RAIRO-Oper. Res. 59 (2025) 3423–3454
https://doi.org/10.1051/ro/2025128

RAIRO Operations Research
www.rairo-ro.org

ANALYSIS AND OPTIMISATION OF A HETEROGENEOUS SERVERS WITH VACATION SCHEME, SYSTEM DISASTER AND REPAIR

A. Mohammed shapique[1,*] and Ramupillai Sudhesh[2]

Abstract. This paper investigates a heterogeneous server system that incorporates vacation schemes, system disasters, server failures, and repairs. Servers are categorized into two types based on their service capabilities. Vacation durations vary for each server, and jobs are assigned to servers using a fastest-first policy. Time-dependent solutions are derived using generating functions and Laplace transforms, alongside steady-state solutions. The study explores key performance measures and presents a comprehensive cost-revenue analysis for the proposed model. To optimize system performance, the particle swarm optimization algorithm is employed to minimize the total expected cost function. Numerical results are presented to illustrate the impact of system and cost parameters, providing insights into the dynamic behavior and economic implications of the system.

Mathematics Subject Classification. 60K25, 60K30.

Received December 6, 2024. Accepted September 10, 2025.

1. Introduction

Optimizing resource utilization and maintaining high service quality are critical in today's computing, communication, and service industries. The analysis of multi-server queueing systems is fundamental for understanding resource allocation and service delivery. In recent times, heterogeneous service systems have emerged, where servers possess diverse processing capabilities and operational profiles. These systems reflect the increasing complexity of contemporary environments. Researchers can gain valuable insights into the interplay of system components and factors influencing performance by delving into these systems.

Heterogeneous queueing systems with vacation policies have been extensively studied, typically focusing on scenarios where only one server takes a vacation while the other remains continuously active. However, these models often assume a static vacation assignment. In this study, we consider a heterogeneous server system where the two servers are classified based on their service rates: the Faster Server (FS) and the Slower Server (SS). We propose a dynamic vacation policy in which either server may independently take a vacation under specific conditions. The FS is permitted to take a vacation after completing service and finding no customers waiting, while the SS may go on vacation following a system failure and subsequent repair, provided the system

Keywords. Heterogeneous server, sleep modes, system disaster, repair, transient analysis, steady state probabilities, cost-revenue analysis.

[1] Department of Mathematics, IFET College of Engineering, Villupuram, India.

[2] Department of Mathematics, UCE, BIT Campus, Anna University, Tiruchirappalli 620024, India.

*Corresponding author: shapique@gmail.com

is empty. This flexible policy enhances adaptability and offers improved performance in systems with variable workloads and asymmetric server capabilities.

The proposed model also accounts for system disasters, representing unforeseen events such as hardware failures or cyberattacks that significantly impact service continuity. Modeling such disruptions is especially relevant for high-reliability domains like cloud computing and telecommunication systems. Addressing both server unreliability and system-level disasters allows for more robust queueing models capable of guiding resource allocation and improving fault tolerance. Additionally, this study develops a cost optimization model for the system. Within a defined cost structure, we apply the particle swarm optimization (PSO) technique to determine optimal parameters that minimize total operational cost.

The proposed model can be effectively applied to cloud computing environments. Consider a data center with high-performance servers (HPS) and low-performance servers (LPS). HPS handle complex, computationally intensive tasks, while LPS handle routine tasks. During peak load, both types of servers are active. However, during off-peak periods, HPS can enter low-power states or be powered off, while LPS handle the reduced workload. This dynamic vacation policy, coupled with techniques like dynamic voltage and frequency scaling and power gating, can significantly reduce energy consumption. Furthermore, the model can enhance system reliability and fault tolerance. In case of HPS failures, LPS can take over the workload, albeit with reduced performance. This graceful degradation ensures continuous service availability. By incorporating these strategies, cloud service providers can optimize resource utilization, reduce operational costs, and improve overall system efficiency

1.1. Related work

Prior research in multi-server queuing models has primarily centered around homogeneous server configurations, where all servers in the system offer services at identical rates (see, *e.g.* [2,7,8]). However, this assumption is often not applicable in contemporary real-world scenarios, especially in technologically advanced environments. In modern queuing systems with human servers, electronic devices, or autonomous machines, the assumption of homogeneity is frequently unrealistic. For instance, in a customer service contact center with a mix of human agents and AI chatbots, the AI chatbots may process customer inquiries at different speeds. In automated manufacturing facilities with robot-assisted assembly lines, the robots may have distinct processing capabilities. In cloud computing environments, virtual machines or server instances may exhibit varying performance levels and response times, affecting the overall system's heterogeneity.

Many authors analysed queueing systems with heterogeneous servers. Krishnamoorthy [11] analysed queueing systems with two heterogeneous servers by introducing modified queue disciplines. The author explored two distinct queue disciplines: Queue Discipline-1 and Queue Discipline-2. In Queue Discipline-1, arriving units wait in a line if both servers are busy. When only one server is free, they occupy it. If both servers are free, they choose the fast server with probability π_1 and the slow server with probability π_2, where $\pi_1 + \pi_2 = 1$. Queue Discipline-2 differs in that, when both servers are free, arriving units directly occupy the fast server. If the fast server is busy, they wait for the first (fast) server, regardless of the status of the second (slow) server.

Singh [20] extends the work of Krishnamoorthy [11] by introducing the concept of balking. Singh's study focuses on comparing the performance of heterogeneous server queues with homogeneous server queues. The primary goal is to identify the conditions under which the heterogeneous system exhibits superior performance compared to its homogeneous counterpart. Leemans [15] investigated the distribution of queue lengths in a heterogeneous $M/M/2$ queuing system and presented a technique for computing the distribution of waiting times. Sharma and Dass [19] conducted an analysis of an $M/M/2/N$ queuing system featuring heterogeneous servers. Their investigation aimed to derive the probability density function of the busy period within this queuing system. Additionally, they calculated the mean and variance of the queuing system. Sujatha and Deekshitulu [23] analyzed a multiserver fractional queueing model with heterogeneous servers, acknowledging that servers may operate with varying efficiencies and may not always be fully active. Their work primarily focuses on the transient-state behaviour of the system.

On a related note, Dharmaraja [4] explores a different yet pertinent dimension of the field by investigating a two-processor heterogeneous system. In this system, jobs arrive according to a Poisson distribution and possess exponentially distributed processing times. The author obtained the transient solution for this model, shedding light on its dynamic behavior as it evolves over time. Sakalauskas *et al.* [18] analysed the stalling effect queueing system using Markovian finite capacity queueing model with heterogeneous servers. Melikov [17] examined a queueing system with heterogeneous servers with a Markov Modulated Poisson flow, and instantaneous feedback. In this model, the primary call is initially serviced on a high-speed server, but may require re-servicing after completion. Secondary calls, generated after primary call servicing, also have the possibility of re-servicing.

A vast body of literature in queuing theory exists, focusing on scenarios where servers remain continuously idle while awaiting customer arrivals. However, in numerous real-world applications, servers incorporate break periods when no customers arrive for extended durations. Consider a modern call center handling customer inquiries. Instead of having operators idly wait for calls during slow periods, the center employs an intelligent queuing system that allows operators to take brief breaks when no customer calls are in the queue for a specified time. This approach enhances operator well-being and work efficiency, contributing to improved customer service and overall operational productivity. Levy and Yechiali [16] laid the foundation for the exploration of the concept of vacations in the context of queueing systems and the determination of the optimal vacation duration through decomposition analysis. Subsequently, the review articles authored by Doshi [6] and Takagi [24] shed light on the escalating research endeavors in the realm of queueing systems featuring intermittent service interruptions.

Kothandaraman and Kandaiyan [10] investigated heterogeneous queueing systems with a hybrid vacation policy. In this model, server 1 is continuously accessible, while server 2 is either busy or on a working vacation. Additionally, server 2 undertakes specific tasks when the queue length reaches or exceeds zero, but it must complete its ongoing task before addressing any unusual requests. Kumar and Madheswari [14] applied the matrix geometric method to examine the distribution of stationary queue length in an $M/M/2$ heterogeneous system featuring multiple vacations. Krishnamoorthy and Sreenivasan [12] studied steady state analysis of an $M/M/2$ queueing system with heterogeneous servers. One of these servers remains consistently accessible, while the other takes brief hiatuses when no customers are awaiting service. It is worth noting that the server on vacation eventually resumes service, though at a reduced pace, in response to the arrival of a customer when the other server is occupied. The study primarily focuses on the examination of this system's steady-state behavior, utilizing the matrix geometric method for a comprehensive understanding.

Sridhar and Allah Pitchai [21] conducted an extensive investigation of an $M/M/2$ queueing system that includes a vacation policy. In this system, it's important to note that server 1 takes a single vacation, while server 2 takes multiple vacations. Specifically, if there are no customers in the system, both servers go on vacation. After their vacation period ends, if there are customers waiting for service, the servers return to their active state to serve those customers. If the system subsequently becomes empty, only server 2 goes on vacation, leaving server 1 in an idle state, ready to serve when needed. Yang *et al.* [27] investigated the steady-state analysis of an $M/M/2$ heterogeneous queue with multiple vacation and breakdowns. Within this study, Server 1 is depicted as dependable, with the ability to go on vacation when the system is devoid of customers. In contrast, Server 2 is deemed less reliable and may experience breakdowns while serving customers. Notably, when such breakdowns do occur, Server 2 doesn't entirely cease its service but instead reduces its service rate. Agarwal *et al.* [1] investigate an $M^X/M/2$ queueing system with heterogeneous servers. In this model, Server 1 is reliable and takes a vacation when the system is empty, while Server 2 is unreliable, subject to breakdowns during service. Upon failure, Server 2 enters immediate repair mode and continues to operate at reduced speed, rather than halting completely

System disasters represent catastrophic events that cause immediate termination of all services, destruction of existing customers, and require significant recovery time before normal operations can resume. Ammar [3] analysed system disasters within heterogeneous server queues, notably in the absence of vacation policies. Divya and Indhira [5] analyzed a heterogeneous queueing model featuring intermittent server availability, server catastrophes, and a hybrid vacation policy. In their model, Server 1 remains continuously available, while Server 2 experiences breakdowns and vacations. Using the matrix-geometric approach, they derived stationary probabilities and performance metrics. Recently Sudhesh *et al.* [22] discussed system disaster in dual-stage vacation queueing system. A review of the literature confirms an absence of research addressing queuing systems that simultaneously incorporate all four elements: heterogeneous servers, a dynamic vacation policy, system disasters, and server failures with repair processes. Motivated by this gap, we investigate the transient system size probabilities for a two-heterogeneous server vacation queueing system susceptible to disasters, server failures, and repairs. Furthermore, we study a cost model for the investigated system. To optimize operational costs, this study investigates a cost optimization problem within a predetermined cost structure, employing the particle swarm optimization (PSO) technique. PSO, a renowned evolutionary computation method developed by Kennedy and Eberhart [9], is leveraged to address the cost optimization challenge at hand. Recently, Tian *et al.* [25] examined a dual-server parallel queueing system with heterogeneous service rates and dynamic priority assignments. To analyze the system, the authors employed a truncated-approximation method combined with the matrix-geometric approach, enabling tractable computation of steady-state probabilities. Furthermore, PSO technique is applied to minimize operational costs, providing practical insights for improving system efficiency and resource allocation.

The subsequent sections of the paper are organized as follows: Section 2 offers a detailed description of the model for the investigated system. Section 3 delves into the transient analysis of the system. In Section 4, time-dependent performance measures of the system are discussed. Section 5 presents the steady-state probabilities of the system. The performance measures of the system in the steady state are outlined in Section 6. Section 7 provides a comprehensive cost analysis of the model. Section 8 offers a numerical illustration of the model. The conclusion and future scope of this work are discussed in Section 9.

1.2. Notations

λ:	Arrival rate
μ_1:	Service rate of the fast server
μ_2:	Service rate of the slow server
γ_1:	Idle timer of the fast server
γ_2:	Idle timer of the slow server
θ_1:	Vacation timer of the fast server
θ_2:	Vacation timer of the slow server
$P_{n,i,j}(t)$:	System size probabilities
$M(t)$:	Expected system size at time t
$B(t)$:	Number of busy servers at time t
$E\{B(t)\}$:	Mean of $B(t)$
$P(B(t) = m)$:	Probability that the system has m busy servers
$\pi_{n,i,j}$:	Steady-state probabilities of the system size
$E(N_s)$:	Expected system size in the steady state
$E(W_s)$:	Expected number of jobs waiting in the steady state
TEC:	Total expected cost
TER:	Total expected revenue

2. Description of the model

In this section, a model description of the investigated system is presented.

- **Job Arrival:** Jobs arrive according to a Poisson process with rate λ.
- **Job Service:** Jobs are served according to the Fastest Server First (FSF) discipline. Service times for the Fast Server (FS) and Slow Server (SS) follow exponential distributions with rates μ_1 and μ_2, respectively, where $\mu_1 > \mu_2$.
- **Server Vacation:** The duration of vacations is modelled as an exponentially distributed random variable. This constitutes a standard assumption within the $M/M/1$ and $M/M/c$ vacation queueing literature [24],

primarily due to the resulting memoryless property and analytical tractability. Such models are frequently employed to analyze scenarios where servers undergo rest periods, for instance for maintenance or energy saving, in domains including wireless communications and manufacturing.

- **FS Vacation:**
  - Upon completing all jobs in the system, the FS activates an idle timer with parameter γ_1.
  - If no new job arrives before the timer expires, the FS enters a vacation state. The vacation duration is exponentially distributed with rate θ_1.
  - During the FS vacation, jobs continue to arrive and are served by the SS.
- **SS Vacation:**
  - After a system failure and repair, if no jobs arrive, the SS sets an idle timer γ_2.
  - If no jobs arrive before the timer expires, the SS enters a vacation state. The vacation duration follows an exponential distribution with rate θ_2.
  - During the SS vacation, jobs continue to arrive and are served by the FS.

– **System disaster, failure and repair:** The model also incorporates system disasters, which occur according to a Poisson process with rate ζ, simultaneously disabling both servers. Upon such an event, all jobs are removed, and the system transitions into a failure state. These catastrophic events are modeled using a Poisson process, consistent with prior studies [13]. The system undergoes immediate repair, with the repair time assumed to follow an exponential distribution with mean η^{-1}. This modeling choice-standard in Markovian reliability analysis [26] captures the memoryless nature of time-to-failure and time-to-repair processes commonly observed in computing and telecommunications domains.

We define $\{R(t), S(t), t \geq 0\}$ as the state of the system at any given time t. Additionally, let $N(t)$ represent the number of jobs present in the system.

$$(R(t), S(t)) = \begin{cases} (0,0), & \text{if both FS and SS are idle.} \\ (1,0), & \text{if FS is busy and SS is idle.} \\ (0,1), & \text{if FS is idle and SS is busy.} \\ (1,1), & \text{if both FS and SS are busy.} \\ (2,0), & \text{if FS is in the vacation state and SS is idle.} \\ (2,1), & \text{if FS is in the vacation state and SS is busy.} \\ (0,2), & \text{if FS is idle and SS is in the vacation state.} \\ (1,2), & \text{if FS is busy and SS is in the vacation state.} \end{cases}$$

Then, $\{N(t), R(t), S(t), t \geq 0\}$ represents the three dimensional Markov process with state space $W = (\{0\} \cup Z^+) \times \{0,1,2\} \times \{0,1,2\}$.

$$\text{Let } P_{n,i,j}(t) = P\big\{N(t) = n, R(t) = i, S(t) = j\big\}, i = 0,1,2; j = 0,1,2, n = 0,1,2,\ldots$$

Then $P_{n,i,j}(t)$ satisfies the following forward Kolmogorov equation. Figure 1 presents the investigated model.

$$P'_{0,0,0}(t) = -(\lambda + \gamma_1 + \gamma_2 + \zeta)P_{0,0,0}(t) + \mu_1 P_{1,1,0}(t) + \mu_2 P_{1,0,1}(t) + \eta P_F(t) \tag{2.1}$$
$$P'_{1,1,0}(t) = -(\lambda + \mu_1 + \zeta)P_{1,1,0}(t) + \lambda P_{0,0,0}(t) + \mu_2 P_{2,1,1}(t) \tag{2.2}$$
$$P'_{1,0,1}(t) = -(\lambda + \mu_2 + \zeta)P_{1,0,1}(t) + \mu_1 P_{2,1,1}(t) \tag{2.3}$$
$$P'_{2,1,1}(t) = -(\lambda + \mu + \zeta)P_{2,1,1}(t) + \lambda P_{1,1,0}(t) + \lambda P_{1,0,1}(t) + \mu P_{3,1,1}(t) + \theta_1 P_{2,2,1}(t) + \theta_2 P_{2,1,2}(t)) \tag{2.4}$$
$$P'_{n,1,1}(t) = -(\lambda + \mu + \zeta)P_{n,1,1}(t) + \lambda P_{n-1,1,1}(t) + \mu P_{n+1,1,1}(t) + \theta_1 P_{n,2,1}(t) + \theta_2 P_{n,1,2}(t), \quad n \geq 3 \tag{2.5}$$
$$P'_{0,2,0}(t) = -(\lambda + \zeta)P_{0,2,0}(t) + \gamma_1 P_{0,0,0}(t) + \mu_2 P_{1,2,1}(t) \tag{2.6}$$
$$P'_{1,2,1}(t) = -(\lambda + \mu_2 + \zeta)P_{1,2,1}(t) + \lambda P_{0,2,0}(t) + \mu_2 P_{2,2,1}(t) \tag{2.7}$$
$$P'_{n,2,1}(t) = -(\lambda + \mu_2 + \theta_1 + \zeta)P_{n,2,1}(t) + \lambda P_{n-1,2,1}(t) + \mu_2 P_{n+1,2,1}(t), \quad n \geq 2 \tag{2.8}$$

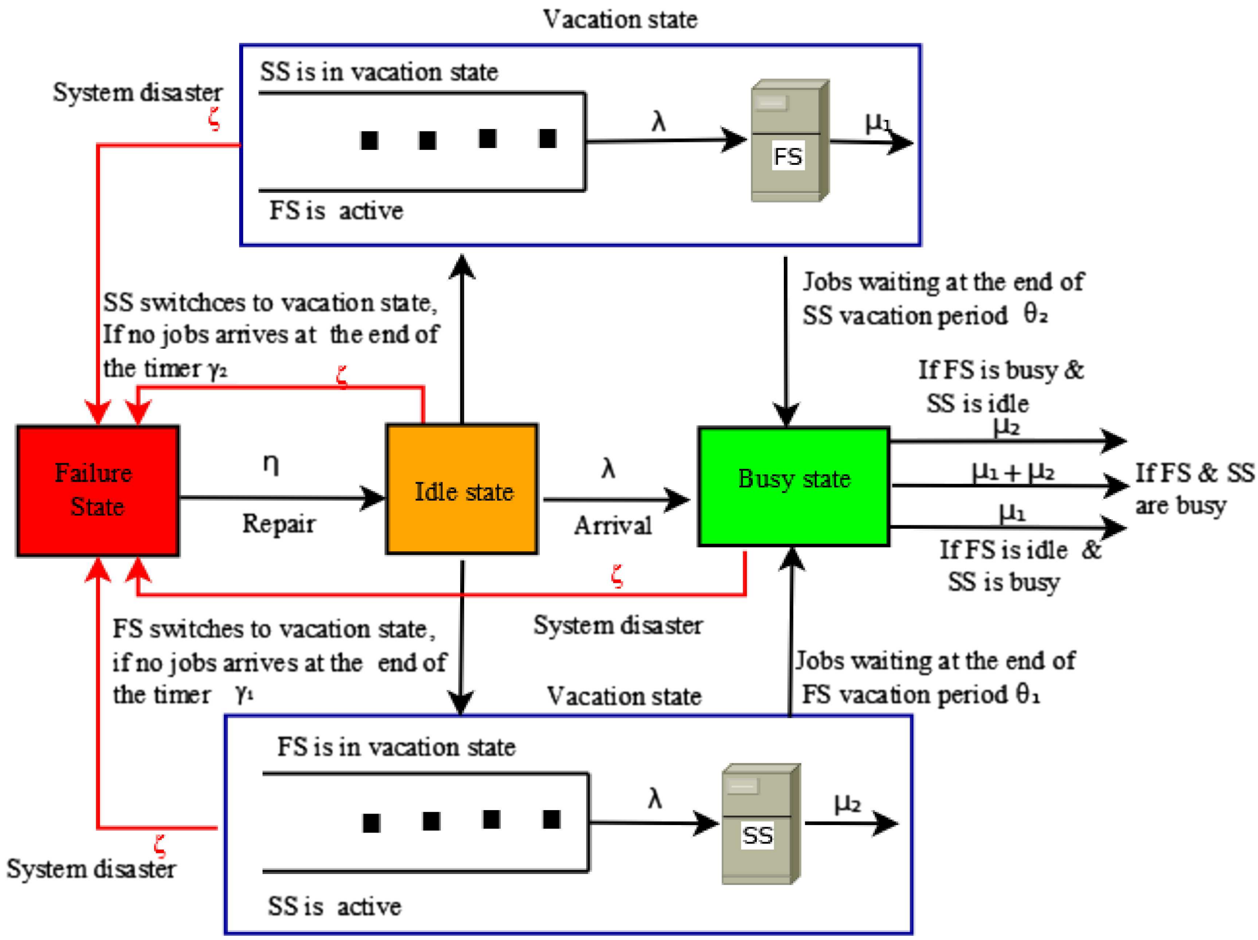

FIGURE 1. Pictorial representation of the investigated model.

$$P'_{0,0,2}(t) = -(\lambda+\zeta)P_{0,0,2}(t) + \gamma_2 P_{0,0,0}(t) + \mu_1 P_{1,1,2}(t) \tag{2.9}$$
$$P'_{1,1,2}(t) = -(\lambda+\mu_1+\zeta)P_{1,1,2}(t) + \lambda P_{0,0,2}(t) + \mu_1 P_{2,1,2}(t) \tag{2.10}$$
$$P'_{n,1,2}(t) = -(\lambda+\mu_1+\theta_2+\zeta)P_{n,1,2}(t) + \lambda P_{n-1,1,2}(t) + \mu_1 P_{n+1,1,2}(t), \quad n \geq 2 \tag{2.11}$$
$$P'_F(t) = \zeta(1-P_F(t)) - \eta P_F(t) \tag{2.12}$$

where $\mu = \mu_1 + \mu_2$. It is assumed that the server is in the idle state initially. Therefore $P_{0,0,0}(0) = 1$. Equations (2.1)–(2.12) represent the forward Kolmogorov differential equations governing the time evolution of the state probabilities in the Markovian queueing model. These equations are derived by balancing the rates of transitions into and out of each state, considering arrivals λ, service completions (μ_1, μ_2), vacation entries (γ_1, γ_2), vacation exits (θ_1, θ_2), server breakdowns ζ, and repairs η.

3. Transient analysis

In this section the transient probabilities of the busy state and vacation state are obtained.

3.1. Evaluation of $P_{n+2,1,1}(t)$, $n \in Z^+$

The transient probability of $P_{n+2,1,1}(t)$, $n \in Z^+$ is obtained using the generating as follows. Let,

$$G_1(z,t) = J(t) + \sum_{n=0}^{\infty} P_{n+3,1,1} z^{n+1}, G_1(z,0) = 1,$$

where

$$J(t) = P_{0,0,0}(t) + P_{1,1,0}(t) + P_{1,0,1}(t) + P_{2,1,1}(t). \tag{3.1}$$

Multiplying z^n, $n = 1,2,3,\ldots$ on both sides of equation (2.5), we get

$$G_1'(z,t) = G_1(z,t)\Big\{-\tau_1 + \lambda z + \frac{\mu}{z}\Big\} + \lambda z P_{2,1,1}(t) - \mu P_{3,1,1}(t) + \theta_1 \sum_{n=1}^{\infty} P_{n+2,2,1}(t) z^n + \theta_2 \sum_{n=1}^{\infty} P_{n+2,1,2}(t) z^n,$$

where, $\tau_1 = \lambda + \mu + \zeta$. Using the above expression and equations (2.1)–(2.4), we get

$$\begin{aligned} G_1'(z,t) &= \lambda(z-1)P_{2,1,1}(t) + \eta P_F(t) - P_{0,0,0}(t)(\gamma_1+\gamma_2) \\ &\quad + \theta_1\left\{P_{2,2,1}(t) + \sum_{n=1}^{\infty} P_{n+2,2,1}(t) z^n\right\} + \theta_2\left\{P_{2,1,2}(t) + \sum_{n=1}^{\infty} P_{n+2,1,2}(t) z^n\right\} \\ &\quad + G_1(z,t)\Big\{-\tau_1 + \lambda z + \frac{\mu}{z}\Big\} - J(t)\Big\{-(\lambda+\mu) + \lambda z + \frac{\mu}{z}\Big\}. \end{aligned}$$

Integrating the above expression yields,

$$\begin{aligned} G_1(z,t) = \int_0^t &\Bigg[\lambda(z-1)P_{2,1,1}(w) + \eta P_F(w) - (\gamma_1+\gamma_2)P_{0,0,0}(w) \\ &+ \theta_1\left\{P_{2,2,1}(w) + \sum_{n=1}^{\infty} P_{n+2,2,1}(w) z^n\right\} + \theta_2\left\{P_{2,1,2}(w) + \sum_{n=1}^{\infty} P_{n+2,1,2}(w) z^n\right\} \\ &+ \Big\{(\lambda+\mu) - \lambda z - \frac{\mu}{z}\Big\}J(w)\Bigg] \exp\{-(\lambda+\mu+\zeta)(t-w)\} \exp\Big\{\Big(\lambda z + \frac{\mu}{z}\Big)(t-w)\Big\}\,\mathrm{d}w \\ &+ \exp\{-\tau_1(t)\} \exp\left\{\Big(\lambda z + \frac{\mu}{z}\Big)(t)\right\}. \end{aligned} \tag{3.2}$$

Let $\kappa_1 = 2\sqrt{\lambda\mu}$ and $\nu_1 = \sqrt{\lambda\mu^{-1}}$ then

$$\exp\Big\{\Big(\lambda z + \frac{\mu}{z}\Big)(t-w)\Big\} = \sum_{n=-\infty}^{\infty} (\nu_1 z)^n I_n(\kappa_1(t-w)).$$

Using the above expression in equation (3.2) and comparing the coefficient z^n, $n \in Z^+$

$$\begin{aligned} {\nu_1}^{-n} P_{n+2,1,1}(t) = \int_0^t &\Bigg[\lambda P_{2,1,1}(w)\big\{{\nu_1}^{-1} I_{n-1}(.) - I_n(.)\big\} \\ &+ \{\eta P_F(w) - (\gamma_1+\gamma_2)P_{0,0,0}(w) + \theta_1 P_{2,2,1}(w) + \theta_2 P_{2,1,2}(w)\} I_n(.) \\ &+ \sum_{m=1}^{\infty} \{\theta_1 P_{m+2,2,1}(w) + \theta_2 P_{m+2,1,2}(w)\} {\nu_1}^{-m} I_{n-m}(.) \end{aligned}$$

$$+ \left\{(\lambda+\mu)I_n(.) - \lambda{\nu_1}^{-1}I_{n-1}(.) - \mu\nu_1 I_{n+1}(.)\right\}J(w)\Bigg] \exp(-\tau_1(t-w))\,\mathrm{d}w$$
$$+ \exp\{-\tau_1(t)\}I_n(\kappa_1(t)), \tag{3.3}$$

where $I_n(\kappa_1(t-w)) = I_n(.)$.

The above equation holds for $n \in Z^-$ and using the property $I_{-n}(.) = I_n(.)$, we obtain

$$\begin{aligned} 0 = \int_0^t \Bigg[&\lambda P_{2,1,1}(w)\left\{{\nu_1}^{-1}I_{n+1}(.) - I_n(.)\right\} \\ &+ \{\eta P_F(w) - (\gamma_1+\gamma_2)P_{0,0,0}(w) + \theta_1 P_{2,2,1}(w) + \theta_2 P_{2,1,2}(w)\}I_n(.) \\ &+ \sum_{m=1}^{\infty}\{\theta_1 P_{m+2,2,1}(w) + \theta_2 P_{m+2,1,2}(w)\}{\nu_1}^{-m}I_{n+m}(.) \\ &+ \left\{(\lambda+\mu)I_n(.) - \lambda{\nu_1}^{-1}I_{n+1}(.) - \mu\nu_1 I_{n-1}(.)\right\}J(w)\Bigg] \exp(-\tau_1(t-w))\,\mathrm{d}w \\ &+ \exp\{-\tau_1(t)\}I_n(\kappa_1(t)). \end{aligned} \tag{3.4}$$

Subtracting equation (3.4) from equation (3.3), we get

$$\begin{aligned} P_{n+2,1,1}(t) = \int_0^t \exp(-\tau_1(t-w))\Bigg[&\lambda{\nu_1}^{n-1}f_n(\kappa_1(t-w))P_{2,1,1}(w) \\ &+ \left(\sum_{m=1}^{\infty}\theta_1 P_{m+2,2,1}(w) + \theta_2\sum_{m=1}^{\infty}P_{m+2,1,2}(w)\right){\nu_1}^{n-m}g_{n,m}(\kappa_1(t-w))\Bigg]\,\mathrm{d}w, \end{aligned} \tag{3.5}$$

where

$$f_n(\kappa_1 t) = (I_{n-1}(\kappa_1 t) - I_{n+1}(\kappa_1 t)) = \frac{2nI_n(\kappa_1 t)}{\kappa_1 t}, \tag{3.6}$$

and

$$g_{n,m}(\kappa_1 t) = (I_{n-m}(\kappa_1 t) - I_{n+m}(\kappa_1 t)). \tag{3.7}$$

3.2. Evaluation of $P_{2,1,1}(t)$

The probability that both the servers are busy when there are two jobs in the system is derived in this section. Equations (2.1)–(2.3) is expressed as

$$K'(t) = AK(t) + \eta P_F(t)e_1 + \mu_2 P_{2,1,1}(t)e_2 + \mu_1 P_{2,1,1}(t)e_3, \tag{3.8}$$

where

$$K(t) = \begin{bmatrix} P_{0,0,0}(t) \\ P_{1,1,0}(t) \\ P_{1,0,1}(t) \end{bmatrix},$$
$$A = \begin{bmatrix} \lambda+\gamma_1+\gamma_2+\zeta & \mu_1 & \mu_2 \\ \lambda & -(\lambda+\mu_1+\zeta) & 0 \\ 0 & 0 & -(\lambda+\mu_2+\zeta) \end{bmatrix},$$

and

$$e_i = \begin{cases} [1,0,0]^T & \text{if } i=1 \\ [0,1,0]^T & \text{if } i=2 \\ [0,0,1]^T & \text{if } i=3. \end{cases}$$

Taking Laplace transform on equation (3.8) yields

$$\hat{K}(s) = (sI-K)^{-1}\Big[K(0) + \eta\hat{P}_F(s)e_1 + \mu_2\hat{P}_{2,1,1}(s)e_2 + \mu_1\hat{P}_{2,1,1}(s)e_3\Big]. \tag{3.9}$$

with $K(0) = [1,0,0]^T$. Equation (3.1) can be expressed as

$$\hat{J}(s) = e^T\hat{K}(s) + \hat{P}_{2,1,1}(s), \tag{3.10}$$

where $e = [1,1,1]^T$. Now, comparing the constant term of equation (3.2), we obtain

$$\begin{aligned} J(t) = \int_0^t \Bigg[&\lambda\{{\nu_1}^{-1}I_1(.) - I_0(.)\}P_{2,1,1}(w) + I_0(.)\Big\{\eta P_F(w) - (\gamma_1+\gamma_2)P_{0,0,0}(w) \\ &+ \theta_1 P_{2,2,1}(w) + \theta_2 P_{2,1,2}(w)\Big\} + \left\{\theta_1\sum_{m=1}^{\infty}P_{m+2,2,1}(w) + \theta_2\sum_{m=1}^{\infty}P_{m+2,1,2}(w)\right\}{\nu_1}^{-m}I_n(.) \\ &+ \left\{(\lambda+\mu)I_0(.) - \lambda{\nu_1}^{-1}I_1(.) - \mu\nu_1 I_1(.)\right\}J(w)\Bigg]\exp(-\tau_1(t-w))\,\mathrm{d}w + \exp(-\tau_1(t))I_0(\kappa_1 t). \end{aligned} \tag{3.11}$$

Taking Laplace transform on equation (3.11), we obtain

$$\begin{aligned} \hat{J}(s)(s+\zeta) = \frac{1}{2}&\left(\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2} - 2\lambda\right)\hat{P}_{2,1,1}(s) + \eta\hat{P}_F(s) - (\gamma_1+\gamma_2)\hat{P}_{0,0,0}(s) \\ &+ \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s) + \theta_1\sum_{m=1}^{\infty}\hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m \\ &+ \theta_2\sum_{m=1}^{\infty}\hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m + 1, \end{aligned} \tag{3.12}$$

where $\omega_1 = s+\lambda+\mu+\zeta$. Using equations (3.9) and (3.10) in (3.12), we obtain

$$\begin{aligned} &e^T(sI-K)^{-1}(s+\zeta)\Big\{K(0) + \eta\hat{P}_F(s)e_1 + \mu_2\hat{P}_{2,1,1}(s)e_2 + \mu_1\hat{P}_{2,1,1}(s)e_3\Big\} + \hat{P}_{2,1,1}(s)(s+\zeta) \\ &= \frac{1}{2}\left(\omega_1 - \sqrt{\omega_1^2 - \kappa_1^2} - 2\lambda\right)\hat{P}_{2,1,1}(s) + \Big\{\eta\hat{P}_F(s) - \hat{P}_{0,0,0}(s)(\gamma_1+\gamma_2)\Big\} \\ &\quad + \theta_1\sum_{m=1}^{\infty}\hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - \kappa_1^2}}{2\lambda}\right)^m + \theta_2\sum_{m=1}^{\infty}\hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - \kappa_1^2}}{2\lambda}\right)^m \\ &\quad + \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s) + 1. \end{aligned}$$

After some algebraic manipulation, we get

$$\hat{P}_{2,1,1}(s) = \left[e^T\{sI-K\}^{-1}(s+\zeta)\{e_2\mu_2 + e_3\mu_1\} + (s+\zeta) - \frac{1}{2}\left(\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2} - 2\lambda\right)\right]^{-1}$$

$$\times \left[\theta_1 \sum_{m=1}^{\infty} \hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m + \theta_2 \sum_{m=1}^{\infty} \hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m \right.$$
$$+ 1 + \eta\hat{P}_F(s) - (\gamma_1+\gamma_2)\hat{P}_{0,0,0}(s) + \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s) - e^T(sI-K)^{-1}$$
$$\left. \times (s+\zeta)\Big(1+\eta\hat{P}_F(s)e_1\Big)\right]. \tag{3.13}$$

Let

$$(sI-K)^{-1} = \left[\hat{k}_{ij}(s)\right]_{3\times 3} = \frac{1}{|C(s)|}\begin{bmatrix} k_1(s)k_2(s) & \mu_1 k_2(s) & \mu_2 k_1(s) \\ \lambda k_2(s) & k_2(s) & \lambda\mu_2 \\ 0 & 0 & l(s)k_1(s) - \lambda\mu_1 \end{bmatrix}, \tag{3.14}$$

where

$$k_1(s) = (s+\lambda+\mu_1+\zeta), k_2(s) = (s+\lambda+\mu_2+\zeta), l(s) = (s+\lambda+\gamma_1+\gamma_2+\zeta),$$

and

$$|C(s)| = \left[s^2 + s(2\lambda+\mu_1+\gamma_1+\gamma_2+2\zeta) + (\lambda+\gamma_1+\gamma_2+\zeta)(\lambda+\mu_1+\zeta) - \mu_1\lambda\right] \times (\lambda+\mu_2+\zeta).$$

The characteristic roots of (3.14) are given by

$$s_1 = -(\lambda+\mu_2+\zeta),$$
$$s_2, s_3 = \frac{-(2\lambda+\mu_1+\gamma_1+\gamma_2+2\zeta) \pm \sqrt{(\mu_1+\gamma_1+\gamma_2)^2 + 4\mu_1\lambda}}{2}.$$

We notice that the functions $\hat{k}_{ij}(s)$ are rational algebraic expressions in s. The cofactor of the (i,j)th element of $(sI-K)$ is a polynomial with a degree of $2-|i-j|$. As the characteristic roots of matrix K are all distinct, the inverse transform $k_{ij}(t)$ of $\hat{k}_{ij}(s)$ can be determined through partial fraction decomposition. Using equation (3.14), we get

$$e^T(sI-K)^{-1}e_i = \sum_{j=1}^{3}\hat{k}_{ji}(s), \qquad i=1,2,3, \tag{3.15}$$

and

$$e^T(sI-K)^{-1}(e_2\mu_2 + e_3\mu_1) = \mu_2\sum_{j=1}^{3}\hat{k}_{j2}(s) + \mu_1\sum_{j=1}^{3}\hat{k}_{j3}(s). \tag{3.16}$$

Applying the equations (3.15) and (3.16) in equation (3.13), we get

$$\hat{P}_{2,1,1}(s) = \left[\mu_2\hat{h}_2(s) + \mu_1\hat{h}_3(s) + (s+\zeta) - \frac{1}{2}\left(\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2} - 2\lambda\right)\right]^{-1}$$
$$\times \left[\theta_1 \sum_{m=1}^{\infty} \hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m + \theta_2 \sum_{m=1}^{\infty} \hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - {\kappa_1}^2}}{2\lambda}\right)^m \right.$$
$$\left. + 1 + \eta\hat{P}_F(s) - (\gamma_1+\gamma_2)\hat{P}_{0,0,0}(s) + \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s) - \hat{h}_1(s) - \eta\hat{P}_F(s)\hat{h}_1(s)\right], \tag{3.17}$$

where

$$\hat{h}_i(s) = (s+\zeta)\sum_{j=1}^{2}\hat{k}_{ji}(s), \qquad i=1,2,3,$$

and

$$\mu_2\hat{h}_2(s) + \mu_1\hat{h}_3(s) = e^T(sI-K)^{-1}(s+\zeta)(e_2\mu_2 + e_3\mu_1).$$

Equation (3.17) can be expressed as

$$\hat{P}_{2,1,1}(s) = \frac{2}{\omega + \sqrt{\omega^2 - \kappa_1{}^2}}\left[1 - \frac{\sqrt{\mu}\left(\omega_1 - \sqrt{\omega_1{}^2 - \kappa_1{}^2}\right)}{\kappa_1\sqrt{\lambda}}\left\{1 - \frac{\left(\mu_2\hat{h}_2(s) + \mu_1\hat{h}_3(s)\right)}{\mu}\right\}\right]^{-1}$$
$$\times\left[\theta_1\sum_{m=1}^{\infty}\hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1{}^2 - \kappa_1{}^2}}{2\lambda}\right)^m + \theta_2\sum_{m=1}^{\infty}\hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1{}^2 - \kappa_1{}^2}}{2\lambda}\right)^m\right.$$
$$\left. + 1 + \eta\hat{P}_F(s) - (\gamma_1+\gamma_2)\hat{P}_{0,0,0}(s) + \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s) - \hat{h}_1(s) \; - \eta\hat{P}_F(s)\hat{h}_1(s)\right].$$

After some manipulation

$$\hat{P}_{2,1,1}(s) = \frac{2}{\kappa_1}\left[1 + \eta\hat{P}_F(s)(1-\hat{h}_1(s)) - \hat{P}_{0,0,0}(s)(\gamma_1+\gamma_2) - \hat{h}_1(s) + \theta_1\hat{P}_{2,2,1}(s) + \theta_2\hat{P}_{2,1,2}(s)\right.$$
$$\left. + \theta_1\sum_{m=1}^{\infty}\hat{P}_{m+2,2,1}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - \kappa_1{}^2}}{2\lambda}\right)^m + \theta_2\sum_{m=1}^{\infty}\hat{P}_{m+2,1,2}(w)\left(\frac{\omega_1 - \sqrt{\omega_1^2 - \kappa_1{}^2}}{2\lambda}\right)^m\right]$$
$$\times\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(-1)^j\frac{1}{\nu_1{}^i}\left(\frac{\omega_1 - \sqrt{\omega_1^2 - \kappa_1{}^2}}{\kappa_1}\right)^{i+1}\left(\frac{\mu_2\hat{h}_2(s) + \mu_1\hat{h}_3(s)}{\mu}\right)^j. \tag{3.18}$$

On inversion, we get

$$P_{2,1,1}(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(-1)^j\frac{1}{\nu_1{}^i}\frac{1}{\mu^j}\int_0^t\left[\int_0^w\left\{\delta(v) + \int_0^v\eta P_F(u)(\delta(v-u) - h_1(v-u))\,\mathrm{d}u\right.\right.$$
$$\left. - (\gamma_1+\gamma_2)P_{0,0,0}(v) - h_1(v) + \theta_1P_{2,2,1}(v) + \theta_2P_{2,1,2}(v)\right\}f_{i+1}(\kappa_1(w-v))$$
$$\times\exp(-\tau_1(w-v))\,\mathrm{d}v + \int_0^w\sum_{m=0}^{\infty}\{\theta_1P_{m+2,2,1}(v) + \theta_2P_{m+2,1,2}(v)\}\nu_1{}^{-m}$$
$$\left.\times f_{m+i+1}(\kappa_1(w-v))\exp(-\tau_1(w-v))\,\mathrm{d}v\right](\mu_2h_2(t-w) + \mu_1h_3(t-w))^{*j}\,\mathrm{d}w. \tag{3.19}$$

3.3. Evaluation of $P_{0,0,0}(t)$, $P_{1,0,1}(t)$ and $P_{1,1,0}(t)$

The probability that both the servers are idle $P_{0,0,0}(t)$, FS is idle and SS is busy $P_{1,0,1}(t)$ and FS is busy and SS is idle $P_{1,0,1}(t)$ is derived in this section. Using equation (3.14) in equation (3.9), we get

$$\hat{P}_{0,0,0}(s) = \left\{1 + \eta\hat{P}_F(s)\right\}\hat{K}_{11}(s) + \left\{\mu_2\hat{P}_{2,1,1}(s)\right\}\hat{K}_{12}(s) + \left\{\mu_1\hat{P}_{2,1,1}(s)\right\}\hat{K}_{13}(s), \tag{3.20}$$
$$\hat{P}_{1,0,1}(s) = \left\{1 + \eta\hat{P}_F(s)\right\}\hat{K}_{21}(s) + \left\{\mu_2\hat{P}_{2,1,1}(s)\right\}\hat{K}_{22}(s) + \left\{\mu_1\hat{P}_{2,1,1}(s)\right\}\hat{K}_{23}(s), \tag{3.21}$$
$$\hat{P}_{1,1,0}(s) = \left\{1 + \eta\hat{P}_F(s)\right\}\hat{K}_{31}(s) + \left\{\mu_2\hat{P}_{2,1,1}(s)\right\}\hat{K}_{32}(s) + \left\{\mu_1\hat{P}_{2,1,1}(s)\right\}\hat{K}_{33}(s). \tag{3.22}$$

Inverting equations (3.20)–(3.22), we get

$$P_{0,0,0}(t) = K_{11}(t) + \eta \int_0^t P_F(u)K_{11}(t-w)\,\mathrm{d}w + \mu_2 \int_0^t P_{2,1,1}(w)K_{12}(t-w)\,\mathrm{d}w$$
$$+ \mu_1 \int_0^t P_{2,1,1}(w)K_{13}(t-w)\,\mathrm{d}w, \tag{3.23}$$

$$P_{1,0,1}(t) = K_{21}(t) + \eta \int_0^t P_F(u)K_{21}(t-w)\,\mathrm{d}w + \mu_2 \int_0^t P_{2,1,1}(w)K_{22}(t-w)\,\mathrm{d}w$$
$$+ \mu_1 \int_0^t P_{2,1,1}(w)K_{23}(t-w)\,\mathrm{d}w, \tag{3.24}$$

$$P_{1,1,0}(t) = K_{31}(t) + \eta \int_0^t P_F(u)K_{31}(t-w)\,\mathrm{d}w + \mu_2 \int_0^t P_{2,1,1}(w)K_{32}(t-w)\,\mathrm{d}w$$
$$+ \mu_1 \int_0^t P_{2,1,1}(w)K_{33}(t-w)\,\mathrm{d}w. \tag{3.25}$$

3.4. Evaluation of $P_{n,2,1}(t)$ and $P_{n,1,2}(t)$

The probability that FS is in vacation while the SS is busy $P_{n,2,1}(t)$ and FS is busy while SS is vacation $P_{n,1,2}(t)$ is derived in this section. We define generating functions as follows.

$$G_2(z,t) = \sum_{n=1}^{\infty} P_{n,2,1}z^n, G_2(z,0) = 0,$$

and

$$G_3(z,t) = \sum_{n=1}^{\infty} P_{n,1,2}z^n, G_3(z,0) = 0.$$

Using equations (2.7) and (2.8), we obtain

$$G_2'(z,t) = \left\{-(\lambda + \mu_2 + \theta_1 + \zeta) + \lambda z + \frac{\mu_2}{z}\right\}G_2(z,t) - \mu_2 P_{1,2,1}(t) + \lambda z P_{0,2,0}(t) + \theta_1 z P_{1,2,1}(t). \tag{3.26}$$

Integrating the above equation, we get

$$G_2(z,t) = \int_0^t \left[\lambda z P_{0,2,0}(t) - \mu_2 P_{1,2,1}(t) + \theta_1 z P_{1,2,1}(t)\right] \exp\{-(\lambda + \mu_2 + \theta_1 + \zeta)(t-w)\}$$
$$\times \exp\left(\lambda z + \mu_2 z^{-1}\right)(t-w)\,\mathrm{d}w. \tag{3.27}$$

Let $\kappa_i = 2\sqrt{\lambda\mu_i}, \nu_i = \sqrt{\frac{\lambda}{\mu_i}}, i = 2, 3$, $\tau_2 = \lambda + \mu_2 + \theta_1 + \zeta$ and $\tau_3 = \lambda + \mu_1 + \theta_2 + \zeta$. then

$$\exp\left\{\left(\lambda z + \frac{\mu_2}{z}\right)(t-w)\right\} = \sum_{n=-\infty}^{\infty} (\nu_2 z)^n I_n \kappa_2(t-w). \tag{3.28}$$

Using equation (3.28) in equation (3.27) and applying the method [22], we obtain

$$P_{n,2,1}(t) = \int_0^t \{\lambda P_{0,2,0}(t) + \theta_1 P_{1,2,1}(t)\} {\nu_2}^{n-1} f_n(\kappa_2 t) \exp(-\tau_2(t)) \,\mathrm{d}w. \tag{3.29}$$

Similarly, using equations (2.10) and (2.11), we get

$$P_{n,1,2}(t) = \int_0^t \{\lambda P_{0,0,2}(t) + \theta_2 P_{1,1,2}(t)\} {\nu_3}^{n-1} f_n(\kappa_3 t) \exp(-\tau_3(t)) \,\mathrm{d}w. \tag{3.30}$$

The expression for $P_{0,0,2}(t)$, $P_{0,2,0}(t)$, $P_{1,2,1}(t)$ and $P_{1,1,2}(t)$ is derived in Section 3.5.

3.5. Evaluation of $P_{0,0,2}(t)$, $P_{0,2,0}(t)$, $P_{1,2,1}(t)$ and $P_{1,1,2}(t)$

Taking Laplace transform on equations (3.29) and (3.30), we get

$$\hat{P}_{n,2,1}(s) = \left\{\lambda \hat{P}_{0,2,0}(s) + \theta_1 \hat{P}_{1,2,1}(s)\right\} {\nu_2}^{n-1} \hat{f}_n(\omega_2 \kappa_2), \tag{3.31}$$

$$\hat{P}_{n,1,2}(s) = \left\{\lambda \hat{P}_{0,0,2}(s) + \theta_2 \hat{P}_{1,1,2}(s)\right\} {\nu_3}^{n-1} \hat{f}_n(\omega_3 \kappa_3). \tag{3.32}$$

Taking Laplace transform of (2.6), we get

$$\hat{P}_{0,2,0}(s) = \frac{\gamma_1}{(s+\lambda+\zeta)} \hat{P}_{0,0,0}(s) + \frac{\mu_2}{(s+\lambda+\zeta)} \hat{P}_{1,2,1}(s). \tag{3.33}$$

Setting $n = 1$ in equation (3.31), after some manipulation, we obtain

$$\hat{P}_{1,2,1}(s) = \frac{\lambda}{\theta_1} \sum_{k=0}^{\infty} \left\{\theta_1 \hat{f}_1(\omega_2 \kappa_2)\right\}^{k+1} \hat{P}_{0,2,0}(s). \tag{3.34}$$

Using equation (3.34) in equation (3.33) after some manipulation, we get

$$\hat{P}_{0,2,0}(s) = \frac{\gamma_1}{\theta_1(s+\lambda+\zeta)} \sum_{i=0}^{\infty} \left(\frac{\lambda\mu_2}{s+\lambda+\zeta}\right)^i \sum_{j=0}^{\infty} \binom{i+j-1}{j} \left\{\theta_1 \hat{f}_1(\omega_2 \kappa_2)\right\}^j \hat{P}_{0,0,0}(s). \tag{3.35}$$

Using equations (3.34) and (3.35) in equation (3.31), we get

$$\hat{P}_{n,2,1}(s) = \hat{\psi}_1(s)\, {\nu_2}^{n-1} \hat{f}_n(\omega_2 \kappa_2) \hat{P}_{0,0,0}(s). \tag{3.36}$$

Similarly, using equation (3.32), we obtain

$$\hat{P}_{n,1,2}(s) = \hat{\psi}_2(s)\, {\nu_3}^{n-1} \hat{f}_n(\omega_3 \kappa_3) \hat{P}_{0,0,0}(s), \tag{3.37}$$

where $\omega_2 = s + \tau_2$, $\omega_3 = s + \tau_3$,

$$\hat{\psi}_1(s) = \frac{\lambda\gamma_1}{(s+\lambda+\zeta)} \left[1 + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \binom{i+j-1}{j} \left\{\frac{\lambda\mu_2}{(s+\lambda+\zeta)\theta_1}\right\}^i \left\{\theta_1 \hat{f}_1(\omega_2 \kappa_2)\right\}^{i+j}\right] \times \left[1 + \sum_{k=0}^{\infty} \left\{\theta_1 \hat{f}_1(\omega_2 \kappa_2)\right\}^{k+1}\right],$$

and

$$\hat{\psi}_2(s) = \frac{\lambda\gamma_2}{(s+\lambda+\zeta)}\left[1+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j-1}{j}\left\{\frac{\lambda\mu_1}{(s+\lambda+\zeta)\theta_2}\right\}^i\left\{\theta_2\hat{f}_1(\omega_3\kappa_3)\right\}^{i+j}\right] \times\left[1+\sum_{k=0}^{\infty}\left\{\theta_2\hat{f}_1(\omega_3\kappa_3)\right\}^{k+1}\right].$$

On inverting

$$P_{n,2,1}(t) = \int_0^t\left[\int_0^u \psi_1(v){\nu_2}^{n-1}f_n((u-v)\kappa_2)\exp(-\tau_2(u-v))\,\mathrm{d}v\right]P_{0,0,0}(t-u)\,\mathrm{d}u, \tag{3.38}$$

$$P_{n,1,2}(t) = \int_0^t\left[\int_0^u \psi_2(v){\nu_3}^{n-1}f_n((u-v)\kappa_3)\exp(-\tau_3(u-v))\,\mathrm{d}v\right]P_{0,0,0}(t-u)\,\mathrm{d}u, \tag{3.39}$$

where

$$\psi_1(v) = \lambda\gamma_1\int_0^v\left[\exp\{-(\lambda+\zeta)x\}+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j-1}{j}\left(\frac{\lambda\mu_2}{\theta_1}\right)^i \times\int_0^x \exp\{-(\lambda+\zeta)y\}^{*(i+1)}\times\{\theta_1 f_1((x-y)\kappa_2)\exp(-\tau_2(x-y))\}^{*(i+j)}\,\mathrm{d}y\right] \times\left[\delta(v-x)+\sum_{k=0}^{\infty}\{\theta_1 f_1((v-x)\kappa_2)\exp(-\tau_2(v-x))\}^{*(k+1)}\right]\mathrm{d}x,$$

and

$$\psi_2(v) = \lambda\gamma_2\int_0^v\left[\exp\{-(\lambda+\zeta)x\}+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j-1}{j}\left(\frac{\lambda\mu_1}{\theta_2}\right)^i \times\int_0^x \exp\{-(\lambda+\zeta)y\}^{*(i+1)}\{\theta_2 f_1((x-y)\kappa_3)\exp(-\tau_3(x-y))\}^{*(i+j)}\,\mathrm{d}y\right] \times\left[\delta(v-x)+\sum_{k=0}^{\infty}\{\theta_2 f_1((v-x)\kappa_3)\exp(-\tau_3(u-x))\}^{*(k+1)}\right]\mathrm{d}x,$$

where, δ and $*^k$ denote Dirac delta function and k-fold convolution respectively. Thus we have expressed the probabilities $P_{n,2,1}(t)$ and $P_{n,1,2}(t)$ interms of $P_{0,0,0}(t)$. Using equations (3.36) and (3.37) in equation (3.18), we get

$$\hat{P}_{2,1,1}(s) = \sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(-1)^j\frac{1}{{\nu_1}^i}\frac{1}{\mu^j}\left[\left\{1+\frac{\eta\zeta}{\zeta+\eta}\left(\frac{1}{s}-\frac{1}{s+\zeta+\eta}\right)\left(1-\hat{h}_1(s)\right)-\hat{h}_1(s) - \hat{P}_{0,0,0}(s)(\gamma_1+\gamma_2)\right\}\hat{f}_{i+1}(\omega_1\kappa_1)+\sum_{m=0}^{\infty}\left\{\theta_1\hat{\psi}_1(s)\,{\nu_2}^{m+1}\hat{f}_{m+2}(\omega_2\kappa_2)\right.\right.$$

$$+ \theta_2\hat{\psi}_2(s)\,{\nu_3}^{m+1}\hat{f}_{m+2}(\omega_3\kappa_3)\Bigg\}\frac{1}{{\nu_1}^m}\hat{f}_{m+i+1}(\omega_1\kappa_1)\hat{P}_{0,0,0}(s)\Bigg]\Big(\mu_2\hat{h}_2(s)+\mu_1\hat{h}_3(s)\Big)^j. \tag{3.40}$$

Inverting the above yields

$$P_{2,1,1}(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(-1)^j\frac{1}{{\nu_1}^i}\frac{1}{\mu^j}\int_0^t\Bigg[\Bigg\{\delta(u)+\frac{\eta\zeta}{\zeta+\eta}\{1-\exp(-(\zeta+\eta)u)\}*(\delta(u)-h_1(u))$$

$$-h_1(u)-(\gamma_1+\gamma_2)P_{0,0,0}(u)\Bigg\}*f_{i+1}(\kappa_1 u)\exp(-\tau_1 u)+\sum_{m=0}^{\infty}\Bigg\{\theta_1\psi_1(u)*{\nu_2}^{m+1}f_{m+2}(\kappa_2 u)$$

$$\times\exp(-\tau_2 u)+\theta_2\psi_2(u)*{\nu_3}^{m+1}f_{m+2}(\kappa_3 u)\exp(-\tau_3 u)\Bigg\}*\frac{1}{{\nu_1}^m}f_{m+i+1}(\kappa_1 u)\exp(-\tau_1 u)$$

$$*P_{0,0,0}(u)\Bigg](\mu_2 h_2(t-u)+\mu_1 h_3(t-u))^{*j}\,\mathrm{d}u.$$

The probability $P_{2,1,1}(t)$ is expressed interms of $P_{0,0,0}(t)$. Using equation (3.40) in equation (3.20) and inverting, we get

$$P_{0,0,0}(t) = \Bigg\{\delta(t)+\frac{\eta\zeta}{\zeta+\eta}(1-\exp\{-(\zeta+\eta)t\})\Bigg\}*K_{11}(t)+\{\mu_2K_{12}(t)+\mu_1K_{13}(t)\}$$

$$*\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}\frac{(-1)^j}{{\nu_1}^i\mu^j}\Bigg\{\delta(t)+\frac{\eta\zeta}{\zeta+\eta}(1-\exp\{-(\zeta+\eta)t\})*(\delta(t)-h_1(t))-h_1(t)\Bigg\}$$

$$*f_{i+1}(t\kappa_1)\exp(-\tau_1 t)*(\mu_2h_2(t)+\mu_1h_3(t))^{*j}*\sum_{n=0}^{\infty}\Bigg[-(\gamma_1+\gamma_2)f_{i+1}(t\kappa_1)\exp(-\tau_1 t)$$

$$*\{\mu_2K_{12}(t)+\mu_1K_{13}(t)\}*\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}(-1)^j\frac{1}{{\nu_1}^i}\frac{1}{\mu^j}(\mu_2h_2(t)+\mu_1h_3(t))^{*j}$$

$$+\sum_{m=0}^{\infty}\Bigg\{\theta_1\psi_1(t){\nu_2}^{m+1}*f_{m+2}(t\kappa_2)\exp(-\tau_2 t)+\theta_2\psi_2(t){\nu_3}^{m+1}*f_{m+2}(t\kappa_3)\exp(-\tau_3 t)\Bigg\}$$

$$*\frac{1}{{\nu_1}^m}f_{m+i+1}(\kappa_1 t)\exp(-\tau_1 t)\Bigg]^{*n}.$$

3.6. Special cases

For $\theta_1=0$, $\theta_2=0$, $\zeta=0$, $\gamma_1=0$, $\gamma_2=0$, $\kappa_1=\alpha$, $\nu_1=\beta$, equation (3.5) reduces to

$$P_{n+2,1,1}(t)=\lambda\int_0^t {\nu_1}^{n-1}f_n(\kappa_1(t-w))P_{2,1,1}(w)\,\mathrm{d}w.$$

Using (3.6), the above equation becomes

$$P_{n+2,1,1}=n\beta^n\int_0^t\exp\{-(\lambda+\mu)(t-y)\}\frac{I_n(\alpha(t-y))}{t-y}P_{2,1,1}\,\mathrm{d}y, \tag{3.41}$$

and equation (3.18) reduces to

$$\hat{P}_{2,1,1}(s) = \frac{2}{\kappa_1}\sum_{i=0}^{\infty}\sum_{j=0}^{i}(-1)^j\frac{1}{\nu^i}\left(\frac{b_2^*(s)}{\mu}\right)^j(1-h_1^*(s))\left(\frac{\omega_1-\sqrt{\omega^2-{\kappa_1}^2}}{\kappa_1}\right)^{i+1}.$$

On inversion, we get

$$P_{2,1,1}(t) = \frac{2}{\alpha}\sum_{i=0}^{\infty}\sum_{j=0}^{i}(-1)^j\frac{1}{\beta^i}\frac{1}{\mu^j}\int_0^t b_2^{*^j}(t-u)[\exp\{-(\lambda+\mu)w\}\{I_i(\alpha w)-I_{i+2}(\alpha w)\}$$
$$-\left.\int_0^w h_1(w-u)\exp\{-(\lambda+\mu)u\}\{I_i(\alpha u)-I_{i+2}(\alpha u)\}\,\mathrm{d}u\right]\mathrm{d}w. \tag{3.42}$$

The results (3.41) and (3.42) coincides with equations (3.40) and (3.41) respectively of Dharmaraja [4].

4. Performance measures

The time dependent performance measures such as expected system size, variance, number of busy servers at a given time, expected busy servers and probability that an arriving job joins the queue are studied in this section.

4.1. Expected system size

Let $M(t)$ denote the expected system size at time t. Then,

$$M(t) = P_{1,0,1}(t) + P_{1,1,0}(t) + \sum_{n=1}^{\infty} nP_{n,2,1}(t) + \sum_{n=1}^{\infty} nP_{n,1,2}(t) + \sum_{n=1}^{\infty}(n+1)P_{(n+1),1,1}(t).$$

To facilitate the substitution of the Kolmogorov forward equations (2.1)–(2.12), we first differentiate the expression for $M(t)$

$$M'(t) = \lambda(1-P_F(t)) - \sum_{n=1}^{\infty}\{\mu P_{n+1,1,1}(t) + \mu_1 P_{n,1,2}(t) + \mu_2 P_{n,2,1}(t)\} - \mu_1 P_{1,1,0}(t) - \mu_2 P_{1,0,1}(t) - \zeta M(t).$$

On integrating, we get

$$M(t) = \frac{\lambda}{\zeta}(1-\exp\{-\zeta t\}) - \lambda\int_0^t P_F(u)\exp\{-\zeta(t-u)\}\,\mathrm{d}u$$
$$-\int_0^t\{\mu_1 P_{1,1,0}(u) + \mu_2 P_{1,0,1}(u)\}\exp\{-\zeta(t-u)\}\,\mathrm{d}u$$
$$-\sum_{n=1}^{\infty}\int_0^t\{\mu P_{n+1,1,1}(u) + \mu_1 P_{n,1,2}(u) + \mu_2 P_{n,2,1}(u)\}\exp\{-\zeta(t-u)\}\,\mathrm{d}u.$$

4.2. Variance of the system

The variance of the system at time t is given by

$$V(N(t)) = E\big(N^2(t)\big) - [E(N(t))]^2,$$

where

$$Q(t) = E\big(N^2(t)\big) = P_{1,0,1}(t) + P_{1,1,0}(t) + \sum_{n=1}^{\infty} n^2 P_{n,2,1}(t) + \sum_{n=1}^{\infty} n^2 P_{n,1,2}(t) + \sum_{n=1}^{\infty} (n+1)^2 P_{(n+1),1,1}(t).$$

Differentiating the above

$$Q'(t) = 2\lambda E[X(t)] + \lambda(1 - P_F(t)) - \mu_1 P_{1,1,0}(t) - \mu_2 P_{1,0,1}(t) - \zeta Q(t)$$
$$- \sum_{n=1}^{\infty} \{\mu(2n+1)P_{n+1,1,1}(t) + \mu_1(2n-1)P_{n,1,2}(t) + \mu_2(2n-1)P_{n,2,1}(t)\}.$$

On integration

$$Q(t) = \frac{\lambda}{\zeta}\{1 - \exp(-\zeta t)\} + \int_0^t \{\lambda P_F(u) - 2\lambda M(u)\} \exp\{-\zeta(t-u)\}\,\mathrm{d}u$$
$$- \sum_{n=1}^{\infty} \int_0^t \{\mu(2n+1)P_{n+1,1,1}(u) + \mu_1(2n-1)P_{n,1,2}(u) + \mu_2(2n-1)P_{n,2,1}(u)\}$$
$$\times \exp\{-\zeta(t-u)\}\,\mathrm{d}u - \int_0^t \{\mu_1 P_{1,1,0}(u) + \mu_2 P_{1,0,1}(u)\} \exp\{-\zeta(t-u)\}\,\mathrm{d}u.$$

4.3. Busy servers

Let $B(t)$ represents the number of busy servers at time t. The probability that the system has m busy servers is given by

$$P(B(t) = m) = \begin{cases} P_{1,0,1}(t) + P_{1,1,0}(t) + \sum_{n=1}^{\infty} P_{n,2,1}(t) + \sum_{n=1}^{\infty} P_{n,1,2}(t), m = 1 \\ \sum_{n=0}^{\infty} P_{(n+2),1,1}, m = 2. \end{cases}$$

4.4. Expected busy servers

Let $E\{B(t)\}$ denote the expected number of busy servers at time t. Then

$$E\{B(t)\} = P_{1,0,1}(t) + P_{1,1,0}(t) + \sum_{n=1}^{\infty} P_{n,2,1}(t) + \sum_{n=1}^{\infty} P_{n,1,2}(t) + 2\sum_{n=0}^{\infty} P_{n+2,1,1}(t).$$

4.5. Probability that an arriving job joining the queue

The probability that an arriving customer joins the queue is given by

$$P(N(t) \geq 2) = P_{2,1,1}(t) + \lambda \sum_{n=0}^{\infty} \Bigg[\int_0^t {\nu_1}^{n-1} f_n(\kappa_1(t-w)) P_{2,1,1}(w)\,\mathrm{d}w$$

$$+\int_0^t \left(\sum_{m=1}^{\infty}\theta_1 P_{m+2,2,1}(w)+\theta_2\sum_{m=1}^{\infty}P_{m+2,1,2}(w)\right){\nu_1}^{n-m}g_{n,m}(\kappa_1(t-w))\,\mathrm{d}w\Bigg].$$

5. Steady state analysis

Let $\pi_{0,0,0}$, $\pi_{n,1,1}$, $\pi_{n,2,1}$, $\pi_{n,1,2}$ denote the steady state probability of the idle state, busy state, vacation state of the fast sever and vacation state of the slow server respectively. The governing equations of the proposed model in the steady state is given by

$$0=-(\lambda+\gamma_1+\gamma_2+\zeta)\pi_{0,0,0}+\mu_1\pi_{1,1,0}+\mu_2\pi_{1,0,1}+\eta\pi_F, \tag{5.1}$$
$$0=-(\lambda+\mu_1+\zeta)\pi_{1,1,0}+\lambda\pi_{0,0,0}+\mu_2\pi_{2,1,1}, \tag{5.2}$$
$$0=-(\lambda+\mu_2+\zeta)\pi_{1,0,1}+\mu_1\pi_{2,1,1}, \tag{5.3}$$
$$0=-(\lambda+\mu+\zeta)\pi_{2,1,1}+\lambda\pi_{1,1,0}+\lambda\pi_{1,0,1}+\mu\pi_{3,1,1}+\theta_1\pi_{2,2,1}+\theta_2\pi_{2,1,2}, \tag{5.4}$$
$$0=-(\lambda+\mu+\zeta)\pi_{n,1,1}+\lambda\pi_{n-1,1,1}+\mu\pi_{n+1,1,1}+\theta_1\pi_{n,2,1}+\theta_2\pi_{n,1,2},\quad n=3,4,5,\dots \tag{5.5}$$
$$0=-(\lambda+\zeta)\pi_{0,2,0}+\gamma_1\pi_{0,0,0}+\mu_2\pi_{1,2,1}, \tag{5.6}$$
$$0=-(\lambda+\mu_2+\zeta)\pi_{1,2,1}+\lambda\pi_{0,2,0}+\mu_2\pi_{2,2,1}, \tag{5.7}$$
$$0=-(\lambda+\mu_2+\theta_1+\zeta)\pi_{n,2,1}+\lambda\pi_{n-1,2,1}+\mu_2\pi_{n+1,2,1},\quad n=2,3,4,\dots, \tag{5.8}$$
$$0=-(\lambda+\zeta)\pi_{0,0,2}+\gamma_2\pi_{0,0,0}+\mu_1\pi_{1,1,2}, \tag{5.9}$$
$$0=-(\lambda+\mu_1+\zeta)\pi_{1,1,2}+\lambda\pi_{0,0,2}+\mu_1\pi_{2,1,2}, \tag{5.10}$$
$$0=-(\lambda+\mu_1+\theta_2+\zeta)\pi_{n,1,2}(t)+\lambda\pi_{n-1,1,2}+\mu_1\pi_{n+1,1,2},\quad n=2,3,4,\dots, \tag{5.11}$$
$$0=\zeta(1-\pi_F)-\eta\pi_F. \tag{5.12}$$

Let

$$\beta=\frac{1}{\lambda+\mu+\zeta},\beta_1=\frac{1}{\lambda+\mu_1+\zeta}\ \text{ and }\ \beta_2=\frac{1}{\lambda+\mu_2+\zeta}\cdot$$

We define generating functions as follows:

$$H_1(z,n)=\sum_{n=1}^{\infty}\pi_{n,2,1}z^n,\ H_2(z,n)=\sum_{n=1}^{\infty}\pi_{n,1,2}z^n\ \text{ and }\ H_3(z,n)=B+\sum_{n=0}^{\infty}\pi_{n+3,1,1}z^{n+1},$$

where

$$B=\pi_{0,0,0}+\pi_{1,1,0}+\pi_{1,0,1}+\pi_{2,1,1}.$$

Using equations (5.8) and (5.9), multiplying suitable powers of z, we obtain

$$H_1(z,n)\big\{\lambda z^2-(\lambda+\mu_2+\theta_1+\zeta)z+\mu_2\big\}=(\mu_2-\theta_1)z\pi_{1,2,1}-\lambda z^2\pi_{0,2,0},$$

leading to

$$H_1(z,n)=\frac{(\mu_2-\theta_1)z\pi_{1,2,1}-\lambda z^2\pi_{0,2,0}}{(z-\alpha_1)(z-\bar{\alpha}_1)}\cdot \tag{5.13}$$

Let $\alpha_1>1$ and $0<\bar{\alpha}_1<1$ be two distinct roots of the denominator of equation (5.13). We have

$$\alpha_1,\bar{\alpha}_1=\frac{\lambda+\mu_2+\theta_1+\zeta\pm\sqrt{(\lambda+\mu_2+\theta_1+\zeta)^2-4\lambda\mu_2}}{2\lambda}\cdot$$

Setting $z=\bar{\alpha}_1$ in equation (5.13) and comparing the coefficient of z^n, we get

$$\pi_{n,2,1}=\frac{\lambda}{(\alpha_1)^n}\pi_{0,2,0},\quad n=1,2,3,\dots \tag{5.14}$$

Similarly, using equations (5.10) and (5.11), we get

$$H_2(z,n) = \frac{(\mu_1-\theta_2)z\pi_{1,1,2} - \lambda z^2\pi_{0,0,2}}{(z-\alpha_2)(z-\bar{\alpha}_2)}\cdot$$

Setting $z=\bar{\alpha}_2$ and equating the coefficient of z^n, we obtain

$$\pi_{n,1,2} = \frac{\lambda}{(\alpha_2)^n}\pi_{0,0,2}, \quad n=1,2,3,\ldots \tag{5.15}$$

where

$$\alpha_2, \bar{\alpha}_2 = \frac{\lambda+\mu_1+\theta_2+\zeta \pm \sqrt{(\lambda+\mu_1+\theta_2+\zeta)^2 - 4\lambda\mu_1}}{2\lambda}\cdot$$

Setting $n=1$ in the results (5.14) and (5.15) using it in equations (5.5) and (5.9) respectively, we get

$$\pi_{0,2,0} = \chi_1\pi_{0,0,0} \tag{5.16}$$

$$\pi_{0,0,2} = \chi_2\pi_{0,0,0}, \tag{5.17}$$

where

$$\chi_1 = \frac{\gamma_1\alpha_1}{(\lambda+\zeta)\alpha_1 - \lambda\mu_2}, \quad \left|\frac{\lambda\mu_2}{(\lambda+\zeta)\alpha_1}\right| < 1,$$

and

$$\chi_2 = \frac{\gamma_2\alpha_2}{(\lambda+\zeta)\alpha_2 - \lambda\mu_1}, \quad \left|\frac{\lambda\mu_1}{(\lambda+\zeta)\alpha_2}\right| < 1.$$

Substituting equations (5.16) and (5.17) in (5.14) and (5.15) respectively, we obtain

$$\pi_{n,2,1} = \frac{\lambda\chi_1}{{\alpha_1}^n}\pi_{0,0,0}, \quad n=1,2,3,\ldots \tag{5.18}$$

$$\pi_{n,1,2} = \frac{\lambda\chi_2}{{\alpha_2}^n}\pi_{0,0,0}, \quad n=1,2,3,\ldots \tag{5.19}$$

Thus, the vacation state probabilities $\pi_{n,2.1}$ and $\pi_{n,1,2}$ are expressed interms of $\pi_{0,0,0}$. Multiplying suitable powers of z in equation (5.5), after some manipulation, we get

$$G_3(z)(z-\alpha_3)(z-\bar{\alpha}_3) = z\lambda\pi_{2,1,1}(1-z) - \eta\pi_F - \theta_1\left\{\pi_{2,2,1} + \sum_{n=1}^{\infty}\pi_{n+2,2,1}z^n\right\}$$
$$- \theta_2\left\{\pi_{2,1,2} + \sum_{n=1}^{\infty}\pi_{n+2,1,2}z^n\right\} + B\left\{-(\lambda+\mu) + \lambda z + \frac{\mu}{z}\right\} + \pi_{0,0,0}\{\gamma_1+\gamma_2\},$$

leading to

$$B + \sum_{n=0}^{\infty}\pi_{n+3,1,1}z^{n+1} = z\frac{\lambda}{(\alpha_3 - z)}\pi_{2,1,1} + \frac{\theta_1}{(\alpha_3-z)}\sum_{n=1}^{\infty}\pi_{n+2,2,1}\left(\frac{z^n - \bar{\alpha}_3^n}{z-\bar{\alpha}_3}\right) + \frac{\theta_2}{(\alpha_3-z)}$$
$$\times\sum_{n=1}^{\infty}\pi_{n+2,1,2}\left(\frac{z^n-\bar{\alpha}_3^n}{z-\bar{\alpha}_3}\right) + \frac{B}{(\alpha_3-z)}\left\{-\lambda + \frac{\mu}{z\bar{\alpha}_3}\right\}.$$

Comparing the coefficient of z^n, $n=0,1,2,\ldots$

$$\pi_{n+3,1,1} = \lambda\left(\frac{1}{\alpha_3}\right)^{n+1}\pi_{2,1,1} + \{\theta_1\lambda\chi_1 Q_n(\alpha_1,\alpha_3,\bar{\alpha}_3) + \theta_2\lambda\chi_2 Q_n(\alpha_2,\alpha_3,\bar{\alpha}_3)\}\pi_{0,0,0} \tag{5.20}$$

where

$$Q_n(\alpha_i,\alpha_3,\bar{\alpha}_3)=\left[\frac{1-\left(\frac{\bar{\alpha}_3}{\alpha_3}\right)^{n+1}}{\left(1-\frac{\bar{\alpha}_3}{\alpha_3}\right)}-\left\{\frac{1-\left(\frac{\bar{\alpha}_3}{\alpha_3}\right)^{n}}{1-\left(\frac{\bar{\alpha}_3}{\alpha_3}\right)}-\frac{\left(\frac{\bar{\alpha}_3}{\alpha_1}\right)\left(\frac{\bar{\alpha}_3}{\alpha_3}\right)^{n}-\left(\frac{\bar{\alpha}_3}{\alpha_i}\right)^{n+1}}{\frac{\bar{\alpha}_3}{\alpha_3}-\frac{\bar{\alpha}_3}{\alpha_i}}\right\}\right]$$
$$\times\frac{\left(\frac{1}{\bar{\alpha}_3}\right)^{n+2}\left(\frac{1}{\alpha_i}\right)^{2}\frac{\bar{\alpha}_3^2}{\alpha_i\alpha_3}}{1-\frac{\bar{\alpha}_3}{\alpha_i}},\quad i=1,2.$$

and

$$\alpha_3,\bar{\alpha}_3=\frac{\lambda+\mu+\zeta\pm\sqrt{(\lambda+\mu+\zeta)^2-4\lambda\mu}}{2\lambda}.$$

Thus we have obtained the busy state probability $\pi_{n+3,1,1}$ in terms of $\pi_{0,0,0}$.

Using equations (5.2), (5.3) and (5.20) in equation (5.4), after some algebra, we get

$$\pi_{2,1,1}=\left[\frac{\lambda\beta_1+\{\theta_1\chi_1Q_0(\alpha_1,\alpha_3,\bar{\alpha}_3)+\theta_2\chi_2Q_0(\alpha_2,\alpha_3,\bar{\alpha}_3)\}\mu+\frac{\chi_1\theta_1}{\alpha_1{}^2}+\frac{\chi_2\theta_2}{\alpha_2{}^2}}{1-\lambda\beta\left(\mu_2\beta_1+\mu_1\beta_2+\frac{\mu}{\alpha_3}\right)}\right]\lambda\beta\pi_{0,0,0}. \tag{5.21}$$

Using the result (5.21) in equations (5.2) and (5.3), we get

$$\pi_{1,1,0}=\left[\frac{\lambda\beta_1+\{\theta_1\chi_1Q_0(\alpha_1,\alpha_3,\bar{\alpha}_3)+\theta_2\chi_2Q_0(\alpha_2,\alpha_3,\bar{\alpha}_3)\}\mu+\frac{\chi_1\theta_1}{\alpha_1{}^2}+\frac{\chi_2\theta_2}{\alpha_2{}^2}}{1-\lambda\beta\left(\mu_2\beta_1+\mu_1\beta_2+\frac{\mu}{\alpha_3}\right)}\right]\lambda\beta\pi_{0,0,0}\times\left\{\mu_2\beta_1+\lambda\beta_1\right\}, \tag{5.22}$$

and

$$\pi_{1,0,1}=\left[\frac{\lambda\beta_1+\{\theta_1\chi_1Q_0(\alpha_1,\alpha_3,\bar{\alpha}_3)+\theta_2\chi_2Q_0(\alpha_2,\alpha_3,\bar{\alpha}_3)\}\mu+\frac{\chi_1\theta_1}{\alpha_1{}^2}+\frac{\chi_2\theta_2}{\alpha_2{}^2}}{1-\lambda\beta\left(\mu_2\beta_1+\mu_1\beta_2+\frac{\mu}{\alpha_3}\right)}\right]\lambda\beta\mu_1\beta_2\pi_{0,0,0}. \tag{5.23}$$

The normalisation condition is given by

$$B+\sum_{n=0}^{\infty}\pi_{n+3,1,1}+\pi_{0,2,0}+\sum_{n=1}^{\infty}\pi_{n,2,1}+\pi_{0,0,2}+\sum_{n=1}^{\infty}\pi_{n,1,2}=1-\pi_F.$$

Using equations (5.18)–(5.23) in the normalisation condition, we get

$$\pi_{0,0,0}=\frac{1-\frac{\zeta}{\zeta+\eta}}{D}, \tag{5.24}$$

where

$$D=\lambda\beta_1+\lambda\beta\left\{\frac{\lambda\beta_1+\{\theta_1\chi_1Q_0(\alpha_1,\alpha_3,\bar{\alpha}_3)+\theta_2\chi_2Q_0(\alpha_2,\alpha_3,\bar{\alpha}_3)\}\mu+\frac{\chi_1\theta_1}{\alpha_1{}^2}+\frac{\chi_2\theta_2}{\alpha_2{}^2}}{1-\lambda\beta\left(\mu_2\beta_1+\mu_1\beta_2+\frac{\mu}{\alpha_3}\right)}\right\}$$
$$\times\left\{1+\mu_2\beta_1+\mu_1\beta_2+\frac{\lambda}{\alpha_3-1}\right\}+\lambda\sum_{n=0}^{\infty}\{\theta_1\chi_1Q_n(\alpha_1,\alpha_3,\bar{\alpha}_3)+\theta_2\chi_2Q_n(\alpha_2,\alpha_3,\bar{\alpha}_3)\}$$
$$+\chi_1+\frac{\lambda\chi_1}{\alpha_1-1}+\chi_2+\frac{\lambda\chi_2}{\alpha_2-1},$$

$$Q_0(\alpha_i,\alpha_3,\bar{\alpha}_3)=\frac{1}{(\alpha_1-\bar{\alpha}_3)(\alpha_i)^2\alpha_3},\quad i=1,2.$$

6. Performance measures in the steady state

The performance measures of the system in the steady state are discussed in this section.

6.1. Expected system size

Let $E(N_s)$ denote the expected system size, then

$$E(N_s) = \pi_{1,1,0} + \pi_{1,0,1} + \sum_{n=1}^{\infty} n\pi_{n,2,1} + \sum_{n=1}^{\infty} n\pi_{n,1,2} + \sum_{n=0}^{\infty} (n+2)\pi_{n+2,1,1}.$$

Using Little's formula, the expected number of jobs waiting in the system is given by

$$E(W_s) = \frac{E(N_s)}{\lambda}.$$

6.2. Expected number of jobs served

Let $E(J_s)$ denote the expected number of jobs served, then

$$E(J_s) = \mu_1\pi_{1,1,0} + \mu_2\pi_{1,0,1} + \mu_2\sum_{n=1}^{\infty} \pi_{n,2,1} + \mu_1\sum_{n=1}^{\infty} \pi_{n,1,2} + \mu\sum_{n=0}^{\infty} \pi_{n+2,1,1}.$$

6.3. Busy servers

Let π_{b_i} represents the probability that the server is in the busy state. Then

$$\pi_{b_i} = \begin{cases} \pi_{1,1,0} + \sum_{n=1}^{\infty} \pi_{n,1,2}, & i=1 \quad \text{(FS is busy)} \\ \pi_{1,0,1} + \sum_{n=1}^{\infty} \pi_{n,2,1}, & i=2 \quad \text{(SS is busy)} \\ \pi_{2,1,1} + \sum_{n=0}^{\infty} \pi_{n+3,1,1}, & i=3 \quad \text{(FS and SS are busy)}, \end{cases}$$

where

$$\pi_{b_1} = \mu_2\beta_1\pi_{2,1,1} + \left(\lambda\beta_1 + \frac{\lambda\chi_2}{\alpha_2 - 1}\right)\pi_{0,0,0},$$
$$\pi_{b_2} = \mu_1\beta_2\pi_{2,1,1} + \frac{\lambda\chi_1}{\alpha_1 - 1}\pi_{0,0,0},$$

and

$$\pi_{b_3} = \left[\frac{\alpha_3}{\alpha_3 - 1} + \sum_{n=0}^{\infty} \{\theta_1\chi_1 Q_n(\alpha_1, \alpha_3, \bar{\alpha}_3) + \theta_2\chi_2 Q_n(\alpha_2, \alpha_3, \bar{\alpha}_3)\}\right]\lambda\pi_{0,0,0} + \pi_{2,1,1}.$$

6.4. Expected busy servers

Let $E\{b_s\}$ denote the expected number of busy servers. Then

$$E\{b_s\} = \pi_{1,0,1} + \pi_{1,1,0} + \sum_{n=1}^{\infty} \pi_{n,2,1} + \sum_{n=1}^{\infty} \pi_{n,1,2} + 2\sum_{n=0}^{\infty} \pi_{n+2,1,1}.$$

Using equations (5.18)–(5.23), we obtain

$$E\{b_s\} = (\mu_1\beta_2 + \mu_2\beta_1 + 2)\pi_{2,1,1} + \left(\beta_1 + \frac{\chi_1}{\alpha_1 - 1} + \frac{\chi_2}{\alpha_2 - 1}\right)\lambda\pi_{0,0,0}$$
$$+ 2\left[\frac{\alpha_3}{\alpha_3 - 1} + \sum_{n=0}^{\infty}\{\theta_1\chi_1 Q_n(\alpha_1, \alpha_3, \bar{\alpha}_3) + \theta_2\chi_2 Q_n(\alpha_2, \alpha_3, \bar{\alpha}_3)\}\right]\lambda\pi_{0,0,0}.$$

6.5. Servers vacation

Let π_{v_1} and π_{v_2} represents probability that the FS and SS are in vacation respectively, then

$$\pi_{v_i} = \begin{cases} \pi_{0,2.0} + \sum_{n=0}^{\infty}\pi_{n,2.1}, i = 1 \\ \pi_{0,0.2} + \sum_{n=0}^{\infty}\pi_{n,1.2}, i = 2 \end{cases}$$

where

$$\pi_{v_1} = \left(\chi_1 + \frac{\lambda\chi_1}{\alpha_1 - 1}\right)\pi_{0,0,0} \text{ and } \pi_{v_2} = \left(\chi_2 + \frac{\lambda\chi_2}{\alpha_2 - 1}\right)\pi_{0,0,0}.$$

7. Cost analysis

Cost and revenue analysis is crucial in studying queuing systems, offering valuable insights into their economic implications. Our study aims to identify optimal service rates to minimize the total expected cost function, enhancing the efficiency and economic viability of queuing systems. We carefully consider various cost parameters throughout the analysis, recognizing their significance in shaping the system's overall economic interpretation. The cost parameters are as follows:

- C_1: Cost per job served by the mean service rate μ_1.
- C_2: Cost per job served by the mean service rate μ_2.
- C_h: Holding cost per unit time.
- C_b: Cost incurred per unit time when both the servers are busy.
- C_I: Cost incurred per unit time when the FS and SS are idle.
- C_{v_1}: Cost incurred per unit time during the vacation period of the FS.
- C_{v_2}: Cost incurred per unit time during the vacation period of the SS.
- C_R: Cost incurred per unit time for repair.

Using the aforementioned cost parameters and associated system performance metrics, the Total expected cost (TEC) is given by

$$\text{TEC} = F(\mu_1, \mu_2) = C_1\mu_1 + C_2\mu_2 + C_h E(N_s) + C_I\pi_{0,0,0} + C_b\pi_{b_3} + C_{v_1}\pi_{b_2} + C_{v_2}\pi_{b_1} + C_R\pi_F.$$

It is important to highlight that the cost function seems to be an intricate and nonlinear expression. This is due to the fact that there are many parameters such as λ, μ_1, μ_2, θ_1, θ_2, γ_1, γ_2, and ζ. Our goal is to ascertain the optimal service rate of the FS, denoted as μ_1^*, and the optimal service rate of the SS, denoted as μ_2^*, to minimize the TEC function. The mathematical representation of the cost minimization problem is as follows:

$$F(\mu_1^*, \mu_2^*) = \min F(\mu_1, \mu_2),$$
$$\text{subject to: } \mu_1 > \mu_2, \left|\frac{\lambda\mu_2}{(\lambda+\zeta)\alpha_1}\right| < 1 \text{ and } \left|\frac{\lambda\mu_1}{(\lambda+\zeta)\alpha_2}\right| < 1.$$

Algorithm 1. PSO for Cost Optimization with $\mu_1 > \mu_2$.

Initialize particle positions (μ_1^i, μ_2^i) and velocities randomly for $i = 1$ to 5 such that $\mu_1^i > \mu_2^i$
Evaluate the cost function for each particle
Set personal best $pBest_i$ and global best $gBest$
for iteration $t = 1$ to 40 **do**
  for each particle $i = 1$ to 5 **do**
    Update velocity:

$$v^i = w \cdot v^i + c_1 \cdot \text{rand}() \cdot (pBest_i - \text{position}_i) + c_2 \cdot \text{rand}() \cdot (gBest - \text{position}_i)$$

    Update position:

$$\text{position}_i = \text{position}_i + v^i$$

    Enforce bounds and ensure $\mu_1^i > \mu_2^i$
    Evaluate cost at new position
    if new cost < cost at $pBest_i$ **then**
      $pBest_i \leftarrow$ new position
    end if
    if new cost < cost at $gBest$ **then**
      $gBest \leftarrow$ new position
    end if
  end for
end for
return optimal $(\mu_1^*, \mu_2^*) = gBest$, and minimum cost

Consider R as the revenue generated from delivering services to a customer, then total expected revenue (TER) is given by

$$\text{TER} = R \times E(N_s).$$

In this study, the Particle Swarm Optimization (PSO) algorithm is employed to solve the cost optimization problem. Due to the intricate nature of the cost function, explicit expressions for the optimal service rates (μ_1^*, μ_2^*) are challenging to derive. The objective function is nonlinear and involves multiple performance measures such as expected system size and server utilization, making analytical based methods impractical. PSO, a population-based, derivative-free metaheuristic, is well-suited for this problem. It efficiently explores complex, high-dimensional spaces, converging to near-optimal solutions with reasonable computational effort, and has proven successful for similar queueing optimizations in the literature. We therefore adopt PSO as a practical, efficient optimization method to find the service rates minimizing TEC.

8. Numerical illustration

In this section, numerical findings are showcased through graphs and tables, providing a visual representation of how different parameters impact the system's performance measures. The values of the parameters are chosen as follows: $\lambda = 0.7$, $\mu_1 = 1.2$, $\mu_2 = 1$, $\gamma_1 = 0.2$, $\gamma_2 = 0.3$, $\theta_1 = 0.2$, $\theta_2 = 0.1$, $\eta = 0.1$, and $\zeta = 0.01$. The cost parameters are selected as follows: $C_1 = 100$, $C_2 = 75$, $C_h = 50$, $C_I = 20$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 40$, and $C_F = 30$. The following values are chosen for the PSO algorithm: the number of particles is 5, the maximum number of iterations is 40, the cognitive coefficient c_1 is set to 1.5, the social coefficient c_2 is set to 1.5, and the initial inertia weight w is set to 0.7

Figure 2 illustrates probability curves depicting the system's behaviour under different scenarios: both servers being idle, and one being idle while the other is busy. The figure reveals that, over time, the probability consistently rises until reaching a steady state. When the FS is in operation and only one customer is present in the system, it tends to serve that customer swiftly, resulting in a high probability of having a lone customer in the system. Conversely, with a SS, the likelihood of having one customer in the system remains reasonably high, especially if the job arrival rate is relatively low. However, it is conceivable that the SS might take a longer time to serve the single customer, potentially leading to a slightly lower probability compared to the scenario with a fast server.

Figure 3 illustrates the behavior of the system when both servers are busy. In such a scenario, incoming jobs are promptly served and exit the system, leading to a decrease in the probability curve over time. Figures 4 and 5 illustrate the probability curves of the system when the FS and SS are on vacation, respectively. It is observed that as the parameter n increases, the probability curves decrease and then attain a steady state. Figures 6 and 7 depict the expected system size and variance of the system. The graphs are plotted against time by varying the disaster rate. An increase in the disaster rate leads to a decrease in the system size. Notably, as the parameter ζ increases, all jobs are removed from the system, resulting in a decrease in the expected system size.

In Figure 8, we can observe the probability curves representing the busy state, denoted as $\pi_{n+3,1,1}$. These curves are plotted against the parameter n for different arrival rates. It is noticeable that as n increases while maintaining a fixed arrival rate of λ, the probability curves for $\pi_{3,1,1}$ decrease before eventually reaching a stable state. Furthermore, when the arrival rate increases, the probability values for $\pi_{3,1,1}$ initially rise and then gradually decrease, eventually stabilizing. Figures 9 and 10 depict the probability curves for vacation states, namely $\pi_{n,2,1}$ and $\pi_{n,1,2}$, respectively. These graphs illustrate how these vacation states change as n varies under different disaster rates ζ. It's evident that as the disaster rate increases, the curves decline and eventually reach a steady state.

From Figures 11 and 12, it is observed that as θ_1 increases (faster return of the fast server from vacation), the values of $\pi_{n,2,1}$ decrease for each n. This is expected, as a quicker return to service reduces the chance that the fast server is in vacation. A similar trend is observed for $\pi_{n,1,2}$ with respect to θ_2: increasing θ_2 leads to a decrease in the probability that the slow server is in vacation. For both cases, the probabilities decline as n increases, indicating lower chances of having a high number of packets in the system when one server is inactive. Notably, the values of $\pi_{n,1,2}$ are generally higher than those of $\pi_{n,2,1}$, suggesting that the system more frequently experiences the scenario where the slow server is in vacation and the fast server is active, likely due to the higher processing capability of the fast server. This numerical illustration highlights the sensitivity of system state probabilities to the vacation switching rates, particularly θ_2, and underscores the importance of optimizing these parameters to achieve desirable performance in heterogeneous server environments. In Figure 13, we examine the expected system size, which shows a decreasing trend as the disaster rate increases.

Figure 14 illustrates the impact of μ_1 and μ_2 on the TEC. The graph takes the form of a bowl-shaped surface, indicating that the cost function initially decreases and subsequently increases as both μ_1 and μ_2 increase. This visual representation highlights specific values of μ_1 and μ_2 that lead to the minimization of the total expected cost function, providing clarity on the combinations resulting in the optimal reduction of overall expected costs. Figure 15 illustrates the impact of λ on both TEC and TER. Notably, as λ increases, both costs show an upward trend. The figure distinctly reveals a breakeven point, occurring at approximately $\lambda = 0.7$, where TEC equals TER, indicating a balance between costs and revenue – no loss or profit at this point. Furthermore, for the specified set of cost values and other model parameters, a loss is incurred when $\lambda < 0.7$, while the system becomes profitable when $\lambda > 0.7$. This insight empowers system managers to make informed decisions based on the arrival rate of customers, allowing them to take proactive measures to minimize TEC and maximize TER.

Using PSO, the optimal values for the service rates μ_1^* and μ_2^*, which lead to the minimization of the TEC, are determined for various values of λ, considering the specified cost values.

- Case (i) $C_1 = 100$, $C_2 = 75$, $C_h = 50$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 40$, $C_F = 30$;
- Case (ii) $C_1 = 110$, $C_2 = 75$, $C_h = 50$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 40$, $C_F = 30$;
- Case (iii) $C_1 = 100$, $C_2 = 65$, $C_h = 50$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 40$, $C_F = 30$;
- Case (iv) $C_1 = 100$, $C_2 = 75$, $C_h = 60$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 40$, $C_F = 30$;
- Case (v) $C_1 = 100$, $C_2 = 75$, $C_h = 60$, $C_b = 140$, $C_{v_1} = 50$, $C_{v_2} = 40$, $C_F = 30$;
- Case (vi) $C_1 = 100$, $C_2 = 75$, $C_h = 60$, $C_b = 150$, $C_{v_1} = 40$, $C_{v_2} = 40$, $C_F = 30$;
- Case (vii) $C_1 = 100$, $C_2 = 75$, $C_h = 60$, $C_b = 150$, $C_{v_1} = 50$, $C_{v_2} = 30$, $C_F = 30$.

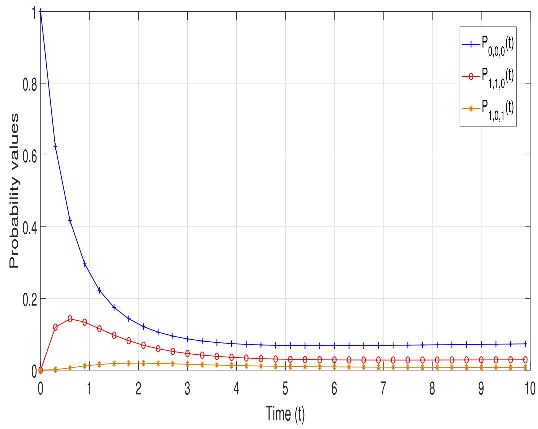

FIGURE 2. Probability curves of the idle server *vs.* time.

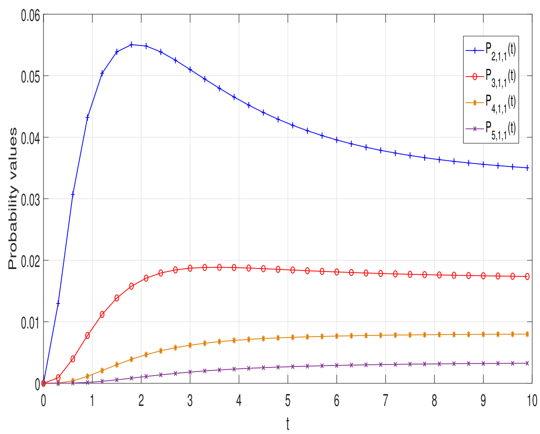

FIGURE 3. Probability cures of the busy state *vs.* time for different values of n.

Table 1 presents the optimal values of μ_1^* and μ_2^*, along with their corresponding TEC*. An observation from the table reveals that μ_1^*, μ_2^* and TEC* exhibit an increase with the rising values of λ, irrespective of the specific set of cost values. From Table 2 it is observed that an increase in service rates μ_1 and μ_2 leads to increase in TER and decrease in TEC, respectively.

While this study specifically analyzes an $M/M/2$ heterogeneous system, the underlying model and optimization framework possess significant breadth. The methodology readily extends to $M/M/c$ systems with an arbitrary number of heterogeneous servers and can incorporate features like threshold policies, priority queueing, customer impatience, and cost-sensitive resource allocation. The use of PSO further broadens the approach's applicability. As PSO's structure is model-independent, it can optimize diverse decision variables – such as vacation thresholds, repair rates, or service configurations – aligned with various operational goals. Consequently, these results not only depict cost-performance trade-offs in the studied system but also present a scalable, generalizable methodology for analyzing complex multi-server queueing environments.

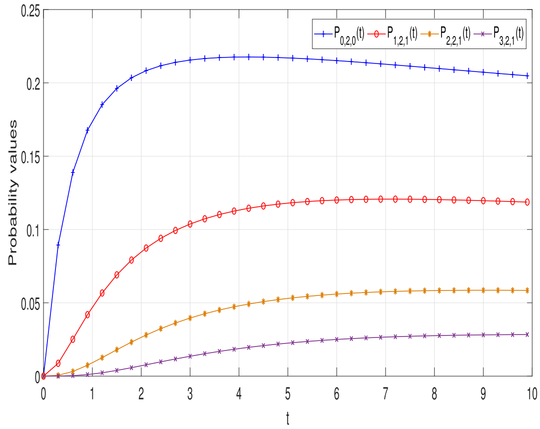

FIGURE 4. robability curves of $P_{n,2,1}(t)$ *vs.* time for different values of n.

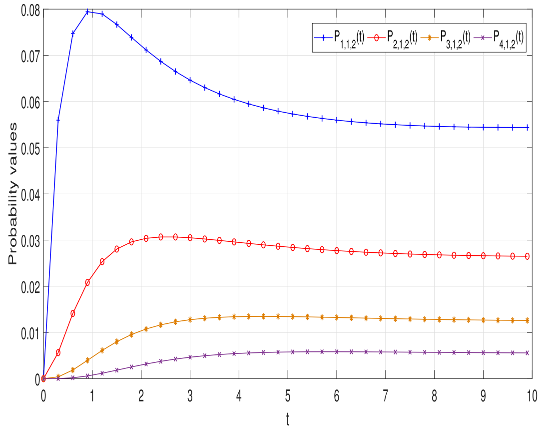

FIGURE 5. Probability curves of $P_{n,1,2}(t)$ *vs.* time for different values of n.

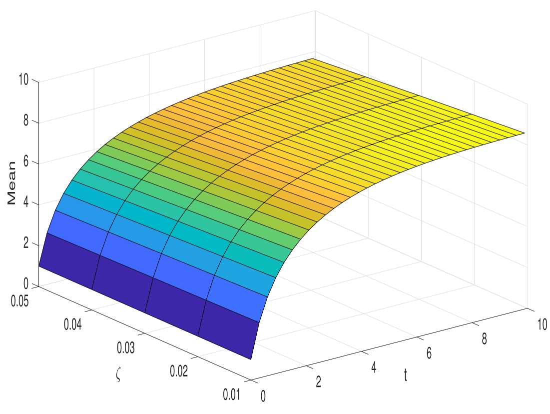

FIGURE 6. Expected system size.

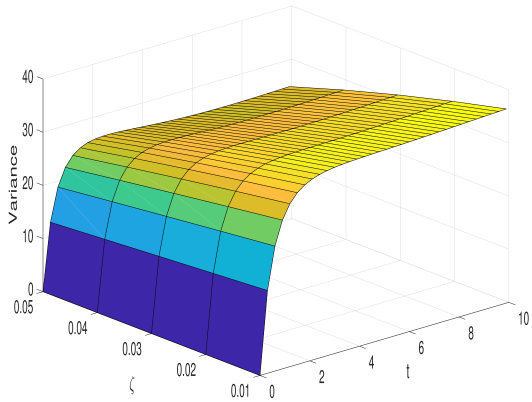

Figure 7. Variance.

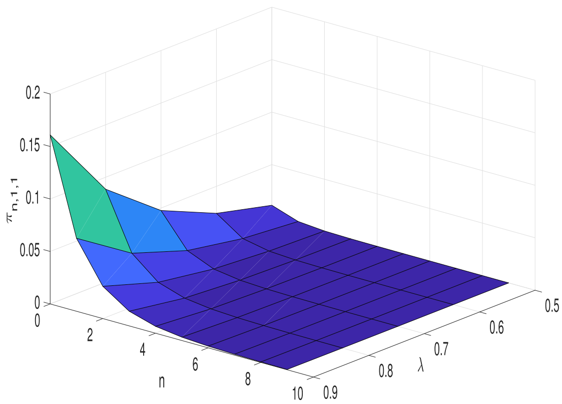

Figure 8. Probabilities of the busy state when both FS and SS are busy.

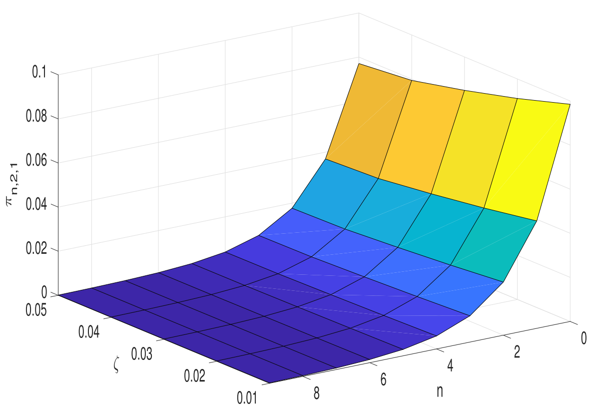

Figure 9. Probability curves when the FS is on vacation and the SS is active.

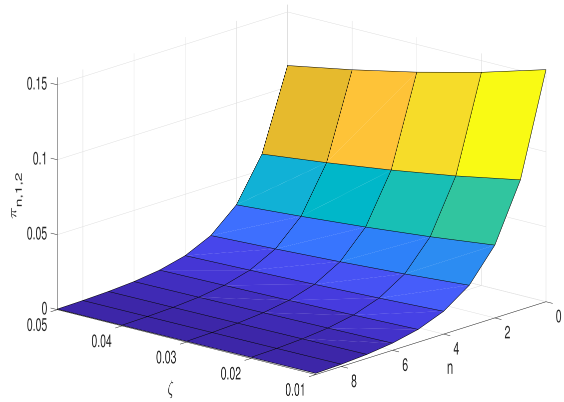

FIGURE 10. Probability curves when the SS is on vacation and the FS is active.

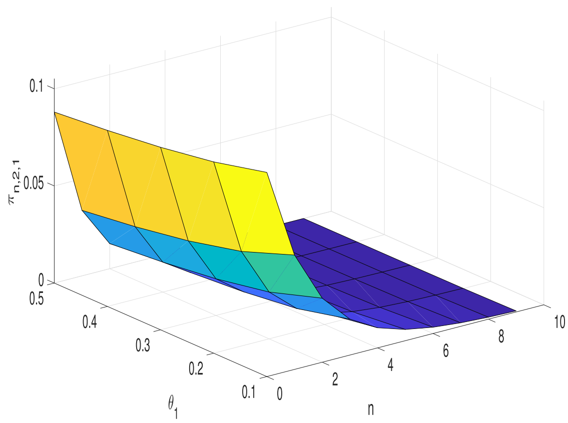

FIGURE 11. Probability curves of $\pi_{n,2,1}$ *vs.* θ_1 for various values of n.

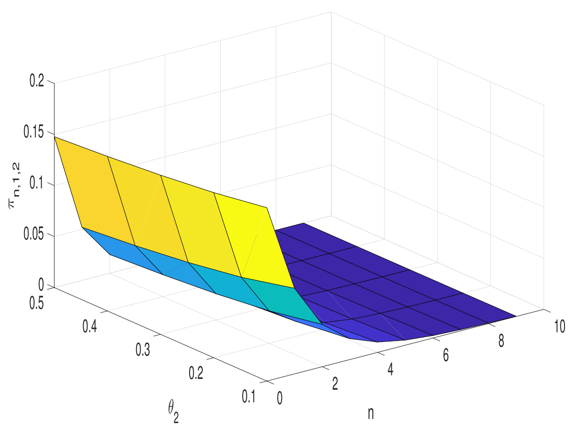

FIGURE 12. Probability curves of $\pi_{n,1,2}$ *vs.* θ_2 for various values of n.

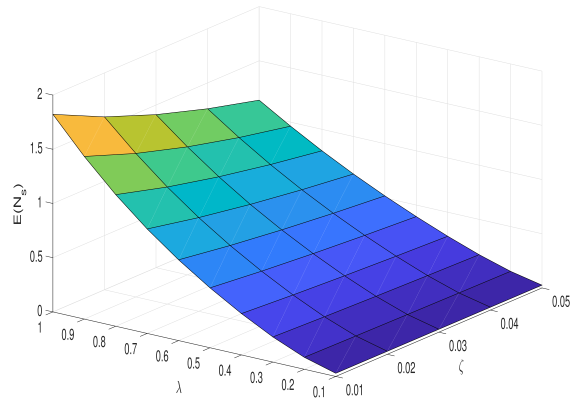

FIGURE 13. Expected system size in the steady state.

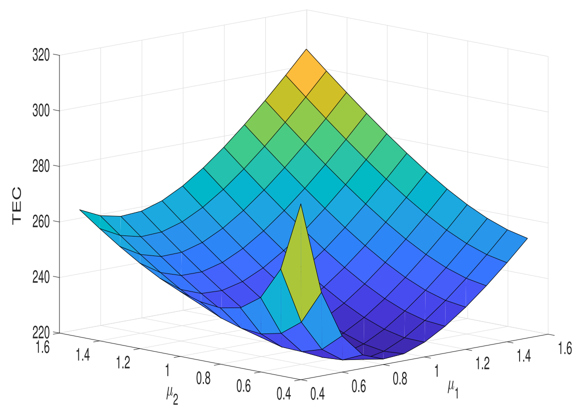

FIGURE 14. Convexity of the TEC for combinations of μ_1 and μ_2.

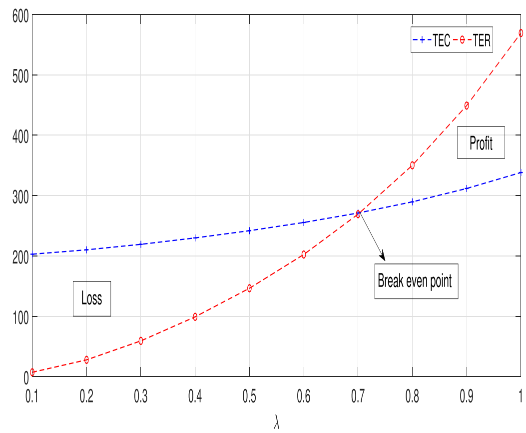

FIGURE 15. Effect of arrival rate λ *vs.* TER and TEC.

TABLE 1. Optimum values of μ_1^* and μ_2^*, that minimsse TEC*.

Case	λ	0.5	0.7	0.9
(i)	(μ_1^*, μ_2^*)	$(0.6468, 0.4379)$	$(0.9067, 0.6175)$	$(1.212, 0.7367)$
	TEC*	188.6734	250.8072	308.2055
(ii)	(μ_1^*, μ_2^*)	$(0.6100, 0.4563)$	$(0.866, 0.6385)$	$(1.1538, 0.7676)$
	TEC*	194.9637	259.6545	319.9942
(iii)	(μ_1^*, μ_2^*)	$(0.8022, 0.1869)$	$(0.8849, 0.6938)$	$(1.1885, 0.8389)$
	TER*	193.9335	244.2673	300.335
(iv)	(μ_1^*, μ_2^*)	$(0.6758, 0.4668)$	$(1.0147, 0.4885)$	$(1.2685, 0.7777)$
	TEC*	199.1436	266.5861	324.5224
(v)	(μ_1^*, μ_2^*)	$(0.6797, 0.4549)$	$(0.9984, 0.4876)$	$(1.2571, 0.7728)$
	TEC*	188.6734	250.8072	308.2055
(vi)	(μ_1^*, μ_2^*)	$(0.6857, 0.4491)$	$(0.9818, 0.6055)$	$(1.2693, 0.7832)$
	TEC*	188.6734	250.8072	323.3958
(vii)	(μ_1^*, μ_2^*)	$(0.5398, 0.2062)$	$(0.9171, 0.558)$	$(1.4117, 0.7256)$
	TEC*	226.8434	263.7763	323.7839

TABLE 2. Service rates *vs.* TER and TEC.

μ_1	TER	TEC	μ_2	TER	TEC
1.2	278.3607474	269.1986406	0.8	276.6821976	279.268131
1.3	283.8056614	251.7168539	0.9	279.8922671	264.1754851
1.4	290.0146716	237.2945451	1	283.8056614	251.7168539
1.5	296.818668	225.2313312	1.1	288.2882034	241.3766903
1.6	304.0921601	215.0116534	1.2	293.230504	232.7336171

9. Conclusion and future work

This paper analyzed an $M/M/2$ heterogeneous queueing system with dynamic vacations, system disasters, and server failure and repair, showcasing potential applications in cloud computing. We derived transient and steady-state probabilities and determined an optimal service rate minimizing total expected cost. The PSO technique has been employed to obtain the optimal service rates that minimize the total expected cost under a specific cost function. Numerical results visually demonstrated the impact of system and cost parameters. Our findings lay a foundation for extending the model to $M/M/C$ systems with a working vacation, offering insights into diverse queuing scenarios. Overall, this research contributes valuable understanding to complex systems, informing optimization strategies and resource management.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author contribution statement

A. Mohammed shapique: Identified the problem, designed the model and analysed the data. ***R. Sudhesh***: Interpretation of the data and Critical revision of the article.

Competing interests

The authors has no relevant financial or non-financial interests to disclose.

References

[1] R. Agarwal, D. Agarwal and S. Upadhyaya, Cost optimisation of a heterogeneous server queueing system with working breakdown using PSO. *Int. J. Math. Oper. Res.* **26** (2023) 410–424.

[2] R.O. Al-Seedy, A.A. El-Sherbiny, S.A. El-Shehawy and S.I. Ammar, Transient solution of the $M/M/c$ queue with balking and reneging. *Comput. Math. App.* **57** (2009) 1280–1285.

[3] S.I. Ammar, Transient behavior of a two-processor heterogeneous system with catastrophes, server failures and repairs. *Appl. Math. Modell.* **38** (2014) 2224–2234.

[4] S. Dharmaraja, Transient solution of a two-processor heterogeneous system. *Math. Comput. Model.* **32** (2000) 1117–1123.

[5] K. Divya and K. Indhira, Performance analysis and ANFIS computing of a Markovian queuing model with intermittently accessible server under a hybrid vacation policy. *RAIRO-Oper. Res.* **58** (2024) 1257–1279.

[6] B.T. Doshi, Queueing systems with vacations – a survey. *Queueing Syst.* **1** (1986) 29–66.

[7] V. Goswami, Analysis of discrete-time multi-server queue with balking. *Int. J. Manage. Sci. Eng. Manage.* **9** (2014) 21–32.

[8] J.C. Ke, C.H. Wu and W.L. Pearn, Analysis of an infinite multi-server queue with an optional service. *Comput. Ind. Eng.* **65** (2013) 216–225.

[9] J. Kennedy and R. Eberhart, Particle swarm optimization, in Proceedings of ICNN'95 – International Conference on Neural Networks. Vol. 4. IEEE (1995) 1942–1948.

[10] D. Kothandaraman and I. Kandaiyan, Analysis of a heterogeneous queuing model with intermittently obtainable servers under a hybrid vacation schedule. *Symmetry* **15** (2023) 1304.

[11] B. Krishnamoorthi, On Poisson queue with two heterogeneous servers. *Oper. Res.* **11** (1963) 321–330.

[12] A. Krishnamoorthy and C. Sreenivasan, An $M/M/2$ queueing system with heterogeneous servers including one with working vacation. *Int. J. Stoch. Anal.* **2012** (2012) 145867.

[13] B.K. Kumar and D. Arivudainambi, Transient solution of an $M/M/1$ queue with catastrophes. *Comput. Math. Appl.* **40** (2000) 1233–1240.

[14] B.K. Kumar and S.P. Madheswari, An $M/M/2$ queueing system with heterogeneous servers and multiple vacations. *Math. Comput. Model.* **41** (2005) 1415–1429.

[15] H. Leemans, Waiting time distribution in a two-class two-server heterogeneous priority queue. *Perform. Eval.* **43** (2001) 133–150.

[16] Y. Levy and U. Yechiali, Utilization of idle time in an $M/G/1$ queueing system. *Manage. Sci.* **22** (1975) 202–211.

[17] A. Melikov, S. Aliyeva and J. Sztrik, Analysis of instantaneous feedback queue with heterogeneous servers. *Mathematics* **8** (2020) 2186.

[18] L. Sakalauskas, L. Kaklauskas and R. Macaitiene, Stalling in queuing systems with heterogeneous channels. *Appl. Sci.* **14** (2024) 773.

[19] O.P. Sharma and J. Dass, Initial busy period analysis for a multichannel Markovian queue. *Optimization* **20** (1989) 317–323.

[20] V.P. Singh, Two-server Markovian queues with balking: heterogeneous vs. homogeneous servers. *Oper. Res.* **18** (1970) 145–159.

[21] A. Sridhar and R. Allah Pitchai, Analyses of a Markovian queue with two heterogeneous servers and working vacation. *Int. J. Appl. Oper. Res.* **5** (2015) 1–15.

[22] R. Sudhesh, A. Mohammed Shapique and S. Dharmaraja, Analysis of a multiple dual-stage vacation queueing system with disaster and repairable server. *Methodol. Comput. Appl. Probab.* **24** (2022) 2485–2508.

[23] N. Sujatha and G. Deekshitulu, A study of transient solution of fractional $M/M/2$ queue with homogeneous and heterogeneous servers. *Int. J. Math. Oper. Res.* **31** (2025) 95–109.

[24] H. Takagi, Queueing Analysis: A Foundation of Performance Analysis, Volume 1: Vacation and Priority Systems, Part 1. Elsevier Science Publishers B.V., Amsterdam (1991).

[25] R. Tian, X. Chen, T. Song and T. Wang, Cost optimization of queueing systems with flexible priorities and heterogeneous servers. *Heterogeneous Serv.* **33** (2025) 2100–2107.

[26] K.S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications. John Wiley & Sons (2001).

[27] D.Y. Yang, Y.H. Chen and C.H. Wu, Modelling and optimisation of a two-server queue with multiple vacations and working breakdowns. *Int. J. Prod. Res.* **58** (2020) 3036–3048.