \\ &\quad + \psi_{1,0,1}(s)(H_3(s + \theta_1)H_1(s) + H_1(s + \theta_3)H_3(s)) \\ &\quad + \psi_{0,1,1}(s)(H_3(s + \theta_2)H_2(s) + H_2(s + \theta_3)H_3(s)) \end{aligned}$$

Therefore, after n iterations we get

$$\psi(s) = \mathbf{p}_{m-1} \sum_{i+j+k < n} D_{i,j,k}(s)C_{i,j,k}(s) + \sum_{i+j+k=n} \psi_{i,j,k}(s)C_{i,j,k}(s) \tag{11}$$

The $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ matrices $C_{i,j,k}(s)$ are defined as follows: A sequence of grid points $\mathbf{p} = \{(i_0, j_0, k_0), (i_1, j_1, k_1), \dots, (i_n, j_n, k_n)\}$ is called path from (i_0, j_0, k_0) to (i_n, j_n, k_n) if each of steps $(i_{l+1}, j_{l+1}, k_{l+1}) - (i_l, j_l, k_l)$, $l = 0, 1, \dots, (n - 1)$, is either $(1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$. For the path \mathbf{p}

$$\begin{aligned} C_p(s) &= H_{(i_{n-1}, j_{n-1}, k_{n-1})}(s) \cdots H_{(i_1, j_1, k_1)}(s)H_{(i_0, j_0, k_0)}(s) \\ H_{(i,j,k)} &= \begin{cases} H_1(s + i\theta_1 + j\theta_2 + k\theta_3), & \text{if } (i_{l+1}, j_{l+1}, k_{l+1}) - (i_l, j_l, k_l) = (1, 0, 0) \\ H_2(s + i\theta_1 + j\theta_2 + k\theta_3), & \text{if } (i_{l+1}, j_{l+1}, k_{l+1}) - (i_l, j_l, k_l) = (0, 1, 0) \\ H_3(s + i\theta_1 + j\theta_2 + k\theta_3), & \text{if } (i_{l+1}, j_{l+1}, k_{l+1}) - (i_l, j_l, k_l) = (0, 0, 1) \end{cases} \end{aligned}$$

For the path $\mathbf{p} = \{i_0, j_0, k_0\}$, set $C_p(s) = 1$. Let $\mathcal{P}(i, j, k)$ be the set of all paths from $(0, 0, 0)$ to (i, j, k) . Then $C_{i,j}(s)$ is defined by

$$C_{i,j,k}(s) = \sum_{p \in \mathcal{P}(i,j,k)} C_p(s)$$

For $i + j + k > 0$, the $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ matrices $C_{i,j,k}(s)$ can be recursively calculated from

$$\begin{aligned} C_{i,j,k}(s) &= \sum_{p \in \mathcal{P}(i-1,j,k)} H_{i-1,j,k}C_p(s) \\ &\quad + \sum_{p \in \mathcal{P}(i,j-1,k)} H_{i,j-1,k}C_p(s) \\ &\quad + \sum_{p \in \mathcal{P}(i,j,k-1)} H_{i,j,k-1}C_p(s) \\ &= H_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)C_{i-1,j,k}(s) \\ &\quad + H_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)C_{i,j-1,k}(s) \\ &\quad + H_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)C_{i,j,k-1}(s) \end{aligned} \tag{12}$$

where $C_{0,0,0}(s) = I$ and $C_{i,j,k}(s)$ is all zero matrix if $i < 0$ or $j < 0$ or $k < 0$.

Lemma 3.1 For each $\delta > 0$, the series $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} C_{i,j,k}(s)$ is absolutely and uniformly convergent for all $s > \delta$.

Proof Fix $\delta > 0$. It suffices to prove that there are constants M and $r < 1$ such that for all $s > \delta$ and $i + j + k \geq 0$,

$$|C_{i,j,k}(s)| \leq Mr^{i+j+k}E, \tag{13}$$

where E is the all-one matrix and the inequality is component-wise. In this bound $s \neq 0$, since $H_1(s)$ and $H_2(s)$, thus by (12) also $C_{i,j,k}(s)$, are unbounded as s approaches to 0. Now fix $r < 1$. Since $H_l(s) \leq \frac{2E}{s}$ ($l = 1, 2, 3$), there exists an $N \geq 0$ such that, for $s > \delta$ and $i + j + k \geq N$,

$$|H_l(s + i\theta_1 + j\theta_2 + k\theta_3)| \leq \frac{2r}{3(m)(m+1)}E, \quad l = 1, 2, 3 \tag{14}$$

Recurrence relation (12) implies that, for each $i, j, k \geq 0$, $C_{i,j,k}(s)$ is bounded for $s > \delta$. Hence, there is a (sufficiently large) M such that (13) is valid for $s > \delta$ and the finitely many $i + j + k \leq N$. By induction we now prove that (13) is valid for all $i + j + k \geq N$. Suppose it holds for all $i + j + k = n$ (also true for $n = N$). From (12) and (14), we get, for $s > \delta$ and $i + j + k = n + 1$,

$$\begin{aligned} |C_{i,j,k}(s)| &\leq |H_1(s + (i-1)\theta_1 + j\theta_2 + k\theta_3)| |C_{i-1,j,k}(s)| \\ &\quad + |H_2(s + i\theta_1 + (j-1)\theta_2 + k\theta_3)| |C_{i,j-1,k}(s)| \\ &\quad + |H_3(s + i\theta_1 + j\theta_2 + (k-1)\theta_3)| |C_{i,j,k-1}(s)| \\ &\leq \frac{2r}{3m(m+1)}E(|C_{i-1,j,k}(s)| + |C_{i,j-1,k}(s)| + |C_{i,j,k-1}(s)|) \\ &\leq \frac{2r}{3m(m+1)}3Mr^nE^2 = Mr^{n+1}E, \end{aligned}$$

where the last inequality follows from the induction hypothesis. \square

Since $D_{i,j,k}(s)$ are uniformly bounded for every $s > \delta > 0$ and $i + j + k \geq 0$, we get the following.

Corollary 3.1.1 The series $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} D_{i,j,k}(s)C_{i,j,k}(s)$ is absolutely and uniformly convergent for all $s > \delta > 0$.

Proof Using that $|\psi_{i,j,k}(s)| \leq 1$, the second term in (11) is bounded and given by

$$\left| \sum_{i+k+j=n} \psi_{i,j,k}(s)C_{i,j,k}(s) \right| \leq \sum_{i+k+j=n} |C_{i,j,k}(s)|$$

Therefore it vanishes as $n \rightarrow \infty$ by virtue of the absolute convergence of the series of $C_{i,j,k}(s)$. Hence, taking as $n \rightarrow \infty$ in (11), we get, from Corollary 3.1.1,

$$\psi(s) = \mathbf{p}_{m-1}\mathbf{C}(s), \tag{15}$$

where

$$\mathbf{C}(s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} D_{i,j,k}(s)C_{i,j,k}(s). \tag{16}$$

in particular, we have

$$\psi(\theta_l) = \mathbf{p}_{m-1}\mathbf{C}(\theta_l), \quad l = 1, 2, 3. \tag{17}$$

To complete the LST of the virtual queueing time, we need to evaluate \mathbf{p}_{m-1} . For this, first we set $s = 0$ in (7), and we obtain

$$\begin{aligned} 0 &= \mathbf{p}_{m-1}\Delta_{m-1} - \mathbf{p}_{m-1}\mathbf{R}_{m-1}\Lambda_{m-2} + \sum_{l=1}^3 \psi(\theta_l)A_l(0) \\ &= \mathbf{p}_{m-1}\Delta_{m-1} - \mathbf{p}_{m-1}\mathbf{R}_{m-1}\Lambda_{m-2} + \mathbf{p}_{m-1} \sum_{l=1}^3 \mathbf{C}(\theta_l)A_l(0) \end{aligned} \tag{18}$$

where the second equality follows from (17). To uniquely determine \mathbf{p}_{m-1} , we finally need the normalizing equation

$$\sum_{n=0}^{m-1} \mathbf{p}_n \mathbf{e} + \phi(0)\mathbf{e} = 1, \tag{19}$$

where \mathbf{e} is the vector of all-one and \mathbf{p}_n is given by (5) for $0 \leq n < m - 1$. However, equation (19) requires the computation of $\phi(0)$, which is the complicated step. Taking the derivatives on both sides of (7) and setting $s = 0$, we get

$$\phi(0) = \sum_{l=1}^3 (\psi(\theta_l)A'_l(0) + \psi'(\theta_l)A_l(0)). \tag{20}$$

Here, prime indicates derivative with respect to s . Thus, to calculate $\phi(0)$ we need $\psi'(s)$ at $s = \theta_1, s = \theta_2$ and $s = \theta_3$. For this, we can use (15,16). After differentiating (15), we obtain

$$\begin{aligned} \psi'(s) &= \mathbf{p}_{m-1}\mathbf{C}'(s) = \mathbf{p}_{m-1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \\ &\quad \left(D'_{i,j,k}(s)C_{i,j,k}(s) + D_{i,j,k}(s)C'_{i,j,k}(s) \right). \end{aligned} \tag{21}$$

The terms $C_{i,j,k}(s)$ can be recursively computed by taking the derivative of (12):

$$\begin{aligned}
 C'_{i,j,k}(s) = & H'_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)C_{i-1,j,k}(s) \\
 & + H_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)C'_{i-1,j,k}(s) \\
 & + H'_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)C_{i,j-1,k}(s) \\
 & + H_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)C'_{i,j-1,k}(s) \\
 & + H'_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)C_{i,j,k-1}(s) \\
 & + H_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)C'_{i,j,k-1}(s),
 \end{aligned}$$

where $C'_{i,j,k}(s)$ is all zero matrix if $i = j = k = 0$ or if $i < 0$ or $j < 0$ or $k < 0$. Term-by-term differentiation of (16) is justified by the following two lemmas. \square

Lemma 3.2 For each $\delta > 0$, the series $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} C'_{i,j,k}(s)$ is absolutely and uniformly convergent for all $s > \delta$.

Proof The proof is similar to the proof of Lemma 3.1. It is sufficient to show that there are constants M and $r < 1$ such that, for all $s > \delta$ and $i + j + k \geq 0$,

$$|C'_{i,j,k}(s)| \leq Mr^{i+j+k}E, \tag{22}$$

Now fix $r < 1$. Since $H_l(s) \leq \lambda_l E/s$ ($l = 1, 2, 3$), and $H'_l(s) \leq \lambda_l E/s^2$ ($l = 1, 2, 3$) there is an $N \geq 0$ such that, for $s > \delta$ and $i + j + k \geq N$,

$$|H_l(s + i\theta_1 + j\theta_2 + k\theta_3)| \leq \frac{2r}{6(m)(m+1)}E, \quad l = 1, 2, 3 \tag{23}$$

and

$$|H'_l(s + i\theta_1 + j\theta_2 + k\theta_3)| \leq \frac{2r}{6(m)(m+1)}E, \quad l = 1, 2, 3 \tag{24}$$

Recursions (12) and (22) imply that, for each $i, j, k \geq 0$ $C_{i,j,k}(s)$ and $C'_{i,j,k}(s)$ are bounded for $s > \delta$. Hence, there is a (sufficiently large) M such that both (13) and (22) are valid for $s > \delta$ and the finitely many $i + j + k \leq N$. Following the induction steps in the proof of Lemma 3.1, it follows that (13) is valid for all $i + j + k \geq N$. We now show that (22) also holds for all $i + j + k \geq N$. Suppose that (22) holds for all $i + j + k = N$ (also holds for $n = N$). From (22), (23) and (24), we get, for $i + j + k = n + 1$,

$$\begin{aligned}
 |C_{i,j,k}(s)| \leq & |H'_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)||C_{i-1,j,k}(s)| \\
 & + |H_1(s + (i - 1)\theta_1 + j\theta_2 + k\theta_3)||C'_{i-1,j,k}(s)| \\
 & + |H'_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)||C_{i,j-1,k}(s)| \\
 & + |H_2(s + i\theta_1 + (j - 1)\theta_2 + k\theta_3)||C'_{i,j-1,k}(s)| \\
 & + |H'_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)||C_{i,j,k-1}(s)| \\
 & + |H_3(s + i\theta_1 + j\theta_2 + (k - 1)\theta_3)||C'_{i,j,k-1}(s)| \\
 \leq & \frac{2r}{6m(m+1)}E(|C_{i-1,j,k}(s)| + |C'_{i-1,j,k}(s)| \\
 & + |C_{i,j-1,k}(s)| \\
 & + |C'_{i,j-1,k}(s)| + |C_{i,j,k-1}(s)| + |C'_{i,j,k-1}(s)|) \\
 \leq & \frac{2r}{6m(m+1)}6Mr^nE^2 = Mr^{n+1}E,
 \end{aligned}$$

where the last inequality follows from the mathematical induction hypothesis. \square

Lemma 3.3 For each $s > 0$, the derivative of $C(s)$ exists and is equal to

$$C'(s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (D'_{i,j,k}(s)C_{i,j,k}(s) + D_{i,j,k}(s)C'_{i,j,k}(s)).$$

Proof Fix $s > 0$. First note that the series converges by Lemmas 3.1-3.2 and the fact that $D_{i,j,k}$ and $D'_{i,j,k}$ are uniformly bounded for all $i + j + k \geq 0$. It suffices to prove that, for each sequence $\{h_n\}$ converging to 0 such that $s + h_n > 0$ for all n ,

$$\lim_{n \rightarrow \infty} \frac{C(s + h_n) - C(s)}{h_n} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} B'_{i,j,k}(s)$$

where $B_{i,j,k}(s) = D_{i,j,k}(s)C_{i,j,k}$. Let h_n be a sequence. Thus, there is a $\delta > 0$ such that $s + h_n > \delta$ for all n . According to (13) and (22) and that the fact that $D_{i,j,k}(s + h_n)$ and $D'_{i,j,k}(s + h_n)$ are uniformly bounded for all n and $i + j + k \geq 0$, there are constants M and $r < 1$ such that, for all n and $i + j + k \geq 0$,

$$|B'_{i,j,k}(s)| \leq Mr^{i+j+k}E. \tag{25}$$

We need to show that for each $\epsilon > 0$ there is an N such that, for $n > N$,

$$\left| \frac{C(s+h_n) - C(s)}{h_n} - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} B'_{i,j,k}(s) \right| = \left| \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left(\frac{B_{i,j,k}(s+h_n) - B_{i,j,k}(s)}{h_n} - B'_{i,j,k}(s) \right) \right| < \epsilon. \tag{26}$$

Let $\epsilon > 0$. For given n and $i, j, k \geq 0$, it follows from the mean value theorem that there is an $0 < \eta < 1$ such that $(B_{i,j,k}(s+h_n) - B_{i,j,k}(s))/h_n = B'_{i,j,k}(s + \eta h_n)$. Hence by (25)

$$\left| \frac{B_{i,j,k}(s+h_n) - B_{i,j,k}(s)}{h_n} - B'_{i,j,k}(s) \right| \leq |B'_{i,j,k}(s + \eta h_n)| + |B'_{i,j,k}(s)| \leq 2Mr^{i+j+k}E.$$

Note that M and r do not depend on n, i, j, k . So there is a constant K such that, for all n ,

$$\sum_{i+j+k \geq K} \left| \frac{B_{i,j,k}(s+h_n) - B_{i,j,k}(s)}{h_n} - B'_{i,j,k}(s) \right| < \frac{\epsilon}{2}. \tag{27}$$

Further, for given $i, j, k \geq 0$, $(B_{i,j,k}(s+h_n) - B_{i,j,k}(s))/h_n$ converges to $B'_{i,j,k}(s)$ as n tends to infinity. Hence, there is an N such that, for $n > N$

$$\sum_{0 \leq i+j+k < K} \left| \frac{B_{i,j,k}(s+h_n) - B_{i,j,k}(s)}{h_n} - B'_{i,j,k}(s) \right| < \frac{\epsilon}{2}. \tag{28}$$

Combining (27) and (28) yields (26).

Substituting (17) and (21) with $s = \theta_l$ in (20) gives

$$\phi(0) = p_{m-1} \sum_{l=1}^3 (C(\theta_l)A'_l(0) + C'(\theta_l)A_l(0))$$

and the normalization equation (19) can be rewritten as

$$\sum_{n=0}^{m-1} p_n e + p_{m-1} \sum_{l=1}^3 (C(\theta_l)A'_l(0) + C'(\theta_l)A_l(0)) e = 1. \tag{29}$$

□

The following theorem summarizes the above findings.

Theorem 3.4 The steady-state LST $\psi(s)$ of the virtual queuing time satisfies

$$\psi(s) = p_{m-1} C(s)$$

where $C(s)$ is defined by (16) and the probability vectors p_n for $0 \leq n \leq m - 1$ are the solution to the system of linear equations (5), (18) and (29).

3.4 Special case $\mu_1 = \mu_2 = \mu_3 = \mu$

We now suppose $\mu_1 = \mu_2 = \mu_3 = \mu$. Now we are dealing with easy problem, since we do not need to keep track of $N_1(t), N_2(t)$ and $N_3(t)$ separately, only $N(t) = N_1(t) + N_2(t) + N_3(t)$. Define

$$\psi(s) = \lim_{t \rightarrow \infty} E(s^{-sW(t)}; N(t) = m - 1)$$

$$= E(s^{-sW}; N = m - 1),$$

$$p_i = \lim_{t \rightarrow \infty} P(W(t) = 0, N(t) = i)$$

$$= P(W = 0, N = i), \quad 0 \leq i \leq m - 1,$$

$$\phi(s) = \psi(s) - p_{m-1}.$$

Then the balance equation (4) can be written as

$$p_{n-1}(\lambda_1 + \lambda_2 + \lambda_3) = p_n n \mu, \quad 1 \leq n \leq m - 1, \tag{30}$$

and (9) reduces to

$$\psi(s) = p_{m-1} + \psi(s + \theta_1) \frac{\lambda_1}{s + m\mu} + \psi(s + \theta_2) \frac{\lambda_2}{s + m\mu} + \psi(s + \theta_3) \frac{\lambda_3}{s + m\mu}.$$

The solution of this equation is given by (15)

$$\psi(s) = p_{m-1} c(s),$$

where

$$c(s) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_{i,j,k}(s)$$

For $i + j + k > 0$, the terms $c_{i,j,k}(s)$ are determined from recursion

$$c_{i,j,k}(s) = \frac{\lambda_1}{s + (i-1)\theta_1 + j\theta_2 + k\theta_3 + m\mu} c_{i-1,j,k}(s) + \frac{\lambda_2}{s + i\theta_1 + (j-1)\theta_2 + k\theta_3 + \mu} c_{i,j-1,k}(s) + \frac{\lambda_3}{s + i\theta_1 + j\theta_2 + (k-1)\theta_3 + \mu} c_{i,j,k-1}(s),$$

with $c_{0,0,0} = 1$ and $c_{i,j,k} = 0$, if $i < 0$ or $j < 0$ or $k < 0$. The normalization equation becomes

$$1 = \sum_{n=0}^{m-1} p_n + \phi(0) = \sum_{n=0}^{m-1} p_n + \psi(\theta_1) \frac{\lambda_1}{k\mu} + \psi(\theta_2) \frac{\lambda_2}{k\mu} + \psi(\theta_3) \frac{\lambda_3}{k\mu}.$$

Together with (30), this yields, for $n = 0, 1, \dots, m - 1$,

$$p_n = \left(1 - \psi(\theta_1) \frac{\lambda_1}{k\mu} - \psi(\theta_2) \frac{\lambda_2}{k\mu} - \psi(\theta_3) \frac{\lambda_3}{k\mu} \right) \frac{\rho^n}{\sum_{j=0}^{m-1} \frac{\rho^j}{j!}}$$

where $\rho = \frac{\lambda_1 + \lambda_2 + \lambda_3}{\mu}$.

4 Performance measures

Now we show how many useful performance measures in steady-state can be computed in terms of the LST evaluated at θ_1, θ_2 and θ_3 . Suppose the $M/M/m + M$ system is in a steady state. An arrival faces a queuing time of W . If the arrival is from Class i , the customer will enter service if his/her impatience time T_i is longer than W . For Class 1, this probability is given by the following (and similarly for Class 2 and 3, by replacing θ_1 by θ_2 and θ_3 , respectively):

$$\begin{aligned}
 P(T_1 > W) &= E(e^{-\theta_1 W}) \\
 &= \sum_{i+j+k \leq m-1} E(e^{-\theta_1 W}; N_1 = i, N_2 = j, N_3 = k) \\
 &= \sum_{n < m-1} p_n e^{-\theta_1 W} \\
 &+ \psi(\theta_1) e^{-\theta_1 W}.
 \end{aligned}$$

Here, we are using that $W=0$ and $E(e^{\theta W}; N_1 = i, N_2 = j, N_3 = k) = p_{ijk}$ when $i + j + k < m - 1$. Next, using Little’s law, we see that the expected number of servers busy serving Class 1 customers is given by

$$\lambda_1 P(T_1 > W) \frac{1}{\mu_1} = \frac{\lambda_1}{\mu_1} \left(\sum_{n < m-1} p_n e^{-\theta_1 W} + \psi(\theta_1) e^{-\theta_1 W} \right)$$

and the steady state throughput equals

$$\sum_{i=1}^3 \lambda_i P(T_i > W) = \sum_{i=1}^3 \lambda_i \left(\sum_{n < m-1} p_n e^{-\theta_i W} + \psi(\theta_i) e^{-\theta_i W} \right).$$

Now we compute the expected time of a Class 1 customer waiting for service. This is given by

$$\begin{aligned}
 E(\min(W, T_1)) &= E(E(\min(W, T_1)|W)) = E\left(\frac{1 - e^{-\theta_1 W}}{\theta_1}\right) \\
 &= \frac{1 - E(e^{-\theta_1 W})}{\theta_1} = \frac{1 - P(T_1 > W)}{\theta_1}.
 \end{aligned}$$

By Little’s law, we get the following for the expected number of Class i customers waiting for service

$$E(L_i^q) = \lambda_i E(\min(W, T_i)) = \frac{\lambda_i}{\theta_i} (1 - P(T_i > W)).$$

At last, we compute the expected conditional waiting time of Class i customers entering service follows from

$$E(W|T_i > W) = \frac{E(W; T_i > W)}{P(T_i > W)},$$

where

$$E(W; T_i > W) = E(W e^{-\theta_i W}) = -\frac{d}{ds} E(e^{-sW}) \Big|_{s=\theta_i} = -\psi'(\theta_i) e^{-\theta_i W}.$$

These formulae simplify significantly when applied to the

$M/G/1 + M$ system. In particular, the probability that the server is busy serving a Class i customer is given by

$$\rho_i = \lambda_i \tau_i \psi(\theta_i),$$

where $\psi(s)$ is as defined in Eq. (2). The probability that the server is busy is given by

$$\rho = \psi(\theta_1) \lambda_1 \tau_1 + \psi(\theta_2) \lambda_2 \tau_2 + \psi(\theta_3) \lambda_3 \tau_3.$$

In steady state, the throughput is equal to

$$\psi(\theta_1) \lambda_1 + \psi(\theta_2) \lambda_2 + \psi(\theta_3) \lambda_3.$$

and the renegeing rate by

$$(1 - \psi(\theta_1)) \lambda_1 + (1 - \psi(\theta_2)) \lambda_2 + (1 - \psi(\theta_3)) \lambda_3.$$

The expected number of Class i , ($i = 1, 2, 3$) customers waiting for service in a steady state is given by

$$E(L_i^q) = \frac{\lambda_i}{\theta_i} (1 - \psi(\theta_i)).$$

The expected number of Class i , ($i = 1, 2, 3$) customers in the system in steady state is given by

$$E(L_i) = E(L_i^q) + \lambda_i \tau_i \psi(\theta_i).$$

This implies that, in the special case when $\tau_i = \frac{1}{\theta_i}$

$$E(L_i) = \frac{\lambda_i}{\theta_i}.$$

This is expected since in this case, the system behaves like an infinite server queue for each class of customers.

5 Numerical analysis

For numerical analysis we have considered three different systems based on their patience time distribution as follows:

System 1: In this system, we considered the distribution of patience time across the classes are same that is $\frac{2}{3}$ unit ($\theta_1 = \theta_2 = \theta_3 = 1.5$).

System 2: In this system, we considered the distribution of patience time are different across the classes. For Class 1 patience time is 1 unit ($\theta_1 = 1$), Class 2 patience time is $\frac{2}{3}$ unit $\theta_2 = 1.5$ and for Class 3 it is $\frac{1}{2}$ unit $\theta_3 = 2$.

System 3: In system 3, we considered that patience time distribution for Class 1 and Class 2 customers are same but differs from Class 3. Here taking ($\theta_1 = 1.5, \theta_2 = 1.5, \theta_3 = 2$).

In all of the systems, we consider requests from Class 1 have a mean service time of 1 unit $\mu_1 = 1$, requests from Class 2 have a mean service time of $\frac{3}{2}$ units ($\mu_2 = 1.5$), and requests from Class 3 have a mean service time of $\frac{1}{2}$ units ($\mu_3 = 2$). The arrival rate of customers across the classes

remain same ($\lambda_1 = \lambda_2 = \lambda_3$) and vary from 6 to 20 throughout the calculations. And we hold number of servers in the system for this analysis is 5.

In Table [1,2,3], we present three steady state performance measures from the above systems. The measures include the percentage of customers who receive service, the system throughput, and the mean waiting time of all customers. We discuss each of these measures:

- Percentage all customers receiving service:** The percentage of all customers who receive service is highest in System 2 while lowest in System 1; see Table 1. We observe that when the arrival rate is high, all systems are approximately equally efficient. Also recall that patience time in System 1 for Class 3 is higher than that is in System 2 and System 3. Consequently, a higher number of Class 3 customers are served in System 1 compared to System 2 and System 3. Since the patience time of Class 1 customers is highest in System 2, a maximum number of customers from Class 1 will be served in System 2. Also for Class 3 customers the patience time is the same in System 2 and System 3; approximately equal proportion of customers from Class will be served in

these two systems. That means the manager of the system can decide for which class of customers should be served at priority.

- System throughput:** We observed in Table 2 that the percentage of all customers receiving service is highest (lowest) in System 2 (System 1). Therefore, it is not surprising at all that the system throughput is highest (lowest) in System 2 (System 1). Initially, increasing the arrival rate in the system increases server utilization and hence system throughput. However, the gains in throughput due to an increase in server utilization diminish and are eventually offset by a reduction in the effective service rate of the system. The service rate of the system is decreasing, as the number of Class 1 customers are increasing and taking longer time to serve. The result has potential implications for systems with limited service capacity that generate revenue based on system throughput.
- Mean waiting time of all customers:** We observed that the mean waiting time is lowest in System 3 while highest in System 2; see Table 3. This is because the patience time of Class 1 customer is highest in System 2 which needs a longer time to serve. At the low arrival rate, all systems serve in approximately equal time,

Table 1 Percentage of all customers receiving service

λ	System 1 ($\theta_1 = \theta_2 = \theta_3 = 1.5$)	System 2 ($\theta_1 = 1, \theta_2 = 1.5, \theta_3 = 2$)	System 3 ($\theta_1 = 1.5, \theta_2 = 1.5, \theta_3 = 2$)
6	89.91741	90.45176	89.92555
8	77.75133	78.28859	77.71657
10	66.07731	66.49367	66.04104
12	56.50889	56.79464	56.4871
14	48.99	49.17586	48.98113
16	43.08472	43.20405	43.08357
18	38.38565	38.46308	38.38809
20	34.58333	34.63463	34.58692

Table 2 System throughput

λ	System 1 ($\theta_1 = \theta_2 = \theta_3 = 1.5$)	System 2 ($\theta_1 = 1, \theta_2 = 1.5, \theta_3 = 2$)	System 3 ($\theta_1 = 1.5, \theta_2 = 1.5, \theta_3 = 2$)
6	5.395044	5.427106	5.395533
8	6.220106	6.263087	6.217326
10	6.607731	6.649367	6.604104
12	6.781067	6.815357	6.778452
14	6.8586	6.88462	6.857358
16	6.893555	6.912648	6.893372
18	6.909417	6.923355	6.909857
20	6.916667	6.926925	6.917384

Table 3 Mean waiting time—all customers

λ	System 1 ($\theta_1 = \theta_2 = \theta_3 = 1.5$)	System 2 ($\theta_1 = 1, \theta_2 = 1.5, \theta_3 = 2$)	System 3 ($\theta_1 = 1.5, \theta_2 = 1.5, \theta_3 = 2$)
6	0.0672173	0.06884892	0.0615518
8	0.1483245	0.1566497	0.1361574
10	0.2261512	0.2418374	0.2075088
12	0.2899407	0.3119091	0.265897
14	0.3400667	0.3669599	0.3117698
16	0.3794352	0.4101124	0.3478129
18	0.4107623	0.4443707	0.3765095
20	0.4361111	0.4720342	0.3997403

while at the higher arrival rate System 3 has significantly less waiting time for service. Here it is interesting that in System 3, the percentage of all customers receiving service is lesser than that is in System 2 while having a lesser mean waiting time. This result again has ramifications for service level forecasting as the average waiting time of all customers is another common measure of service level.

6 Conclusion

In this article, we analyzed a queuing system with three classes of impatient customers who arrive according to $PP(\lambda_i)$ ($i = 1, 2, 3$) and are served on the basis of FCFS. We also determine the performance measures such as the percentage of all customers receiving service in each class, the mean waiting time of customers in each class and the expected conditional waiting time for each class of customers. After numerical analysis, we found that it may very effective in call centers since in call centers usually patience time of customers differs from each other. And a subset of servers can be trained to handle a class of customers. Overall this system has many managerial implications.

Acknowledgements The second author is thankful for the funding by IIT Madras from the IoE project: SB20210848MAMHRD008558.

References

- Adan, I., Boon, M., Weiss, G.: Design and evaluation of overloaded service systems with skill based routing, under fcfs policies. *Perform Evalu* **70**(10), 873–888 (2013)
- Adan, I., Hathaway, B., Kulkarni, V.: On first come, first served queues with two classes of impatient customers. *Queue Sys* **91**, 113–142 (2019)
- Baccelli, F., Boyer, P., Hebuterne, G.: Single-server queues with impatient customers. *Adv Appl Probab* **16**(4), 887–905 (1984)
- Benes, V. E.: *General stochastic processes in the theory of queues*. Addison-Wesley Publishing Company, Inc., (1963)
- Brandt, M., Brandt, A.: On the two-class m/m/1 system under preemptive resume and impatience of the prioritized customers. *Queueing Syst* **47**, 147–168 (2004)
- Brandt, M., Brandt, A.: Workload and busy period for M/GI/1 with a general impatience mechanism. *Queueing Syst* **75**, 189–209 (2013)
- Brill, M.J., Posner, P.: Level crossings in point processes applied to queues: single-server case. *Oper. Res.* **25**, 662–674 (1977)
- Brill, M.J., Posner, P.: The system point method in exponential queues: a level crossing approach. *Math. Oper. Res.* **6**, 31–49 (1981)
- Choi, B., Kim, B.D., Chung, J.: M/M/1 queue with impatient customers of higher priority. *Queueing Syst.* **38**, 49–66 (2001)
- Daley, D.: General customer impatience in the queue GI/G/1. *J. Appl. Probab.* **2**, 186–205 (1965)
- Garnett, A., Mandelbaum, O., Reiman, M.: Designing a call center with impatient customers. *Manuf. Serv. Oper. Manag.* **4**, 208–227 (2002)
- Houdt, B.: Analysis of the adaptive mmap[k]/ph[k]/1 queue: a multi-type queue with adaptive arrivals and general impatience. *Eur J Operat Res* **220**, 695–704 (2012)
- Iravani, B., Balcioglu, F.: Approximations for the M/GI/n+GI type call center. *Queueing Syst* **58**, 137–153 (2008)
- Iravani, B., Balcioglu, F.: On priority queues with impatient customers. *Queueing Syst.* **58**, 239–260 (2008)
- Sakuma, T., Takine, Y.: Multi-class m/ph/1 queues with deterministic impatience times. *Stoch. Models* **33**, 1–33 (2017)
- Stanford, R.E.: Reneging phenomena in single channel queues. *Math. Oper. Res.* **4**, 162–178 (1979)
- Takacs, M.: *Introduction to the Theory of Queues*. Oxford University Press, (1962)
- Zeltyn, A., Mandelbaum, S.: Call centers with impatient customers: many-server asymptotics of the M/M/n+G queue. *Queueing Syst.* **51**, 361–402 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.