RAIRO-Oper. Res. 58 (2024) 2569–2589
https://doi.org/10.1051/ro/2024083

RAIRO Operations Research
www.rairo-ro.org

EQUILIBRIUM BALKING STRATEGIES IN UNOBSERVABLE QUEUES WITH MULTIPLE VACATIONS AND AN OPTIONAL SERVICE

Dong-Yuh Yang[1,*], Yin-Ying Dai[2] and Chia-Huang Wu[2]

Abstract. This paper examines equilibrium mixed strategies in unobservable Markovian queues featuring a second optional service with server vacations, where arriving customers may choose to join or balk the system. All customers arriving at the system receive the essential service, and some customers opt for the second service after the first service has been completed. Once all customers in the system have been served, the server takes the first of multiple vacations. If no customers are waiting upon from the vacation, then the server takes another vacation. In unobservable queues, arriving customers cannot know the queue length; however, the information pertaining to the server state may be available. By examining unobservable queues (fully unobservable and almost unobservable cases), it is possible to formulate an equilibrium joining strategy as well as the socially optimal probability of joining a fully unobservable queue. This paper also presents numerical examples illustrating how system parameters affect mixed equilibrium and socially optimal balking strategies.

Mathematics Subject Classification. 90B22, 60K25.

Received July 17, 2023. Accepted April 4, 2024.

1. Introduction and literature review

In daily life, we often encounter situations, in which all customers receive an essential service (ES), and some customers can then partake in a secondary (optional) service. Madan [23] was the first to consider a second optional service (SOS) in the context of an $M/G/1$ queue using the supplementary variable technique. In that study, the service times of the first essential service followed a general distribution, while that of the SOS followed an exponential distribution. Medhi [26] generalized the findings of Madan [23] by proposing a general SOS time. Based on an in-depth analysis of Madan [23]'s model, Choudhury [5] derived the waiting time distribution and the departure point queue size distribution. Wang [38] further extended the work of Madan [23] to include an unreliable server, where the repair times of both service phases are generally distributed. We refer the readers to Wang *et al.* [39], Ke *et al.* [18], Ghorbani-Mandolakani and Salehi Rad [10] and Gao [9] for related works in queueing systems with SOS.

Vacation queueing is characterized by a server that is unavailable for a stochastic duration, during which secondary work can be completed. Much of the literature on queueing systems, discusses single vacation and

Keywords. Equilibrium strategy, multiple vacations, second optional service, unobservable queue.

[1] Institute of Information and Decision Sciences, National Taipei University of Business, Taipei, Taiwan.

[2] Department of Industrial Engineering and Management, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.

*Corresponding author: yangdy@ntub.edu.tw

multiple vacations. This paper particularly focuses on queues with multiple vacations, in which the server may leave for a vacation as soon as the last customer leaves the system. Furthermore, if there are no customers in the system when a vacation is completed, then a new vacation begins. Researchers have extensively studied vacation queueing models due to their theoretical significance and wide practical applicability in telecommunication systems, production-inventory systems, call centers, computer and communication networks, and manufacturing systems. To gain a thorough understanding of queues with server vacations, please refer to Doshi [6], Takagi [32], Tian and Zhang [34], Ke *et al.* [17] and the references therein. Upadhyaya [36] presented an overview of queues with vacation policies. Wu *et al.* [40] considered $BMAP/G/1$ queues with negative arrivals, SOS and multiple vacations, where the arrival process is described using the batch Markovian arrival process. Maraghi *et al.* [25] studied an $M^x/G/1$ queue with SOS, server breakdowns and Bernoulli vacation. Yue *et al.* [43] considered an $M/M/1$ queue with SOS, impatient customers and multiple vacations. Chakravarthy [3] presented a stationary analysis of a vacation queue with SOS and non-renewal process, where service times obey a phase-type distribution. Madheswari *et al.* [24] and Jain and Kaur [15] studied retrial queues with SOS and Bernoulli vacations. Jin *et al.* [16] modeled the cloud architecture as a multi-server queue with SOS and asynchronous vacations. Using the matrix-geometric method to compute the steady-state probability distribution, they derived a pricing policy that maximizes social welfare. Vijaya Laxmi and Jyothsna [37] recently considered an $M/M/1/N$ queue with SOS, balking, reneging and working vacations, wherein the server continues working at a different rate. They developed a cost-revenue model and used the particle swarm optimization algorithm to determine the optimal service rates, in which the total expected cost is minimized. Despite extensive research on queueing systems with SOS, most previous studies have focused on performance analysis, optimal design and optimal control. To the best of our knowledge, no previous research has focused on the queueing game of vacation queues with SOS. The wide practical applicability of vacation queues prompted us to focus on equilibrium balking strategies for unobservable queues with SOS and multiple vacations. In the following, we provide a brief review of queuing game models.

The development of the social economy has prompted considerable research into queueing systems from an economic perspective. The equilibrium behavior of customers in queueing systems was first introduced by Naor [27], based on research of an observable $M/M/1$ queue, wherein arriving customers are informed of the queue length before deciding whether to join. In extending Naor's model to the case of unobservable queues, Edelson and Hildebrand [8] demonstrated that in an $M/M/1$ queue with no balking, the same toll charged to all customers maximizes Naor's social welfare function and expected revenue per unit time. Hassin and Haviv [13] conducted a comprehensive survey of equilibrium behavior in queues. Please see Hassin [12] for a review of the literature pertaining to customer decisions in queueing systems. A recent survey by Haviv and Ravner [14] provided an overview of developments in dealing with queues involving the strategic timing of arrivals. Burnetas and Economou [2] and Economou and Manou [7] classified queueing systems into four cases based on the availability of information for customers: fully unobservable case, almost unobservable case, fully observable case, and almost observable case. In the current paper, we discuss only equilibrium balking strategies in the two unobservable cases, in which information related to the queue length is unavailable to arriving customers.

In the past decade, there has been much research on queueing game models. Next, we conduct a brief review of critical literature. Sun *et al.* [30] derived equilibrium mixed strategies in an unobservable single-server Markovian queue with multiple vacations. Li *et al.* [21] performed an equilibrium analysis of the fully observable and unobservable $M/M/1$ queues with partial breakdowns, in which customers could be served at a reduced rate after a server failure. Yue *et al.* [44] examined equilibrium strategies in fully observable and unobservable $M/M/1$ queues with setup times and single vacation policy. For the fully unobservable case, they derived a socially optimal strategy to maximize the total expected social benefit per unit time. Lee [19] investigated strategies by which to derive optimal pricing in an unobservable Markovian queue, where the arrival of a negative customer triggers a server breakdown. Xu and Xu [41] conducted a study on an $M/M/1$ queue with partial failures, in which entrance into the system is disallowed whenever the server fails. They derived equilibrium threshold strategies for the fully observable case and the mixed equilibrium strategies for the fully unobservable case. Tian [33] determined the social benefit and optimal pricing strategies in an unobservable $M/M/1$ queue with delayed multiple vacations. Sun and Wang [29] discussed equilibrium joining and socially optimal strategies in a Markovian queue with repairs and two types of arrivals (positive and negative customers) in two information cases. Aghsami and Jolai [1] focused on analyzing customer behavior in $M/M/1$ queues with partial breakdowns and interruptible setup/closedown policy. The authors investigated equilibrium balking strategies, considering service interruptions, and derived conditions for social benefits under various system states. Li and Wang [20] investigated equilibrium and socially optimal strategies in observable and unobservable $M/M/1$ retrial queues with constant retrial policy and catastrophes. Panda and Goswami [28] recently derived equilibrium strategies and computed the social benefit of an $M/M/1$ queue with negative customers and working vacations in fully unobservable and almost unobservable cases. Sun *et al.* [31] analyzed customer strategies in single-server Markovian queues with a B-limited service rule and multiple vacations. It explored behaviors in fully observable and unobservable queues, examining equilibrium and socially optimal joining strategies. Tian *et al.* [35] recently considered a Markovian queue with changeable service rates under N-policy. They investigated how service rate adjustments and threshold N affect customer strategies in observable and unobservable queues, focusing on equilibrium and social optimality.

1.1. Research gaps and contributions

As mentioned earlier, the performance measures of various systems with SOS and vacation policies have been investigated. These previous studies demonstrated the applicability of queueing models with SOS. Table 1 summarizes the literature on queueing game models relative to this paper, which identifies no work on a vacation queueing system with SOS and strategic customers' choices. This research gap motivates us to investigate equilibrium mixed strategies in unobservable Markovian queues featuring a SOS with server vacations. The main contributions and innovations of this paper are summarized as follows:

– This study is the first to analyze the customer equilibrium strategy for queueing systems with multiple vacations (MV) and SOS. The analytical and numerical results show that the customer strategies depend on the cost-reward parameters and information scenarios. Particularly, for the almost observable queues, whether the system manager can discourage arrivals or encourage arrivals depends on which equilibrium appears.
– The investigated queueing model in the current study is similar to the model proposed by Yue *et al.* [43], except for the issue of reneging and underlying objectives. Yue *et al.* [43] used the probability generating function method to enable the explicit derivation of performance measures. In contrast, our analysis focused on the use of game-theoretic analysis to model the strategic behavior of customers and social welfare in situations involving unobservable queues, which is an issue that has never been explored.
– For our study's expected system length and various major performance measures, we derived explicit mathematical expressions that can be reduced to cases with only SOS or without vacation policy when taking the limits. With numerical analysis, this research provides critical managerial insights for practitioners to evaluate system performances with strategic customers' choices. Thus, this study yields benefits in practical management problems.

1.2. Practical applications

The proposed queueing system has a wide range of potential applications. For example, consider a hair salon with a single hairdresser providing necessary and optional services. The necessary service refers to the haircut service that all arriving customers request. Customers can also choose an optional service, such as a shampoo, hair straightening, or perm. In the absence of customers, the hairdresser may clean the workstations, sweep, and perform maintenance. Given that customers arriving at the hair salon are unaware of the number of customers already waiting for service, the scenario when they decide whether to enter upon arrival can be either fully unobservable or almost unobservable. This depends on whether the state of the hairdresser is visible outside. Another example is the administrative office in a school or government. When we go there for document application service, the state of the service personnel can be observed; however, the number of customers

TABLE 1. Summary of previous literature studies on queueing game models.

Study	Types*	Characteristics	Findings/Results*
Yue *et al.* [44]	*fo* & *fu*	Simple vacation & setup times	ES & SS
Sun *et al.* [30]	*fu*	Multiple vacations	ES
Li *et al.* [21]	*fo* & *fu*	Partial breakdown	ES
Lee [19]	*fu* & *au*	Server breakdown	ES & OPS
Xu and Xu [41]	*fo* & *fu*	Partial breakdown	ES & SS
Tian [33]	*fu* & *au*	Delayed multiple vacations	ES & SS & OPS
Sun and Wang [29]	*fu* & *ao*	Server breakdown	ES & SS
Aghsami and Jolai [1]	*fo*	Partial breakdowns & interruptible setup/closedown policy	ES & equilibrium social benefits
Li and Wang [20]	*fo* & *au*	Customer retrial system catastrophes	ES & SS
Panda and Goswami [28]	*fu* & *au* & *fo* & *ao*	Multiple working vacations	ES & equilibrium social benefits
Sun *et al.* [31]	*fo* & *fu*	Batch-limited service rule multiple vacations	ES & SS
Tian *et al.* [35]	*fo* & *fu*	Changeable service rates	ES & SS

Notes. (*)Types: fully observable (*fo*), fully unobservable (*fu*), almost observable (*ao*), almost unobservable (*au*). (*)Findings/Results: equilibrium strategy (ES), optimal pricing strategy (OPS), socially optimal strategy (SS).

in the service queue is often unobservable, which belongs to almost unobservable cases. When the queue is empty, the service personnel may support other business tasks (vacations), and some document application requests need additional approval processes or information confirmation procedures (SOS). Along with the service industry, the proposed model also can be applied in the manufacturing industry. Considering a factory produces chemical textiles, all orders need a standard procedure (ES), including sizing, de-sizing, scouring, bleaching, mercerization, dyeing, printing, etc. Partial of them require customized service (SOS) such as special surface treatment, bespoke tailoring, personalization design, etc. When a potential customer determines whether to place an order, the manufacturer may provide information about the production. Nevertheless, the system length requires more effort to evaluate. An appropriate unobservable queueing model can be used to analyze the system characteristics and the equilibrium strategy, assessing the production profit and benefits for both the manufacturer and the customer. Furthermore, the socially optimal strategy can be derived and examined.

In these applications, the decision of the arriving customer is critical to system profit and social benefit, particularly depending on whether the server state is provided or not. The research question of this paper focuses on examining the customer strategic behavior in a Markovian queue with MV and SOS. We aim to derive crucial system characteristics, investigate the customer equilibrium strategy under different cost-reward structures in two unobservable cases, and develop the socially optimal strategy in the fully unobservable cases. The remainder of this paper is organized as follows. In Section 2, we outline a queueing system with MV and SOS, including the assumptions and notations. Section 3 examines mixed equilibrium strategies for almost unobservable queues. In Section 4, we discuss equilibrium and socially optimal balking strategies for fully unobservable queues. Section 5 presents numerical analysis of equilibrium and socially optimal entrance probabilities. In Section 6, we summarize the paper and draw conclusions.

TABLE 2. The parameters and notations used in the model.

Parameters	
λ	Mean arrival rate
δ	Probability that a customer requests a SOS, $0 \leq \delta \leq 1$
μ_1	Mean service rate for the ES
μ_2	Mean service rate for a SOS, $\mu_2 > \mu_1(2-\delta)$
ν	Mean vacation rate
R	Customer reward from the service
C	Waiting cost per unit time
Notations used in the almost unobservable case	
$p_{au}(i,n)$	Steady-state probability for the state $(i,n) \in \Omega$
(q_0, q_1, q_2)	Joining probability vector of a new arriving customer
$G_i(z)$	Probability generating function as the server is in the state i, $i = 0, 1, 2$
$p(n \mid I = i)$	Conditional steady-state probability of having n customers in the system given that the server is in state i, $i = 0, 1, 2$
$E(N \mid I = i)$	Conditional expected number of customers in the system given that the server is in state i, $i = 0, 1, 2$
$W_{au}(i,n)$	Expected sojourn time of an entrance customer arriving at a system in state $(i,n) \in \Omega$
$q_e(i)$	Joining probability vector in a mixed equilibrium strategy at a system in state i, $i = 0, 1, 2$
Υ	Expected customer service time
S_i	Expected net benefit function for a customer arriving at a system in state i, $i = 0, 1, 2$
Notations used in the fully unobservable case	
q	Customers' joining probability
$E(L_s)$	Expected number of customers in the system
$p(i,n)$	Steady-state probability for the state $(i,n) \in \Omega$
$W(q)$	Expected sojourn time of a customer
$S(q)$	Expected net benefit function for a customer
$\mathrm{SW}(q)$	Social welfare per unit time
q_e	Joining probability for the unique equilibrium strategy
q^*	Joining probability for the social welfare maximization strategy

2. MODEL DESCRIPTION AND FORMULATION

In the following, we consider a Markovian queue with MV and an SOS. The arrival of customers at the system is described by a Poisson process at the rate of λ. After the arriving customers receive the ES, they may request an SOS with probability $\delta(0 \leq \delta \leq 1)$ or leave the system with probability $1-\delta$. The ES and SOS are provided by the same server, thereby limiting service to only one customer at a time on a first-come, first-served (FCFS) basis. The service times of ES and SOS are both exponentially distributed with mean service time μ_1^{-1} and μ_2^{-1}, respectively. It is further assumed that $\mu_2 > \mu_1(2-\delta)$. The system becomes vacant at the instant that an ES or SOS is completed, at which point the server leaves for a vacation of random length following an exponential distribution with mean rate of ν. If no customers are in the system at the completion of a vacation, then the server leaves for another vacation, and the cycle continues until at least one customer appears in the system.

Consider a situation in which customers are unable to leave the queue before they receive the ES. The customer receives a reward R following the completion of service. Each customer is riskneutral and incurs waiting cost C per unit time for as long as they remain in the system. We define a linear cost-reward function $S = R - C \cdot E[W]$, where S and $E[W]$ respectively denote the expected net utility and expected sojourn time in the system. All customers are assumed to be homogeneous with respect to R and C, and seek to maximize S. The inter-arrival times, service times of ES and SOS, and vacation times are assumed to be mutually independent. In accordance with Kendall's notation, and for simplicity, the proposed model is referred to as an $M/M/1/MV$ queue with SOS. To avoid trivial cases, we further assume that the reward obtained following the completion of a service exceeds the expected cost for a customer arriving at a system and finding no customers in the system, *i.e.*,

$$R > C \cdot \left(\mu_1^{-1} + \delta\mu_2^{-1} + \nu^{-1}\right). \tag{1}$$

Let $I(t)$ and $N(t)$ respectively denote the server state and the number of customers in the system at time t. Further, $I(t)$ is defined as follows:

$$I(t) = \begin{cases} 0, & \text{if the server is on vacation at time } t; \\ 1, & \text{if the server is busy and provides ES at time } t; \\ 2, & \text{if the server is busy and provides SOS at time } t. \end{cases}$$

The stochastic process $\{(I(t), N(t)), t \geq 0\}$ is a quasi birth-and-death (QBD) process with state space given by $\Omega = \{(0, n) : n \geq 0\} \cup \{(i, n) : i = 1, 2, n \geq 1\}$. Here, our analysis focuses on strategic behavior in unobservable $M/M/1/MV$ queue with SOS, where arriving customers are unaware of the queue length. In accordance with the availability of information pertaining to the state of the server, we discuss two cases: (1) the almost unobservable case (customers can observe the state of the server); and (2) the fully unobservable case (customers cannot perceive the state of the server). Table 2 summarizes the parameters and notations used in these two unobservable queues.

3. Almost unobservable $M/M/1/MV$ queue with SOS

This section considers an almost unobservable $M/M/1/MV$ queue with SOS, where arriving customers are aware of the server state but not aware of the queue length. In the following, we prove that it is possible to formulate equilibrium mixed strategies for the almost unobservable case. Note that a mixed strategy is represented by a probability vector (q_0, q_1, q_2), $q_0 \in (0, 1]$, $q_i \in [0, 1]$ for $i = 1, 2$, where q_i denotes the joining probability that an arriving customer observes that the server is in state i. This means that the effective arrival rate is equal to λq_i when the server is in state i. Figure 1 shows the state-transition-rate diagram for the almost unobservable $M/M/1/MV$ queue with SOS. Let $p_{au}(i, n)$ be the steady-state probability in the almost unobservable queue, in which the system is in state (i, n). That is, $p_{au}(i, n) = \lim_{t\to\infty} \Pr(I(t) = i, N(t) = n)$, $(i, n) \in \Omega$.

3.1. Stability conditions

Referring to Figure 1, the set of steady-state equations is given as follows:

$$\lambda q_0 p_{au}(0, 0) = \mu_1(1 - \delta)p_{au}(1, 1) + \mu_2 p_{au}(2, 1), \tag{2}$$

$$(\lambda q_0 + \nu)p_{au}(0, n) = \lambda q_0 p_{au}(0, n - 1), \quad n \geq 1, \tag{3}$$

$$(\lambda q_1 + \mu_1)p_{au}(1, 1) = \nu p_{au}(0, 1) + \mu_1(1 - \delta)p_{au}(1, 2) + \mu_2 p_{au}(2, 2), \tag{4}$$

$$(\lambda q_1 + \mu_1)p_{au}(1, n) = \lambda q_1 p_{au}(1, n - 1) + \nu p_{au}(0, n) + \mu_1(1 - \delta)p_{au}(1, n + 1) + \mu_2 p_{au}(2, n + 1), \quad n \geq 2, \tag{5}$$

$$(\lambda q_2 + \mu_2)p_{au}(2, 1) = \mu_1 \delta p_{au}(1, 1), \tag{6}$$

$$(\lambda q_2 + \mu_2)p_{au}(2, n) = \lambda q_2 p_{au}(2, n - 1) + \mu_1 \delta p_{au}(1, n), \quad n \geq 2. \tag{7}$$

The normalization equation is as follows:

$$\sum_{n=0}^{\infty} p_{au}(0, n) + \sum_{n=1}^{\infty} p_{au}(1, n) + \sum_{n=1}^{\infty} p_{au}(2, n) = 1. \tag{8}$$

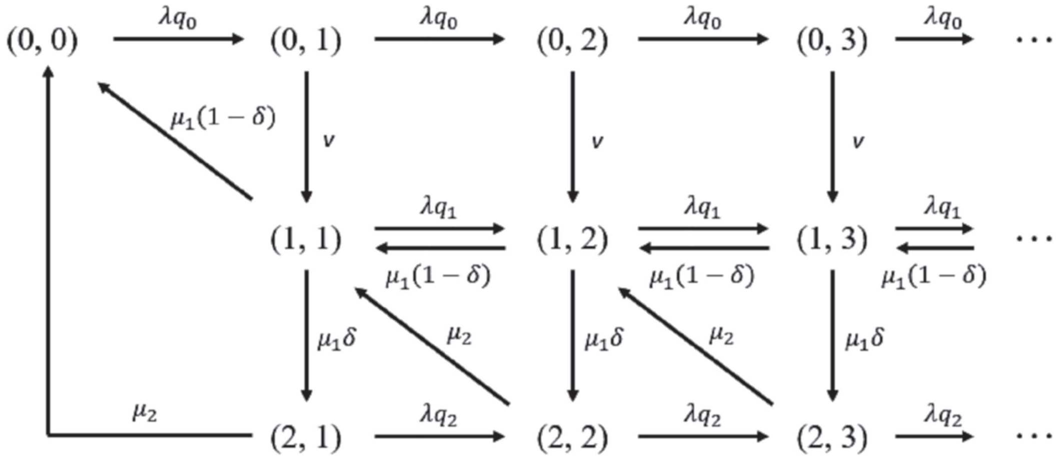


FIGURE 1. State-transition-rate diagram of the almost unobservable $M/M/1/MV$ queue with SOS.

From equations (2) and (6), we have

$$p_{au}(1,1) = \frac{\lambda q_0(\lambda q_2 + \mu_2)}{\mu_1(\lambda q_2(1-\delta) + \mu_2)} p_{au}(0,0), \tag{9}$$

$$p_{au}(2,1) = \frac{\lambda q_0 \delta}{\lambda q_2(1-\delta) + \mu_2} p_{au}(0,0). \tag{10}$$

Let $G_0(z)$, $G_1(z)$, and $G_2(z)$ be the probability generating functions pertaining to the number of customers in the system, defined by $G_0(z) = \sum_{n=0}^{\infty} p_{au}(0,n)z^n$, $G_1(z) = \sum_{n=1}^{\infty} p_{au}(1,n)z^n$, $G_2(z) = \sum_{n=1}^{\infty} p_{au}(2,n)z^n$, $|z| \le 1$. We multiply z^n to both sides of equations (2)–(7) and then sum the results in the following equations:

$$G_0(z) = \frac{\lambda q_0 + \nu}{\lambda q_0(1-z) + \nu} p_{au}(0,0), \tag{11}$$

$$G_2(z) = \frac{\mu_1 \delta}{\lambda q_2(1-z) + \mu_2} G_1(z), \tag{12}$$

$$\begin{aligned}\left[\lambda q_1(1-z) + \mu_1 - \mu_1(1-\delta)z^{-1}\right] G_1(z) = \nu G_0(z) - \nu p_{au}(0,0) - \mu_1(1-\delta) p_{au}(1,1) \\ + \frac{\mu_2}{z} G_2(z) - \mu_2 p_{au}(2,1).\end{aligned} \tag{13}$$

Using equations (9)–(13), $G_1(z)$ and $G_2(z)$ can be written in terms of $p_{au}(0,0)$ as follows:

$$G_1(z) = \frac{\lambda q_0 z(z-1)(\lambda q_0 + \nu)[\lambda q_2(1-z) + \mu_2] p_{au}(0,0)}{[\lambda q_0(1-z) + \nu]\{z[\lambda q_1(1-z) + \mu_1][\lambda q_2(1-z) + \mu_2] - \mu_1[\lambda q_2(1-z)(1-\delta) + \mu_2]\}}, \tag{14}$$

$$G_2(z) = \frac{\lambda q_0 \mu_1 \delta z(z-1)(\lambda q_0 + \nu) p_{au}(0,0)}{[\lambda q_0(1-z) + \nu]\{z[\lambda q_1(1-z) + \mu_1][\lambda q_2(1-z) + \mu_2] - \mu_1[\lambda q_2(1-z)(1-\delta) + \mu_2]\}}. \tag{15}$$

Based on the normalization equation in equation (8), we have

$$G_0(1) + G_1(1) + G_2(1) = 1. \tag{16}$$

By substituting $z = 1$ into equation (11), we get

$$G_0(1) = \frac{\lambda q_0 + \nu}{\nu} p_{au}(0,0). \tag{17}$$

We then evaluate $G_1(z)$ and $G_2(z)$ by taking the limit as $z \to 1$ using equations (14) and (15), respectively. By applying L'Hospital's rule, and let $\rho_{au} = \mu_1\mu_2 - \lambda(q_1\mu_2 + q_2\mu_1\delta)$ for convenience, we obtain

$$G_1(1) = \frac{\lambda q_0 \mu_2(\lambda q_0 + \nu)}{\nu \rho_{au}} p_{au}(0,0), \tag{18}$$

$$G_2(1) = \frac{\lambda q_0 \mu_1 \delta(\lambda q_0 + \nu)}{\nu \rho_{au}} p_{au}(0,0). \tag{19}$$

Substituting equations (17)–(19) into equation (16), and performing simple algebraic manipulations, we can derive $p_{au}(0,0)$ as follows:

$$p_{au}(0,0) = \frac{\nu \rho_{au}}{(\lambda q_0 + \nu)[\lambda q_0(\mu_2 + \mu_1\delta) + \rho_{au}]}. \tag{20}$$

The probability is $0 < p_{au}(0,0) < 1$, such that the sufficient stability condition is given by

$$\rho_{au} = \mu_1\mu_2 - \lambda(q_1\mu_2 + q_2\mu_1\delta) > 0. \tag{21}$$

If the inequality in equation (21) holds, then the system is stable. Note that the stability condition is independent of the parameter ν since the server serves the customer once the system is not empty. Equation (21) can be further represented as

$$\frac{1}{\lambda} > q_1\left(\frac{1}{\mu_1}\right) + q_2\delta\left(\frac{1}{\mu_2}\right). \tag{22}$$

It implies that the system is stable if the expected length of the customer inter-arrival time exceeds the expected service time for an entrance customer, a result that aligns with the intuitive understanding. In the following, we discuss three special cases of the stability condition:

(1) If $q_1 = q_2 = 1$, we obtain the stability condition for the queueing system without balk consideration. This condition is identical to the result of Madan [23] when the service time of the first essential service follows an exponential distribution.
(2) If $q_1 = 0$ or $\mu_1 \to \infty$, the service time for the ES is negligible. In this case, equation (22) simplifies the stability condition for the classical $M/M/1$ queue with mean arrival rate $q_2\delta\lambda$.
(3) Similarly, if $q_2 = 0$, $\delta = 0$, or $\mu_2 \to \infty$, the SOS can be ignored and therefore equation (22) represents the stability condition for an $M/M/1$ queue with mean arrival rate $q_1\lambda$.

3.2. Expected sojourn time

Let $p(n|I = i)$ represent the conditional steady-state probability of having n customers in the system given that the server is in state i. It follows that

$$p(n|I=0) = \frac{p(0,n)}{\sum_{n=0}^{\infty} p(0,n)} = \frac{p(0,n)}{G_0(1)}, \qquad n \geq 0, \tag{23}$$

$$p(n|I=i) = \frac{p(i,n)}{\sum_{n=1}^{\infty} p(i,n)} = \frac{p(i,n)}{G_i(1)}, \qquad n \geq 1, \quad i = 1, 2, \tag{24}$$

where $G_0(1)$, $G_1(1)$ and $G_2(1)$ are given in equations (17)–(19), respectively.

Based on equations (11), (14), and (15), we differentiate $G_i(z)$ and then substitute $z = 1$ into it. This gives the following:

$$G_0'(1) = \frac{\lambda q_0(\lambda q_0 + \nu)}{\nu^2} p_{au}(0,0), \tag{25}$$

$$G_1'(1) = \frac{\lambda q_0(\lambda q_0 + \nu)\left[\lambda q_0 \mu_2 \rho_{au} + \nu\left(\lambda^2 q_2^2 \mu_1 \delta + \mu_2 \rho_{au} + \lambda q_1 \mu_2^2\right)\right]}{\nu^2 \rho_{au}^2} p_{au}(0,0), \tag{26}$$

$$G_2'(1) = \frac{\lambda q_0 \mu_1 \delta(\lambda q_0 + \nu)[\lambda q_0 \rho_{au} + \nu(\lambda q_2(\mu_1 - \lambda q_1) + \rho_{au} + \lambda q_1 \mu_2)]}{\nu^2 \rho_{au}^2} p_{au}(0,0), \tag{27}$$

where $p_{au}(0,0)$ is given in equation (20).

Let $E(N|I=i)$ be the conditional expected number of customers in the system when the server is in state i. Following Li *et al.* [22], we have

$$E(N|I=0) = \sum_{n=0}^{\infty} np(n|I=0) = \frac{\sum_{n=0}^{\infty} np(0,n)}{G_0(1)} = \frac{G_0'(1)}{G_0(1)}, \tag{28}$$

$$E(N|I=i) = \sum_{n=1}^{\infty} np(n|I=i) = \frac{\sum_{n=1}^{\infty} np(i,n)}{G_i(1)} = \frac{G_i'(1)}{G_i(1)}, \quad i=1,2. \tag{29}$$

Substituting equations (17) and (25) into equation (28), we obtain the following:

$$E(N|I=0) = \frac{\lambda q_0}{\nu} \cdot \tag{30}$$

Using equations (18), (19) and (26), (27), equation (29) can be written as

$$E(N|I=1) = \frac{\lambda q_0 \mu_2 \rho_{au} + \nu\left(\lambda^2 q_2^2 \mu_1 \delta + \mu_2 \rho_{au} + \lambda q_1 \mu_2^2\right)}{\mu_2 \nu \rho_{au}}, \tag{31}$$

$$E(N|I=2) = \frac{\lambda q_0 \rho_{au} + \nu\left(\lambda q_2 \mu_1 - \lambda^2 q_1 q_2 + \rho_{au} + \lambda q_1 \mu_2\right)}{\nu \rho_{au}} \cdot \tag{32}$$

Lemma 1. *Under the stability condition* $\rho_{au} = \mu_1\mu_2 - \lambda(q_1\mu_2 + q_2\mu_1\delta) > 0$*, we know that* $E(N|I=0)$*,* $E(N|I=1)$ *and* $E(N|I=2)$ *are strictly increasing in* $q_0 \in (0,1]$*,* $q_i \in [0,1]$ *for* $i=1,2$*.*

Proof. It is obvious that $E(N|I=0)$ in equation (30) is strictly increasing with respect to q_0, where $q_0 \in (0,1]$. By differentiating equation (31) with respect to q_i for $i=0,1,2$, we obtain

$$\frac{\mathrm{d}E(N|I=1)}{\mathrm{d}q_0} = \frac{\lambda}{\nu} > 0, \qquad q_0 \in (0,1], \tag{33}$$

$$\frac{\mathrm{d}E(N|I=1)}{\mathrm{d}q_1} = \frac{\lambda\mu_2(\rho_{au} + \lambda q_1 \mu_2) + \lambda^3 q_2^2 \mu_1 \delta}{\rho_{au}^2} > 0, \qquad q_1 \in [0,1], \tag{34}$$

$$\frac{\mathrm{d}E(N|I=1)}{\mathrm{d}q_2} = \frac{\lambda^2 \mu_1 \delta\left(2q_2\rho_{au} + q_1\mu_2^2 + \lambda q_2^2 \mu_1 \delta\right)}{\mu_2 \rho_{au}^2} > 0, \qquad q_2 \in [0,1]. \tag{35}$$

In equations (33)–(35), it is easy to see that $\mathrm{d}E(N|I=1)/\mathrm{d}q_i > 0$ for $i=0,1,2$. Thus, $E(N|I=1)$ is strictly increasing with respect to q_i. Similarly, by differentiating equation (32) with respect to q_i, it follows that

$$\frac{\mathrm{d}E(N|I=2)}{\mathrm{d}q_0} = \frac{\lambda}{\nu} > 0, \tag{36}$$

$$\frac{\mathrm{d}E(N|I=2)}{\mathrm{d}q_1} = \frac{\lambda\mu_2(\rho_{au} + \lambda q_1\mu_2) + \lambda^3 q_2^2 \mu_1 \delta}{\rho_{au}^2} > 0, \tag{37}$$

$$\frac{\mathrm{d}E(N|I=2)}{\mathrm{d}q_2} = \frac{\lambda\mu_2(\mu_1 - \lambda q_1)^2 + \lambda^2 q_1 \mu_1 \mu_2 \delta}{\rho_{au}^2} > 0. \tag{38}$$

Based on equations (36)–(38), we determine that $E(N|I=2)$ strictly increases in $q_0 \in (0,1]$, $q_i \in [0,1]$ for $i=1,2$. □

In addition to the expected number of customers, note that sojourn time is often used as a measure of system performance, as it includes both the waiting time and service time (*i.e.*, the total time a customer remains in the system). For convenience, we use $\Upsilon = \mu_1^{-1} + \delta\mu_2^{-1}$ to denote the expected customer service time. Let $W_{au}(i,n)$ be the expected sojourn time of a customer arriving at a system in the state (i,n) and deciding to enter, *i.e.*,

$$W_{au}(0,n) = (n+1)\Upsilon + \frac{1}{\nu}, \qquad n \geq 0, \tag{39}$$

$$W_{au}(1,n) = (n+1)\Upsilon, \qquad n \geq 1, \tag{40}$$

$$W_{au}(2,n) = n\Upsilon + \frac{1}{\mu_2}, \qquad n \geq 1. \tag{41}$$

3.3. Equilibrium strategy

In accordance with the methods outlined by Yu *et al.* [42] and Hao *et al.* [11], it is possible to use the reward-cost structure to evaluate the expected net benefits for a customer deciding to enter a system in state i. The expected benefit functions below:

$$S_0(q_0) = R - C \cdot W_{au}(0, E(N|I=0)) = R - C \cdot \left[(E(N|I=0)+1)\Upsilon + \frac{1}{\nu}\right], \tag{42}$$

$$S_1(q_0,q_1,q_2) = R - C \cdot W_{au}(1, E(N|I=1)) = R - C \cdot [E(N|I=1)+1]\Upsilon, \tag{43}$$

$$S_2(q_0,q_1,q_2) = R - C \cdot W_{au}(2, E(N|I=2)) = R - C \cdot \left[E(N|I=2)\Upsilon + \frac{1}{\mu_2}\right], \tag{44}$$

where $W_{au}(0, E(N|I=0))$, $W_{au}(1, E(N|I=1))$, $W_{au}(2, E(N|I=2))$ are the conditional expected sojourn times given that the server is in state i.

Substituting equations (30)–(32) into equations (42)–(44), we obtain the following:

$$S_0(q_0) = R - C\left(\frac{(\lambda q_0 + \nu)\Upsilon}{\nu} + \frac{1}{\nu}\right), \tag{45}$$

$$S_1(q_0,q_1,q_2) = R - C\left(\frac{\lambda q_0}{\nu} + 1 + \frac{\lambda q_2 \delta(\lambda q_2 \mu_1 \mu_2^{-1} - \mu_1) + \mu_1\mu_2}{\rho_{au}}\right)\Upsilon, \tag{46}$$

$$S_2(q_0,q_1,q_2) = R - C\left(\frac{\lambda q_0}{\nu} + \frac{\lambda q_2(\mu_1 - \lambda q_1 - \mu_1\delta) + \mu_1\mu_2}{\rho_{au}}\right)\Upsilon - \frac{C}{\mu_2}. \tag{47}$$

With the above benefit functions, we refer to the literature [2, 4, 22, 42] and derive the customer equilibrium strategy $(q_e(0), q_e(1), q_e(2))$. If the investigated almost unobservable $M/M/1/MV$ queue with SOS is stable, the following theorems provide the mixed equilibrium strategy wherein customers arriving at a system with the server in state i enter the system with probability $q_e(i)$, $i = 0, 1, 2$.

Theorem 1. *When equation (1) is held, the equilibrium strategy for the server in state 0 is*

$$q_e(0) = q_0^e = \min\left\{\frac{(R - C\Upsilon)\nu - C}{\lambda C\Upsilon}, 1\right\}. \tag{48}$$

Proof. A customer finding the server is on vacation upon arriving at the system would prefer to join the system if the expected net benefit $S_0(q_0) > 0$. Solving $S_0(q_0) = 0$ obtains

$$q_e(0) = \frac{(R - C\Upsilon)\nu - C}{\lambda C\Upsilon}. \tag{49}$$

Since $q_e(0) \in (0,1]$, $q_e(0)$ has an upper bound of 1, thereby completing the proof of Theorem 1. □

Subsequently, let us consider a customer who finds the server in state $i(=1,2)$ upon arrival. We substitute $q_0 = q_0^e$ into equations (46) and (47) and discuss the expected net benefit for new arriving customers. To simplify the proof, we first present a lemma relative to the benefit functions in $S_1(q_0^e, q_1, q_2)$ and $S_2(q_0^e, q_1, q_2)$.

Lemma 2. *The customer decision is significantly relative to the relationship between the service revenue R and the waiting cost C. Given the constants*

$$\tau_1 = C\Upsilon\left(\frac{\lambda q_0^e}{\nu}+1\right)+\frac{C}{\mu_2}, \qquad \tau_2 = C\Upsilon\left(\frac{\lambda q_0^e}{\nu}+1+\frac{\lambda}{\mu_2-\lambda\delta}\right)+\frac{C}{\mu_2},$$
$$\tau_3 = C\Upsilon\left(\frac{\lambda q_0^e}{\nu}+2+\frac{\lambda}{\mu_2}\cdot\frac{\lambda\delta}{\mu_2-\lambda\delta}\right), \quad \tau_4 = C\Upsilon\left(\frac{\lambda q_0^e}{\nu}+\frac{\lambda\delta\mu_1(\lambda-\mu_2)+\mu_1\mu_2^2}{\mu_2\rho_{au}^1}+1\right),$$
$$\rho_{au}^1 = \mu_1\mu_2-\lambda(\mu_2+\mu_1\delta),$$

we have

(1) *If $R<\tau_1$, $S_1(q_0^e,0,0)<0$ and $S_2(q_0^e,0,0)<0$;*
(2) *If $\tau_1<R<\tau_2$, $S_2(q_0^e,0,1)<0$ and $S_2(q_0^e,0,0)>0$;*
(3) *If $\tau_2<R<\tau_3$, $S_1(q_0^e,0,1)<0$ and $S_2(q_0^e,0,1)>0$;*
(4) *If $\tau_3<R<\tau_4$, $S_1(q_0^e,0,1)>0$ and $S_1(q_0^e,1,1)<0$;*
(5) *If $\tau_4<R$, $S_1(q_0^e,1,1)>0$ and $S_2(q_0^e,1,1)>0$.*

Proof. If $S_i(q_0^e,q_1,q_2)$ is greater than 0, a new arriving customer prefers to enter the system. If $S_i(q_0^e,q_1,q_2)$ equals 0, an arriving customer exhibits no preference between entering and balking. If $S_i(q_0^e,q_1,q_2)$ is less than 0, an arriving customer prefers to balk. Consequently, we can conclude that

(1) If $R<\tau_1$, from equations (46) and (47), we have $S_2(q_0^e,0,0)<0$ and $S_2(q_0^e,q_1,q_2)<0$, $\forall q_1,q_2\in[0,1]$ by Lemma 1. Moreover, the assumption $\mu_2>\mu_1(2-\delta)$ implies $S_1(q_0^e,0,0)<S_2(q_0^e,0,0)<0$.
(2) By (1), we have $S_2(q_0^e,0,0)>0$ as $\tau_1<R$. Solving $S_2(q_0^e,0,1)=0$ with respect to R can yield the root τ_2 and thus, $S_2(q_0^e,0,1)<0$ if $R<\tau_2$.
(3) By (2), we have $S_2(q_0^e,0,1)>0$ as $\tau_2<R$. Solving $S_1(q_0^e,0,1)=0$ with respect to R can yield the root τ_3 and thus, $S_1(q_0^e,0,1)<0$ if $R<\tau_3$.
(4) By (3), we have $S_1(q_0^e,0,1)>0$ as $\tau_3<R$. Solving $S_1(q_0^e,1,1)=0$ with respect to R can yield the root τ_4 and thus, $S_1(q_0^e,1,1)<0$ if $R<\tau_4$.
(5) By (4), we have $S_1(q_0^e,1,1)>0$ as $\tau_4<R$. The root of $S_2(q_0^e,1,1)=0$ is

$$\tau_5 = C\Upsilon\left(\frac{\lambda q_0^e}{\nu}+\frac{\lambda(\mu_1-\lambda-\mu_1\delta)+\mu_1\mu_2}{\rho_{au}^1}\right)+\frac{C}{\mu_2}\cdot$$

Under the assumption $\mu_2>\mu_1(2-\delta)$, it can be easily proved that $\tau_5<\tau_4$ after some manipulations. Hence, if $R>\tau_4>\tau_5$, we have $S_2(q_0^e,1,1)>0$. □

Based on Lemmas 1 and 2, the customers' decisions under different combinations of R and C can be explicated derived. In the following, we provide several theorems for the mixed equilibrium strategies.

Theorem 2. *If $R<\tau_1$, the mixed strategy can be specified as $(q_0^e,0,0)$.*

Proof. When $R<\tau_1$, we have $S_2(q_0^e,0,0)<0$ by Lemma 2. Furthermore, Lemma 1 implies that

$$S_2(q_0^e,q_1,q_2)\le S_2(q_0^e,0,0)<0, \text{ for } \forall q_i\in[0,1], \quad i=1,2.$$

Thus, the probability of a new arriving customer joining a system with the server in state 2 is $q_2=0$. On the other hand, $\mu_2>\mu_1(2-\delta)$ indicates

$$S_1(q_0^e,0,0)=R-C\left(\frac{\lambda q_0^e}{\nu}+2\right)\Upsilon<S_2(q_0^e,0,0)<0.$$

Similarly, Lemma 1 implies that

$$S_1(q_0^e, q_1, 0) \le S_1(q_0^e, 0, 0) < 0, \quad \forall q_1 \in [0, 1].$$

It means that the new arriving customer will not join the system in state 1. Merging the above two observations, the mixed strategy is specified by $(q_e(0), q_e(1), q_e(2)) = (q_0^e, 0, 0)$. □

Theorem 3. *If $\tau_2 < R < \tau_3$, the mixed strategy can be specified as $(q_0^e, 0, 1)$.*

Proof. When $\tau_2 < R < \tau_3$, Lemmas 1 and 2 imply $S_2(q_0^e, 0, q_2) \ge S_2(q_0^e, 0, 1) > 0$, $\forall q_2 \in [0, 1]$ and $S_1(q_0^e, q_1, 1) \le S_1(q_0^e, 0, 1) < 0$, $\forall q_1 \in [0, 1]$. Consequently, customers always join the system in state 2 but not in state 1. The mixed strategy is specified by $(q_e(0), q_e(1), q_e(2)) = (q_0^e, 0, 1)$. □

Theorem 4. *If $\tau_4 < R$, the mixed strategy can be specified as $(q_0^e, 1, 1)$.*

Proof. When $\tau_4 < R$, Lemmas 1 and 2 imply $S_1(q_0^e, q_1, q_2) \ge S_1(q_0^e, 1, 1) > 0$, $\forall q_1, q_2 \in [0, 1]$ and thus, customers always join the system in state 1. Furthermore, in this case, $S_2(q_0^e, 1, q_2) \ge S_2(q_0^e, 1, 1) > 0$ indicates customers prefer to join the system in state 2 if all other customers adopt the strategy $q_1 = 1$. Therefore, The mixed strategy is specified by $(q_e(0), q_e(1), q_e(2)) = (q_0^e, 1, 1)$. □

Theorem 5. *If $\tau_1 < R < \tau_2$, the mixed strategy can be specified as $(q_0^e, 0, q_2^*)$.*

Proof. From Theorems 2 and 3, we have the equilibrium decision for customers arriving at the system in state 1 is $q_e(1) = 0$. When $\tau_1 < R < \tau_2$, Lemma 2 implies $S_2(q_0^e, 0, 0) > 0 > S_2(q_0^e, 0, 1)$ and then there is an unique solution for the equation $S_2(q_0^e, 0, q_2) = 0$. The root q_2^* can be solved as

$$q_2^* = \frac{\mu_2[C(\mu_1\delta + \mu_2)(\lambda q_0^e + \nu) - \mu_1\nu(R\mu_2 - C)]}{\lambda[C(\mu_1\delta + \mu_2)(\lambda q_0^e\delta + \delta\nu - \nu) - \mu_1\delta\nu(R\mu_2 - C)]}. \tag{50}$$

Consequently, the mixed strategy is $(q_0^e, 0, q_2^*)$. □

Theorem 6. *If $\tau_3 < R < \tau_4$, the mixed strategy can be specified as $(q_0^e, q_1^*, 1)$.*

Proof. From Theorems 3 and 4, we have the equilibrium decision for customers arriving at the system in state 2 is $q_e(2) = 1$. When $\tau_3 < R < \tau_4$, Lemma 2 implies $S_1(q_0^e, 0, 1) > 0 > S_1(q_0^e, 1, 1)$ and then there is an unique solution for the equation $S_1(q_0^e, q_1, 1) = 0$. The root q_1^* can be derived as

$$q_1^* = \frac{\mu_1\mu_2\nu(\mu_2 - \lambda\delta)(R - 2C\Upsilon) - C\Upsilon\lambda\mu_1[q_0^e\mu_2(\mu_2 - \lambda\delta) + \lambda\delta\nu]}{\lambda\mu_2[R\mu_2\nu - C\Upsilon\mu_2(\nu + \lambda q_0^e)]}. \tag{51}$$

Consequently, the mixed strategy is $(q_0^e, q_1^*, 1)$. □

By estimating the customer's waiting cost and the reward, practitioners can evaluate the joining decision based on the above theorems. After the analysis of the queueing system in an almost unobservable case, we focus on the analysis of customers' balking strategies in the fully unobservable case.

4. Fully unobservable $M/M/1/MV$ queue with SOS

This section examines equilibrium and socially optimal balking strategies of a fully unobservable $M/M/1/MV$ queue with SOS, wherein the customers cannot determine the system size or server state. Here, a strategy is based on a probability q of joining the queue, which results in an effective arrival rate of λq. A corresponding state-transition-rate diagram is shown in Figure 2.

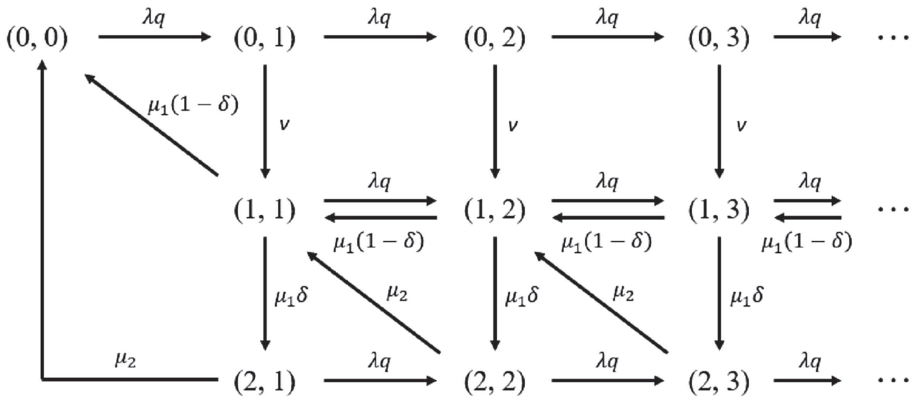


FIGURE 2. State-transition-rate diagram of the fully unobservable $M/M/1/MV$ queue with SOS.

We first derive the expected number of customers in the fully unobservable $M/M/1/MV$ queue with SOS, which is denoted by $E(L_s)$. By substituting $q_0 = q_1 = q_2 = q$ into equation (21), we can refine the stability condition to $\rho_{fu} = \mu_1\mu_2 - \lambda q(\mu_2 + \mu_1\delta) > 0$, and thereby obtain the probability

$$p(0,0) = \frac{\nu\rho_{fu}}{\mu_1\mu_2(\lambda q + \nu)}\cdot \tag{52}$$

Consequently, the expected number of customers in the system is

$$E(L_s) = G_0'(1) + G_1'(1) + G_2'(1) = \lambda q\left(\frac{\mu_2^2 + \lambda q\mu_1\delta}{\mu_2\rho_{fu}} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\right)\cdot \tag{53}$$

Let $W(q)$ be the expected sojourn time of a customer in the fully unobservable $M/M/1/MV$ queue with SOS. Applying Little's formula, $W(q)$ can be easily derived as follows:

$$W(q) = \frac{E(L_s)}{\lambda q} = \frac{\mu_2^2 + \lambda q\mu_1\delta}{\mu_2\rho_{fu}} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\cdot \tag{54}$$

Lemma 3. *Under the stability condition $\rho_{fu} = \mu_1\mu_2 - \lambda q(\mu_2 + \mu_1\delta) > 0$, $W(q)$ strictly increases in $q \in (0, 1]$.*

Proof. By differentiating $W(q)$ with respect to q, we obtain the following:

$$\frac{\mathrm{d}}{\mathrm{d}q}W(q) = \frac{\lambda\mu_1\delta\rho_{fu} - \left(\mu_2^2 + \lambda q\mu_1\delta\right)\rho_{fu}'}{\mu_2\rho_{fu}^2}, \tag{55}$$

which is positive since $\rho_{fu} > 0$ and $\rho_{fu}' = -\lambda(\mu_2 + \mu_1\delta) < 0$. Thus, $\mathrm{d}W(q)/\mathrm{d}q > 0$ and $W(q)$ strictly increases in q, $q \in (0, 1]$. □

If we consider the relationship between the cost to the customer of waiting in the system and the reward that they obtain following the completion of the service, the equilibrium behavior of the customers can be described as follows.

Theorem 7. *In a fully unobservable $M/M/1/MV$ queue with SOS under the stability condition and $\rho_{fu}^1 = \rho_{fu}|_{q=1} = \mu_1\mu_2 - \lambda(\mu_2 + \mu_1\delta)$, the unique equilibrium strategy involves "entering with probability q_e", which is specified by*

$$q_e = \begin{cases} \frac{R\mu_2\nu - C\mu_2(\nu\Upsilon + 1)}{\lambda\left[C\delta\nu\mu_2^{-1} - C(\delta\nu + \mu_2)\Upsilon + R\mu_2\nu\Upsilon\right]}, & C\cdot W(0) < R < C\cdot W(1),\\ 1, & R > C\cdot W(1), \end{cases} \tag{56}$$

where

$$W(0) = \Upsilon + 1/\nu \quad \text{and} \quad W(1) = \frac{\nu\big(\mu_2^2 + \lambda\mu_1\delta\big) + \rho_{fu}^1(\delta\nu + \mu_2)}{\mu_2\nu\rho_{fu}^1}\cdot$$

Proof. To prove this theorem, we refer to the works of Burnetas and Economou [2] and Chen and Zhou [4]. If a new arriving customer decides to enter the system with probability q, then the expected net benefit is

$$S(q) = R - C\cdot W(q) = R - C\left(\frac{\mu_2^2 + \lambda q\mu_1\delta}{\mu_2\rho_{fu}} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\right), \tag{57}$$

where $q \in (0, 1]$. From Lemma 3, it is clear that $S(q)$ strictly decreases in q. Based on equation (57), one can see that

$$S(0) = R - C\left(\Upsilon + \frac{1}{\nu}\right), \tag{58}$$

$$S(1) = R - C\left(\frac{\mu_2^2 + \lambda\mu_1\delta}{\mu_2\rho_{fu}^1} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\right)\cdot \tag{59}$$

From equation (1), we have $R - C\cdot W(0) > 0$. When $R - C\cdot W(1) < 0$, based on the intermediate value theorem, there is a unique root of the equation $S(q) = 0$, $q \in (0, 1]$. Solving $S(q) = 0$, we obtain the solution given by

$$q_e = \frac{R\mu_2\nu - C\mu_2(\nu\Upsilon + 1)}{\lambda\big[C\delta\nu\mu_2^{-1} - C(\delta\nu + \mu_2)\Upsilon + R\mu_2\nu\Upsilon\big]}\cdot \tag{60}$$

By contrast, if $S(1) = R - C\cdot W(1) > 0$, we obtain $S(q) \geq S(1) > 0$ for $\forall q \in (0, 1]$. This means that a new arriving customer always prefers to enter the system, such that the unique equilibrium strategy is $q_e = 1$. □

4.1. Socially optimal strategy

In dressing the socially optimal strategy for a fully unobservable $M/M/1/MV$ queue with SOS, let $\mathrm{SW}(q)$ be the social welfare per unit time for the actual arrival rate λq. Then using equation (57), $\mathrm{SW}(q)$ can be written as follows:

$$\mathrm{SW}(q) = \lambda qR - C\cdot E(L_s) = \lambda q\left[R - C\cdot\left(\frac{\mu_2^2 + \lambda q\mu_1\delta}{\mu_2\rho_{fu}} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\right)\right]\cdot \tag{61}$$

By differentiating equation (61) with respect to q, we obtain

$$\mathrm{SW}'(q) = \lambda R - C\lambda\left[\frac{\mu_1\big(\mu_2^3 + \lambda q\delta\rho_{fu} + \mu_1\mu_2\lambda q\delta\big)}{\mu_2\rho_{fu}^2} + \frac{\delta}{\mu_2} + \frac{1}{\nu}\right]\cdot \tag{62}$$

Solving $\mathrm{SW}'(q) = 0$ implies the root of q as follows:

$$\bar{q} = \frac{1}{\lambda\Upsilon} - \frac{\sqrt{\mu_1\nu C(\mu_1^2\delta + \mu_1\mu_2\delta + \mu_2^2)\{[C(\mu_2 + \delta\nu) - R\mu_2\nu](\mu_2 + \mu_1\delta) - C\mu_1\delta\nu\}}}{\mu_1\lambda\Upsilon\{[C(\mu_2 + \delta\nu) - R\mu_2\nu](\mu_2 + \mu_1\delta) - C\mu_1\delta\nu\}}\cdot \tag{63}$$

In addition, the second derivative of equation (61) with respect to q gives us the following:

$$\mathrm{SW}''(q) = -\frac{2C\lambda^2\mu_1\mu_2}{\rho_{fu}^3}\big[\mu_1^2\delta + \mu_2(\mu_2 + \mu_1\delta)\big] < 0, \tag{64}$$

indicating that $\mathrm{SW}(q)$ is a concave function in q. Thus, $\bar{q}$ is the optimal solution to maximize the social welfare per unit time. Finally, we should consider the socially optimal strategy in a wider context. Let q^* be the socially optimal probability of joining, which ranges from 0 to 1. By summarizing the above results pertaining to socially optimal strategies as $0 \leq \bar{q} \leq 1$, the socially optimal strategy q^* is equal to $\bar{q}$, *i.e.*, $q^* = \bar{q}$. When $\bar{q} > 1$, the socially optimal strategy is $q^* = 1$.

5. Numerical examples

In this section, we initially employ a numerical example to illustrate the theoretical results presented in Sections 3 and 4. Subsequently, we conduct a sensitivity analysis to study the impact of varying parameters on equilibrium and socially optimal entrance probabilities, thereby offering valuable managerial insights. Assuming that the base set of system parameters is $\lambda = 2$, $\mu_1 = 3.5$, $\mu_2 = 6$, $\nu = 1$, $\delta = 0.6$, $R = 6$, $C = 3$, satisfies the stability condition. We separately calculate the numerical results of the expected sojourn time for both the almost unobservable and fully unobservable cases.

From equations (42)–(44), we obtain the conditional expected sojourn time for the almost unobservable case as follows:

$$W_{au}(0, E(N|I=0)) = 0.7714q_0 + 1.3857, \tag{65}$$

$$W_{au}(1, E(N|I=1)) = \frac{97.1964q_0 + 3.2399q_2^2 + 48.5982 - 55.5408q_0q_1 - 19.4393q_0q_2 - 9.7196q_2}{126 - 72q_1 - 25.2q_2} + 0.3857, \tag{66}$$

$$W_{au}(2, E(N|I=2)) = \frac{16.1994q_0 + 1.08q_2 + 8.0997 - 9.2568q_0q_1 - 3.2399q_0q_2 - 1.5428q_1q_2}{21 - 12q_1 - 4.2q_2} + 0.1667. \tag{67}$$

From Theorems 1–6, regarding the reward (R), we have the following five situations, with each corresponding to the equilibrium strategy $(q_e(0), q_e(1), q_e(2))$ as follows:

(1) when $R < 3.5$, $(q_e(0), q_e(1), q_e(2)) = (0.7963, 0, 0)$;
(2) when $3.5 < R < 3.98$, $(q_e(0), q_e(1), q_e(2)) = (0.7963, 0, q_2^*)$, where q_2^* is the root of the equation $R - (63 - 4.5q_2)/(21 - 4.2q_2) - 0.5 = 0$;
(3) when $3.98 < R < 4.25$, $(q_e(0), q_e(1), q_e(2)) = (0.7963, 0, 1)$;
(4) when $4.25 < R < 7.39$, $(q_e(0), q_e(1), q_e(2)) = (0.7963, q_1^*, 1)$, where q_1^* is the root of the equation $R - (312.11 - 132.68q_1)/(100.8 - 72q_1) - 1.1571 = 0$;
(5) when $R > 7.39$, $(q_e(0), q_e(1), q_e(2)) = (0.7963, 1, 1)$.

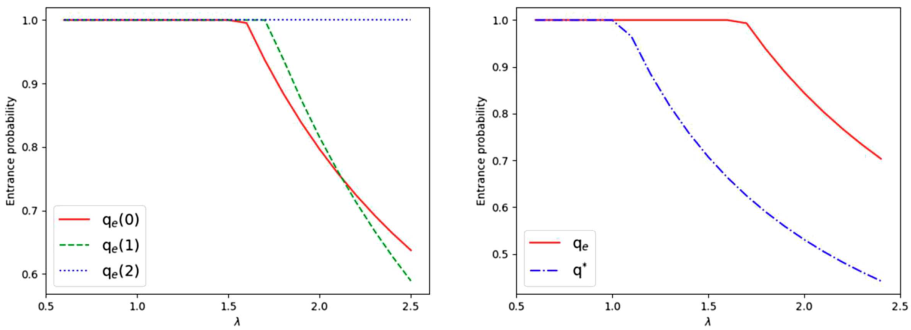
In the base set of system parameters, we have $R = 6$, which is the case (4) mentioned above. Therefore, it can be deduced that $q_1^* = 0.815$.

Equation (54) is the expected sojourn time for the fully unobservable case, which can be calculated as $W(q) = (174.6 - 102.72q)/(126 - 97.2q)$. Then it follows from Theorem 7 that the equilibrium strategy q_e is determined as the root of the equation $R - (523.8 - 308.16q)/(126 - 97.2q) = 0$ when $4.1571 < R < 7.4875$; or it is 1 when $R > 7.4875$. In the default system parameters, $R = 6$; therefore, it can be deduced that $q_e = 0.8442$. According to equations (63) and (64), we can determine the socially optimal strategy $\bar{q} = 0.5308$, which maximizes the social welfare.

5.1. Sensitivity analysis

We present a sensitivity analysis to illustrate the influence of parameters (λ, μ_1, μ_2, ν, δ, R, C) on the equilibrium and socially optimal entrance probabilities. Seven cases are considered as follows:

Case 1. $\mu_1 = 3.5$, $\mu_2 = 6$, $\nu = 1$, $\delta = 0.6$, $R = 6$ and $C = 3$ for various values of λ.
Case 2. $\lambda = 2$, $\mu_2 = 6$, $\nu = 1$, $\delta = 0.6$, $R = 6$ and $C = 3$ for various values of μ_1.
Case 3. $\lambda = 2$, $\mu_1 = 3.5$, $\nu = 1$, $\delta = 0.6$, $R = 6$ and $C = 3$ for various values of μ_2.
Case 4. $\lambda = 2$, $\mu_1 = 3.5$, $\mu_2 = 6$, $\delta = 0.6$, $R = 6$ and $C = 3$ for various values of ν.
Case 5. $\lambda = 2$, $\mu_1 = 3.5$, $\mu_2 = 6$, $\nu = 1$, $R = 6$ and $C = 3$ for various values of δ.
Case 6. $\lambda = 2$, $\mu_1 = 3.5$, $\mu_2 = 6$, $\nu = 1$, $\delta = 0.6$ and $C = 3$ for various values of R.
Case 7. $\lambda = 2$, $\mu_1 = 3.5$, $\mu_2 = 6$, $\nu = 1$, $\delta = 0.6$ and $R = 6$ for various values of C.



(a) (b)

FIGURE 3. Equilibrium and optimal entrance probabilities *versus* λ. (a) Almost unobservable case. (b) Fully unobservable case.

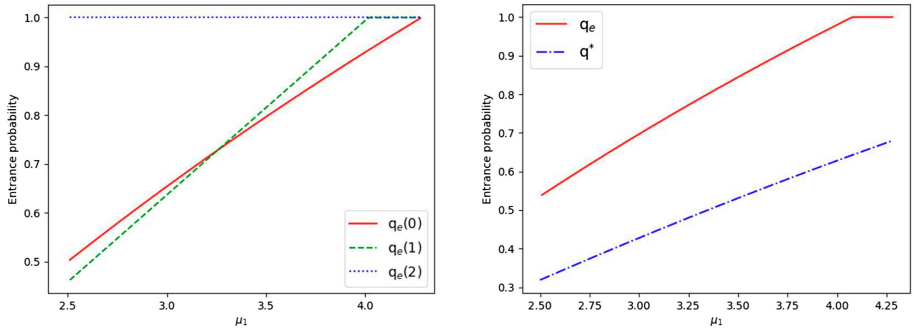
Note that the stability condition is fulfilled in each case. Figures 3–9 respectively present numerical results for Cases 1–7. Figures 3(a)–9(a) show the equilibrium entrance probabilities for the almost unobservable queueing model. Figures 3(b)–9(b) show the equilibrium and optimal entrance probabilities for the fully unobservable queueing model. We can see in Figure 3 that $q_e(0)$, $q_e(1)$, q_e, and q^* decrease with an increase in the mean arrival rate λ. This can be attributed to the fact that a higher λ corresponds to a larger number of customers in the system, which inevitably increases the waiting time for arriving customers. The willingness of customers to enter the system is inversely proportional to waiting cost. As shown in Figures 4 and 5, $q_e(0)$, $q_e(1)$, q_e, and q^* increase with an increase in the mean service rate μ_1 or μ_2. This can be explained by the fact that increasing the service rate will reduce the sojourn time of customers and the corresponding waiting cost, thereby providing an incentive for customers to enter the system. From Figure 6, we can find that (i) $q_e(0)$ and q_e increase with an increase in ν, (ii) $q_e(1)$ first decreases and then increases as ν increases, and (iii) q^* increases as ν increases. These results demonstrate that the mean vacation rate is inversely proportional to the duration of vacations. Reducing the time that customers spend waiting in the system reduces the waiting cost, which increases the willingness of customers to enter the system. The variation in $q_e(1)$ can be attributed to the fact that when the mean vacation rate is lower than the mean service rate, arriving customers are more likely to find the server on vacation and are therefore less likely to enter the system. As shown in Figure 7, $q_e(0)$, $q_e(1)$, q_e and q^* decrease as probability δ increases. The explanation for this is similar to that for the increase in service time. From Figures 8 and 9, we observe that $q_e(0)$, $q_e(1)$, q_e and q^* increase with an increase in R or a decrease in C. Obviously, increasing the service reward or decreasing the waiting cost will increase the net benefit, thereby increasing the probability of entry into the system. As shown in Figures 3–9, $q_e(2)$ remains constant for all parameters, regardless of changes in λ, μ_1, μ_2, ν, δ, R or C. Furthermore, q^* is less than or equal to q_e.

In summary, the theoretical results and numerical calculations outlined in this paper help elucidate the strategies of customers entering a queue, which should benefit managers in formulating rules by which to overcome problems associated with queuing.

5.2. Discussion and managerial insights

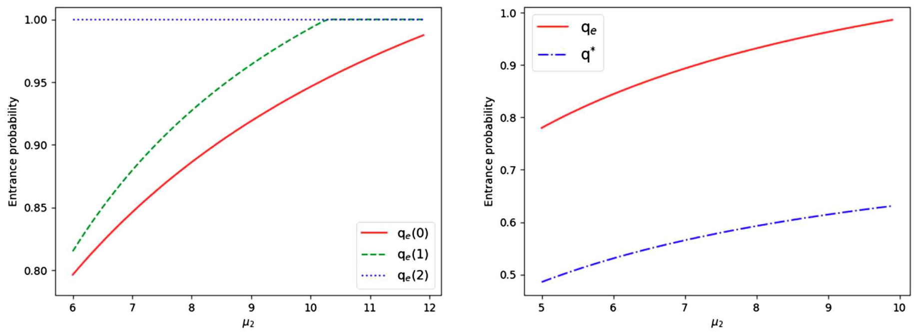
With the above derived mathematical expressions for critical system characteristics, we observe several interesting patterns in our almost and fully unobservable queueing models:

(1) The variation of parameter value influences the waiting time, the expected number of customers, and other system characteristics. The expected waiting time for a customer arriving at the system in state 0, $W_{au}(0, n)$, consists of a vacation period and two-stage service time for queueing customers. Thus, it implies that the joining probability $q_e(0)$ is the most sensitive to parameter changes and is always the first to be changed in



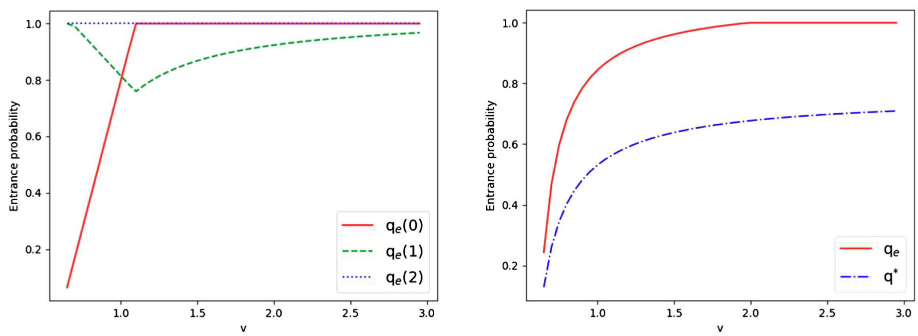
(a) (b)

FIGURE 4. Equilibrium and optimal entrance probabilities *versus* μ_1. (a) Almost unobservable case. (b) Fully unobservable case.



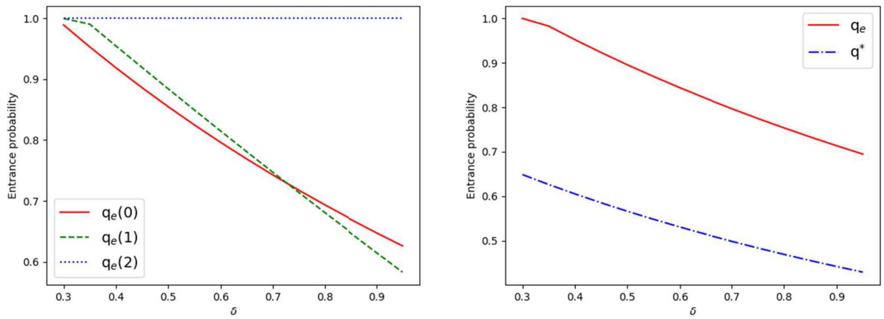
(a) (b)

FIGURE 5. Equilibrium and optimal entrance probabilities *versus* μ_2. (a) Almost unobservable case. (b) Fully unobservable case.



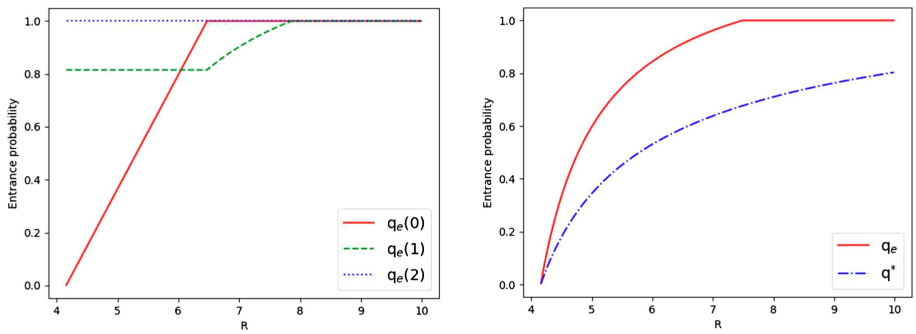
(a) (b)

FIGURE 6. Equilibrium and optimal entrance probabilities *versus* ν. (a) Almost unobservable case. (b) Fully unobservable case.



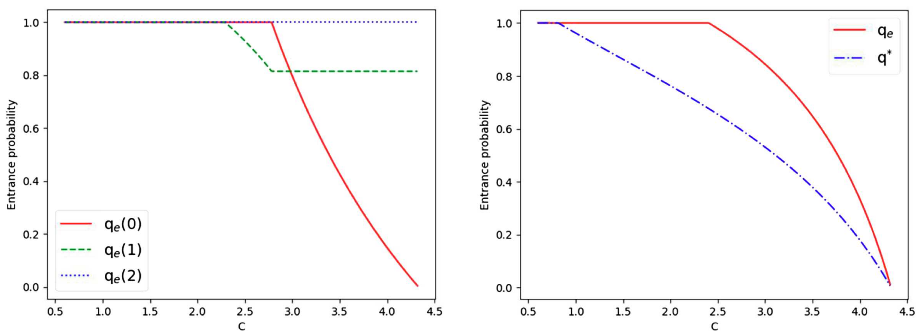
(a) (b)

FIGURE 7. Equilibrium and optimal entrance probabilities *versus* δ. (a) Almost unobservable case. (b) Fully unobservable case.



(a) (b)

FIGURE 8. Equilibrium and optimal entrance probabilities *versus* R. (a) Almost unobservable case. (b) Fully unobservable case.



(a) (b)

FIGURE 9. Equilibrium and optimal entrance probabilities *versus* C. (a) Almost unobservable case. (b) Fully unobservable case.

an almost unobservable model. On the contrary, $q_e(2)$ is insensitive to parameter changes. This phenomenon suggests that managers should focus their analysis on customer balk behavior during vacation rather than busy periods when services are provided.

(2) When the vacation period is very long ($\nu \to 0$), customers arriving at the system in state 0 will not enter the system, *i.e.*, $q_e(0)$ is closae to zero; otherwise, $q_e(1) = q_e(2) = 1$. As ν increases so that $q_e(0)$ increases, the system becomes congested, and customers arriving in state 1, $q_e(1)$, will be less willing to enter the system. However, once $q_e(0)$ achieves the upper bound of 100%, a larger value of ν indicates a shorter vacation period and makes customers more willing to enter the system, *i.e.*, increases of $q_e(1)$. It shows that an appropriate control of the vacation rate can determine the customers' equilibrium entrance probabilities.

(3) In the fully unobservable case, because the equilibrium strategy considers only customer benefit and prefers to enter the system as long as it is beneficial, q_e is no less than q^*. An excessive joining probability only congests the system, especially when the socially optimal strategy takes advantage of customers and service providers. It increases the expected waiting time, which is no additional social welfare. When the system is overburdened with heavy loading, the manager can hide (block) the server state information to discourage customer arrivals.

6. Conclusions and future directions

This paper investigated a queueing model similar to that presented by Yue *et al.* [43]. From a game-theoretic point of view, we studied the strategic behavior of customers in almost and fully unobservable $M/M/1/MV$ queues with SOS. Arriving customers who are not informed about the number of customers in the queue at arrival must decide whether to join the system, based on a linear reward-cost framework. Steady-state analysis was performed to derive equilibrium mixed strategies for both unobservable queues. For the fully unobservable queue, we also addressed the social welfare per unit time to derive the socially optimal joining strategy. Numerical examples demonstrated how equilibrium and socially optimal entrance probabilities are affected by changes in various system parameters, waiting cost and service reward. This paper also provided insights for managers or decision makers seeking to make informed decisions in line with the strategies adopted by their potential customers. The limitations of this study include the assumptions that the server is perfectly reliable and that customers' decisions are made without the possibility of reneging. In the future, we may consider three possible research directions. The first is to examine the queueing model presented in this study under fully and almost observable cases. The second is to investigate corresponding queueing game models considering server breakdowns and customer behaviors. The third is to extend the exploration of the SOS in our model to include multiple optional services.

Acknowledgements

This work was partially supported by the National Science and Technology Council, Taiwan, under Grant Nos. 110-2221-E-141-009-MY2, 112-2221-E-141-009-MY2 and 112-2221-E-A49-124-MY3.

References

[1] A. Aghsami and F. Jolai, Equilibrium threshold strategies and social benefits in the fully observable Markovian queues with partial breakdowns and interruptible setup/closedown policy. *Qual. Technol. Quant. Manage.* **17** (2020) 685–722.

[2] A. Burnetas and A. Economou, Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Syst.* **56** (2007) 213–228.

[3] S.R. Chakravarthy, Analysis of $MAP/PH1$, $PH2/1$ queue with vacations and optional secondary services. *Appl. Math. Modell.* **37** (2013) 8886–8902.

[4] P. Chen and Y. Zhou, Equilibrium balking strategies in the single server queue with setup times and breakdowns. *Oper. Res.* **15** (2015) 213–231.

[5] G. Choudhury, Some aspects of an $M/G/1$ queueing system with optional second service. *Top* **11** (2003) 141–150.

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

[7] A. Economou and A. Manou, Equilibrium balking strategies for a clearing queueing system in alternating environment. *Ann. Oper. Res.* **208** (2013) 489–514.

[8] N.M. Edelson and D.K. Hilderbrand, Congestion tolls for Poisson queuing processes. *Econ.: J. Econ. Soc.* **43** (1975) 81–92.

[9] S. Gao, Availability and reliability analysis of a retrial system with warm standbys and second optional repair service. *Commun. Stat.-Theory Methods* **52** (2023) 1039–1057.

[10] M. Ghorbani-Mandolakani and M.R. Salehi Rad, ML and Bayes estimation in a two-phase tandem queue with a second optional service and random feedback. *Commun. Stat.-Theory Methods* **45** (2016) 2576–2591.

[11] Y. Hao, J. Wang, Z. Wang and M. Yang, Equilibrium joining strategies in the $M/M/1$ queues with setup times under n-policy. *J. Syst. Sci. Syst. Eng.* **28** (2019) 141–153.

[12] R. Hassin, Rational Queueing. CRC Press, Boca Raton (2016).

[13] R. Hassin and M. Haviv, To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems. Vol. 59. Springer Science & Business Media (2003).

[14] M. Haviv and L. Ravner, A survey of queueing systems with strategic timing of arrivals. *Queueing Syst.* **99** (2021) 163–198.

[15] M. Jain and S. Kaur, Bernoulli vacation model for $MX/G/1$ unreliable server retrial queue with bernoulli feedback, balking and optional service. *RAIRO-Oper. Res.* **55** (2021) S2027–S2053.

[16] S. Jin, H. Wu, W. Yue and Y. Takahashi, Performance evaluation and Nash equilibrium of a cloud architecture with a sleeping mechanism and an enrollment service. *J. Ind. Manage. Optim.* **16** (2020) 2407.

[17] J.-C. Ke, C.-H. Wu and Z.G. Zhang, Recent developments in vacation queueing models: a short survey. *Int. J. Oper. Res.* **7** (2010) 3–8.

[18] 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.

[19] D.H. Lee, Optimal pricing strategies and customers' equilibrium behavior in an unobservable $M/M/1$ queueing system with negative customers and repair. *Math. Probl. Eng.* **2017** (2017) 8910819.

[20] K. Li and J. Wang, Equilibrium balking strategies in the single-server retrial queue with constant retrial rate and catastrophes. *Qual. Technol. Quant. Manage.* **18** (2021) 156–178.

[21] L. Li, J. Wang and F. Zhang, Equilibrium customer strategies in Markovian queues with partial breakdowns. *Comput. Ind. Eng.* **66** (2013) 751–757.

[22] X. Li, J. Wang and F. Zhang, New results on equilibrium balking strategies in the single-server queue with breakdowns and repairs. *Appl. Math. Comput.* **241** (2014) 380–388.

[23] K.C. Madan, An $M/G/1$ queue with second optional service. *Queueing Syst.* **34** (2000) 37–46.

[24] S.P. Madheswari, B.K. Kumar and P. Suganthi, Analysis of $M/G/1$ retrial queues with second optional service and customer balking under two types of Bernoulli vacation schedule. *RAIRO-Oper. Res.* **53** (2019) 415–443.

[25] F.A. Maraghi, K.C. Madan and K. Darby-Dowman, Batch arrival vacation queue with second optional service and random system breakdowns. *J. Stat. Theory Pract.* **4** (2010) 137–153.

[26] J. Medhi, A single server Poisson input queue with a second optional channel. *Queueing Syst.* **42** (2002) 239–242.

[27] P. Naor, The regulation of queue size by levying tolls. *Econ.: J. Econ. Soc.* **37** (1969) 15–24.

[28] G. Panda and V. Goswami, Equilibrium joining strategies of positive customers in a Markovian queue with negative arrivals and working vacations. *Methodol. Comput. Appl. Probab.* **24** (2022) 1439–1466.

[29] K. Sun and J. Wang, Equilibrium joining strategies in the single server queues with negative customers. *Int. J. Comput. Math.* **96** (2019) 1169–1191.

[30] W. Sun, S. Li and N. Tian, Equilibrium mixed strategies of customers in an unobservable queue with multiple vacations. *Qual. Technol. Quant. Manage.* **10** (2013) 389–421.

[31] W. Sun, X. Xie, Z. Zhang and S. Li, Customer joining strategies in Markovian queues with B-limited service rule and multiple vacations. *4OR* (2023). DOI: 10.1007/s10288-023-00542-8.

[32] H. Takagi, Queueing Analysis: A Foundation of Performance Evaluation. Vol. I. *Vacation And Priority Systems, Part I.* Elsevier Science Pub. Co., North-Holland, Amsterdam (1991).

[33] R. Tian, Social optimization and pricing strategies in unobservable queues with delayed multiple vacations. *Math. Prob. Eng.* **2019** (2019) 4684957.

[34] N. Tian and Z.G. Zhang, Vacation Queueing Models: Theory and Applications. Vol. 93. Springer Science & Business Media (2006).

[35] R. Tian, S. Su and Z.G. Zhang, Equilibrium and social optimality in queues with service rate and customers' joining decisions. *Qual. Technol. Quant. Manage.* **21** (2024) 1–34.

[36] S. Upadhyaya, Queueing systems with vacation: an overview. *Int. J. Math. Oper. Res.* **9** (2016) 167–213.

[37] P. Vijaya Laxmi and K. Jyothsna, Cost and revenue analysis of an impatient customer queue with second optional service and working vacations. *Commun. Stat.-Simul. Comput.* **51** (2022) 4799–4814.

[38] J. Wang, An $M/G/1$ queue with second optional service and server breakdowns. *Comput. Math. App.* **47** (2004) 1713–1723.

[39] K.-H. Wang, D.-Y. Yang and W.L. Pearn, Comparison of two randomized policy $M/G/1$ queues with second optional service, server breakdown and startup. *J. Comput. Appl. Math.* **234** (2010) 812–824.

[40] J. Wu, Z. Liu and Y. Peng, On the $BMAP/G/1$ G-queues with second optional service and multiple vacations. *Appl. Math. Modell.* **33** (2009) 4314–4325.

[41] B. Xu and X. Xu, Equilibrium strategic behavior of customers in the $M/M/1$ queue with partial failures and repairs. *Oper. Res.* **18** (2018) 273–292.

[42] S. Yu, Z. Liu and J. Wu, Equilibrium strategies of the unobservable $M/M/1$ queue with balking and delayed repairs. *Appl. Math. Comput.* **290** (2016) 56–65.

[43] D. Yue, W. Yue and X. Li, Analysis of a two-phase queueing system with impatient customers and multiple vacations, in The Tenth International Symposium on Operations Research and Its Applications (2011) 292–298.

[44] D. Yue, R. Tian, W. Yue and Y. Qin, Equilibrium strategies in an $M/M/1$ queue with setup times and a single vacation policy, in 11th International Symposium on Operations Research and its Applications in Engineering, Technology and Management 2013. ISORA 2013. IET Digital Library (2013) 152–157.