

Equilibrium balking strategies in the single server queue with setup times and breakdowns

Peishu Chen^{1,2} · Yongwu Zhou¹

Received: 30 September 2014 / Revised: 27 January 2015 / Accepted: 26 February 2015 /
Published online: 11 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This paper studies the equilibrium behavior of customers in the $M/M/1$ queue with setup times, breakdowns and repairs. The server is turned off whenever the system is empty. Once a customer arrives to an empty system, the server begins an exponential setup time to start service again. The lifetime of the server is assumed to be exponentially distributed and once the server breaks down, it will be sent for repair immediately, and the repair time is also exponentially distributed. We consider separately the equilibrium threshold strategies for the fully observable case and mixed strategies for the partially observable and fully unobservable cases. Some numerical examples are presented to illustrate the effect of the information levels and several parameters on the customers' equilibrium and optimal strategies.

Keywords Equilibrium strategies · Queueing · Balking · Setup · Breakdowns · Repairs

1 Introduction

Due to wide applications for management in service systems, there exists an emerging tendency to study customers' behavior impact on the performance of queueing system. Customers in service systems act independently in order to maximize their welfare. However, each customer's optimal behavior is affected by acts taken by the system managers and the other customers. This can be viewed as a game among the customers. Studies about the economic analysis of queueing systems can go back at least to the pioneering work of Naor (1969) who analyzed

✉ Peishu Chen
cps8@sina.com

¹ School of Business Administration, South China University of Technology, Guangzhou 510000, Guangdong, China

² Department of Mathematics, Chaohu University, Chaohu 238000, Anhui, China

customers optimal strategies in an observable $M/M/1$ queue with a simple reward-cost structure. Edelson and Hildebrand (1975) reexamined the work of Naor and considered the unobservable case in which the customers make their decisions without being informed about the state of the queue. Naor's model and results had already been extended by several authors (Chen and Frank 2004; Economou and Kanta 2011). The fundamental results on equilibrium behavior of customers and servers in queueing systems with extensive bibliographical references can be found in the comprehensive monograph of Hassin and Haviv (2003).

Queues with setup times have also been studied by many authors. In such models once the server is reactivated, a random time is required for setup before it can begin serving customers (Reddy et al. 1998; Bischof 2001; Choudhury 2000; Allahverdi et al. 2008). The research on the equilibrium customer behavior in queues with setup times was firstly presented by Burnetas and Economou (2007). Subsequently, Economou and Kanta (2008) considered a Markovian queue that alternates between on and off periods in observable cases.

Server failures which lead to service interruptions are quite common in many real life situations. It is well known that performance measures of unreliable queuing systems are heavily influenced by server failures (Wang et al. 2001; Boudali and Economou 2012; Zhang and Zhu 2013; Zhang et al. 2014).

This motivates us to analyze equilibrium strategies in the single server queue with setup times, breakdowns and repairs. Customers make decisions based on a natural reward-cost structure, which incorporates their desires for service as well as their unwillingness to wait. We consider separately the fully observable case where customers have informed not only about the state of the server but also about the exact number of customers in the system and the partially observable case where an arriving customer knows the state of the server but does not observe the number of customers waiting for service, as well as fully unobservable case where customers do not observe the state of the server and the exact number of customers in the system.

The paper is organized as follows. In Sect. 2, we describe the dynamics of the model and the reward-cost structure. In Sect. 3, we derive the equilibrium threshold strategies for the fully observable case and stationary probabilities of the system. In Sect. 4, we derive the stationary probabilities and equilibrium mixed strategies for the partially observable case. In Sect. 5, we study equilibrium threshold strategies for the fully unobservable case. In Sect. 6, some numerical results are given to illustrate the effect of the information levels and several parameters on the customers' strategies. Finally, some conclusions are given in Sect. 7.

2 Model description

We consider a single-server queue with infinite waiting room in which customers arrive according to a Poisson process with rate λ . The service times are assumed to be exponentially distributed with parameter μ . During the busy periods, the server may break down and once a breakdown occurs, the server will start to be repaired immediately. The lifetime of the server is assumed to be exponentially distributed with parameter ξ , and the repair time is exponentially distributed with parameter η .

The server is deactivated and begins a vacation as soon as the queue becomes empty. When a new customer arrives at an empty system, a setup process starts for the server to be activated. The time required to setup is exponentially distributed with rate θ . In addition, we assume that once customers enter the system, they will not allowed to leave. The customers are not allowed to join the system during the repair time. We further assume that inter-arrival times, service times, setup times, lifetime of the server and repair times are mutually independent.

We denote the state of the system at time t by a random vector $(I(t), N(t))$, where $I(t)$ denotes the state of the server, and $N(t)$ denotes the number of customers in the system. More specifically, $(0, 0)$ implies that the system is deactivated; the vector $(0, n), (1, n), (2, n)$ corresponds to the server in setup, busy or broken periods with n customers in the system. It is clear that the process $\{(I(t), N(t)) : t \geq 0\}$ is a continuous time Markov chain with state space $S = \{0, 0\} \cup (\{0, 1, 2\} \times \{1, 2, \dots, \})$ and the related transition rates are given by

$$\begin{aligned} q_{(0,n)(0,n+1)} &= \lambda, n = 0, 1, 2, \dots, \\ q_{(0,n)(1,n)} &= \theta, n = 1, 2, \dots, \\ q_{(1,1)(0,0)} &= \mu, \\ q_{(1,n)(1,n+1)} &= \lambda, n = 1, 2, \dots, \\ q_{(1,n+1)(1,n)} &= \mu, n = 1, 2, \dots, \\ q_{(1,n)(2,n)} &= \xi, n = 1, 2, \dots, \\ q_{(2,n)(1,n)} &= \eta, n = 1, 2, \dots \end{aligned}$$

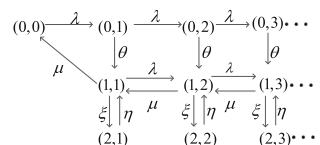
The transition rate diagram is shown in Fig. 1.

In this paper, we assume that every customer receives a reward of R units for completing service. Besides, we assume that there exists a waiting cost of C units per time unit that is continuously accumulated from the time that the customer arrives at the system till the time he leaves the system after being served. From now on, we further assume that

$$R > C \left(\frac{\xi + \eta}{\mu\eta} + \frac{1}{\theta} \right). \tag{1}$$

Once a customer joins an empty system, his mean sojourn time equals the mean setup time plus the mean service time and the mean overall repair time due to failures during the service time, i.e. $\frac{1}{\theta} + \frac{1}{\mu} + \xi \frac{1}{\mu\eta}$. The condition (1) ensures that the reward for service exceeds the expected cost for a customer who finds the system empty. Otherwise, after the system becomes empty for the first time, no customers will ever enter because of the negative net benefit.

Fig. 1 Transition rate diagram of the original model



We focus on studying the behavior of customers when they are allowed to decide whether to join or balk based on the available information upon their arrival, including the queue length and/or the state of the server. We consider separately three information cases in this paper, that is, (1) fully observable case: arriving customers get informed not only about the state of the server but also about the exact number of customers in the system; (2) partially observable case: arriving customers just observe the state of the server and don't know how many customers in the system; (3) fully unobservable case: arriving customers do not observe the state of the system at all. We further assume that decisions are irrevocable: retrials of balking and renegeing of entering customer are not allowed.

3 Equilibrium threshold strategies for the fully observable case

We now focus on the fully observable case, which assumes that customers observe not only the state of the server, but also the exact number of customers in the system. To study the general threshold strategy adopted by all customers in the fully observable case, we will first consider the mean overall sojourn time of a customer who faces different server state upon arrival.

Once the server is activated, the proportion of time that the server is operational is $\frac{\eta}{\xi + \eta}$, thus, the effective service rate is $\frac{\mu\eta}{\xi + \eta}$. A customer who joins the system when he observes state (i, n) has mean sojourn time $T(i, n)$, which equals to

$$T(i, n) = \frac{\mu\eta(n + 1)}{\xi + \eta} + \frac{1 - i}{\theta}, \quad i = 0, 1.$$

His expected net benefit is $B(i, n) = R - CT(i, n)$.

The customer prefers to join if $B(i, n) > 0$; he is indifferent between joining and balking if $B(i, n) = 0$ and prefers to balk if $B(i, n) < 0$. By solving inequality $B(i, n) \geq 0$, we obtain the following results.

Theorem 1 *In the fully observable M/M/1 queue with setup times, server breakdowns and repairs, there exists thresholds*

$$(n_e(0), n_e(1)) = \left(\left\lfloor \frac{(R\theta - C)\mu\eta}{C\theta(\xi + \eta)} \right\rfloor - 1, \left\lfloor \frac{R\mu\eta}{C(\xi + \eta)} \right\rfloor - 1 \right), \quad (2)$$

such that the strategy 'observe $(I(t), N(t))$, enter if $N(t) \leq n_e(I(t))$ and balk otherwise' is a weakly dominant strategy.

We now turn our attention to the social profit problem. For the stationary analysis, note that if all customers follow the threshold strategy in (2), then the system follows a Markov chain $(I(t), N(t))$ with state space

$$S_{fo} = \{(0, n) | 0 \leq n \leq n_e(0) + 1\} \cup \{(i, n) | 1 \leq n \leq n_e(1) + 1, i = 1, 2\}.$$

The transition diagram is depicted in Fig. 2.

The corresponding stationary distribution $(p_{i,j} | (i, j) \in S_{fo})$ is given by the following proposition.

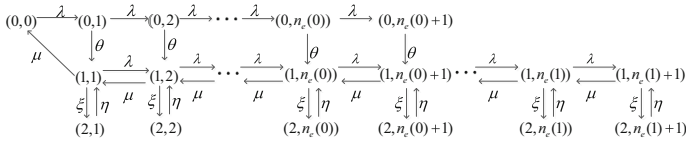


Fig. 2 Transition rate diagram for the $(n_e(0), n_e(1))$ threshold strategy

Proposition 1 Consider an $M/M/1$ queue with setup times, server breakdowns and repairs, $\sigma \neq 1, \rho \neq 1, \sigma \neq \rho$, in which the customers follow the threshold policy $(n_e(0), n_e(1))$. The stationary probabilities $(p_{fo}(i, j) \in S_{fo})$ are as follows:

$$p_{fo}(1, 1) = \eta(1 - \rho)\rho^{1+n_e(0)}(\rho - \sigma)(1 - \sigma) \left\{ \rho^{n_e(0)} \left[(\eta + \xi\rho)(\rho - \sigma) - (\eta + \xi)\rho^{3+n_e(1)} \cdot (1 - \sigma) \right] - (\eta + \xi)\rho \left[\rho^{n_e(0)}(\rho - \sigma) - \rho^{1+n_e(1)}(1 - \sigma) \right] \sigma^{1+n_e(0)} \right\}^{-1}, \tag{3}$$

$$p_{fo}(0, n) = \frac{\sigma^n p_{fo}(1, 1)}{\rho}, \quad n = 0, 1, \dots, n_e(0), \tag{4}$$

$$p_{fo}(0, n_e(0) + 1) = \frac{\sigma^{n_e(0)+1} p_{fo}(1, 1)}{\rho(1 - \sigma)}, \tag{5}$$

$$p_{fo}(1, n) = \frac{(\sigma^n - \rho^n) p_{fo}(1, 1)}{\sigma - \rho}, \quad n = 1, 2, \dots, n_e(0) + 1, \tag{6}$$

$$p_{fo}(1, n) = \frac{(\sigma^{n_e(0)+1} - \rho^{n_e(0)+1}) \rho^{n-n_e(0)-1} p_{fo}(1, 1)}{\sigma - \rho}, \quad n = n_e(0) + 2, \dots, n_e(1) + 1, \tag{7}$$

$$p_{fo}(2, n) = \frac{(\sigma^n - \rho^n) \xi p_{fo}(1, 1)}{\eta(\sigma - \rho)}, \quad n = 1, 2, \dots, n_e(0) + 1, \tag{8}$$

$$p_{fo}(2, n) = \frac{\xi(\sigma^{n_e(0)+1} - \rho^{n_e(0)+1}) \rho^{n-n_e(0)-1} p_{fo}(1, 1)}{\eta(\sigma - \rho)}, \quad n = n_e(0) + 2, \dots, n_e(1) + 1, \tag{9}$$

where

$$\rho = \frac{\lambda}{\mu}, \quad \sigma = \frac{\lambda}{\lambda + \theta}. \tag{10}$$

Proof The stationary distribution is obtained from the following balance equations:

$$\lambda p(0, 0) = \mu p(1, 1), \tag{11}$$

$$(\lambda + \theta)p(0, n) = \lambda p(0, n - 1), \quad n = 1, 2, \dots, n_e(0), \quad (12)$$

$$\lambda p(0, n_e(0)) = \theta p(0, n_e(0) + 1), \quad (13)$$

$$(\lambda + \mu + \xi)p(1, 1) = \theta p(0, 1) + \mu p(1, 2) + \eta p(2, 1), \quad (14)$$

$$(\lambda + \mu + \xi)p(1, n) = \lambda p(1, n - 1) + \theta p(0, n) + \mu p(1, n + 1) + \eta p(2, n), \quad (15)$$

$$n = 2, 3, \dots, n_e(0),$$

$$(\lambda + \mu + \xi)p(1, n) = \lambda p(1, n - 1) + \mu p(1, n + 1) + \eta p(2, n), \quad (16)$$

$$n = n_e(0) + 2, \dots, n_e(1),$$

$$(\mu + \xi)p(1, n_e(1) + 1) = \lambda p(1, n_e(1)) + \eta p(n_e(1) + 1, 2), \quad (17)$$

$$\eta p(2, n) = \xi p(1, n), \quad n = 1, 2, \dots, n_e(1) + 1. \quad (18)$$

By iterating (12) and (16), taking into account (11), (13), (17), (18) and (10), we can easily obtain expressions (4), (5), (6), (7) in Proposition 1, and the following equation:

$$p_{fo}(2, n) = \frac{\xi p_{fo}(1, n)}{\eta}, \quad n = 1, \dots, n_e(1) + 1. \quad (19)$$

From (14), (18) and (10), we get

$$p(1, 2) = (\rho + \sigma)p(1, 1). \quad (20)$$

From (15), (18) and (4), we have

$$\mu p(1, n + 1) - (\lambda + \mu)p(1, n) + \lambda p(1, n - 1) = -\frac{\theta \sigma^n p_{fo}(1, 1)}{\rho}, \quad n = 2, 3, \dots, n_e(0). \quad (21)$$

Combining (20) and the recursive formulas (21), we can obtain (6) by tedious calculations. Then, with the help of the normalizing equation

$$\sum_{n=0}^{n_e(0)+1} p_{fo}(0, n) + \sum_{n=1}^{n_e(1)+1} p_{fo}(1, n) + \sum_{n=1}^{n_e(1)+1} p_{fo}(2, n) = 1,$$

we get the expression of $p_{fo}(1, 1)$. \square

Lemma 1 *Proposition 1 holds for the stationary distribution corresponding to any threshold policy $(n_e(0), n_e(1))$ and not merely to the individually optimal policy specified by (2).*

Because an arrival balks once he finds the system at state $(0, n_e(0) + 1)$ or $(1, n_e(1) + 1)$ or the server is broken, then the effective arrival rate of the system is given by

$$\lambda(n_e(0), n_e(1)) = \lambda \left[\sum_{n=0}^{n_e(0)} p_{fo}(0, n) + \sum_{n=1}^{n_e(1)} p_{fo}(1, n) \right].$$

Hence, the social benefit per time unit when all customers follow the threshold policy $(n_e(0), n_e(1))$ equals

$$S_{fo} = \lambda R \left[\sum_{n=0}^{n_e(0)} p_{fo}(0, n) + \sum_{n=1}^{n_e(1)} p_{fo}(1, n) \right] - C \left[\sum_{n=0}^{n_e(0)+1} n p_{fo}(0, n) + \sum_{n=1}^{n_e(1)+1} n p_{fo}(1, n) + \sum_{n=1}^{n_e(1)+1} n p_{fo}(2, n) \right].$$

4 Equilibrium mixed strategies for the partially observable case

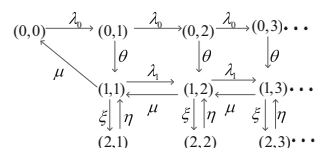
In this section, we turn our attention to the partially observable case, where arrivals only observe the server state, and do not observe the number of customers in the system. We will prove that there exists equilibrium mixed strategies. A mixed strategy for a customer is specified by a vector (q_0, q_1) , where $q_i = q(i)$ is the probability of joining when the server is in state i . If all customers follow the same mixed strategy (q_0, q_1) , then the system follows a Markov chain similar to that described in Fig. 1 except that the arrival rate equals to $\lambda_i = \lambda q_i$. The state space S_{po} for the partially observable case is identical to the original state space S , and transition diagram is illustrated in Fig. 3.

Proposition 2 Consider an $M/M/1$ queue with setup times, server breakdowns and $\sigma_0 \neq \rho_1$, in which the customers enter the system with probability q_i if observe the server in state $i (i = 0, 1)$ upon arrival, and never enter the system whenever the server is broken. The system is stable if and only if $\rho_1 < 1$ and the stationary probabilities of the system are given by:

$$p_{po}(0, 0) = \frac{\eta(1 - \rho_1)(1 - \sigma_0)}{(\eta + \xi)\rho_0 + \eta(1 - \rho_1)}, \tag{22}$$

$$p_{po}(0, n) = \frac{\eta(1 - \rho_1)(1 - \sigma_0)\sigma_0^n}{(\eta + \xi)\rho_0 + \eta(1 - \rho_1)}, \quad n = 0, 1, 2, \dots, \tag{23}$$

Fig. 3 Transition rate diagram for the (q_0, q_1) mixed strategy



$$p_{po}(1, n) = \frac{\rho_0 \eta (1 - \rho_1) (1 - \sigma_0) (\sigma_0^n - \rho_1^n)}{(\sigma_0 - \rho_1) [\eta + (\eta + \xi) \rho_0 - \eta \rho_1]}, \quad n = 1, 2, 3, \dots, \quad (24)$$

$$p_{po}(2, n) = \frac{\rho_0 \xi (1 - \rho_1) (1 - \sigma_0) (\sigma_0^n - \rho_1^n)}{(\sigma_0 - \rho_1) [\eta + (\eta + \xi) \rho_0 - \eta \rho_1]}, \quad n = 1, 2, 3, \dots, \quad (25)$$

where

$$\rho_0 = \frac{\lambda_0}{\mu}, \rho_1 = \frac{\lambda_1}{\mu}, \sigma_0 = \frac{\lambda_0}{\lambda_0 + \theta}. \quad (26)$$

Proof Let $(p_{po}(i, n) : (i, n) \in S)$ be the stationary distribution of the corresponding system. The balance equations are presented below:

$$\lambda_0 p(0, 0) = \mu p(1, 1), \quad (27)$$

$$(\lambda_0 + \theta) p(0, n) = \lambda_0 p(0, n - 1), \quad n = 1, 2, 3, \dots, \quad (28)$$

$$(\lambda_1 + \mu + \xi) p(1, 1) = \theta p(0, 1) + \mu p(1, 2) + \eta p(2, 1), \quad (29)$$

$$(\lambda_1 + \mu + \xi) p(1, n) = \lambda_1 p(1, n - 1) + \theta p(0, n) + \mu p(1, n + 1) + \eta p(2, n), \quad (30)$$

$$n = 2, 3, \dots,$$

$$\eta p(n, 1) = \xi p(1, n), \quad n = 1, 2, 3, \dots \quad (31)$$

By iterating (28), we obtain

$$p_{po}(0, n) = \sigma_0^n p_{po}(0, 0), \quad n = 0, 1, 2, \dots \quad (32)$$

From (28)–(32), we can get

$$p(1, 1) = \rho_0 p(0, 0), \quad (33)$$

$$p(1, 2) = \rho_0 (\sigma_0 + \rho_1) p(0, 0), \quad (34)$$

$$\mu p(1, n + 1) - (\lambda_1 + \mu) p(1, n) + \lambda_1 p(1, n - 1) = -\theta \sigma_0^n p(0, 0), \quad n = 2, 3, \dots \quad (35)$$

Combining (33)–(34) and the recursive formulas (35), we obtain the following equation by tedious calculations

$$p_{po}(1, n) = \frac{\rho_0 (\sigma_0^n - \rho_1^n) p(0, 0)}{\sigma_0 - \rho_1}, \quad n = 1, 2, 3, \dots \quad (36)$$

From (31), we get

$$p_{po}(2, n) = \frac{\xi p_{po}(1, n)}{\eta}, \quad n = 1, 2, 3, \dots \quad (37)$$

Putting (33), (36) and (37) into the normalizing equation

$$p_{po}(0, 1) + p_{po}(1, 1) + p_{po}(2, 1) = 1,$$

we obtain the expression of $p_{po}(0, 0)$ in (22). From the expression $p_{po}(0, 0) > 0$, we immediately get $\rho_1 < 1$. The conclusions (23)–(25) can be easily derived based on the above expressions (32), (36) and (37). \square

From Proposition 2, we can obtain the steady state probabilities of the server

$$p(i = 0) = \sum_{n=0}^{+\infty} p_{po}(0, n) = \frac{\eta(1 - \rho_1)}{\eta(1 - \rho_1) + (\eta + \xi)\rho_0}, \tag{38}$$

$$p(i = 1) = \sum_{n=1}^{+\infty} p_{po}(1, n) = \frac{\eta\rho_0}{\eta(1 - \rho_1) + (\eta + \xi)\rho_0}. \tag{39}$$

Let $E(N^-|i)$ be the expected number of customers in the system found by an arrival, given that the server is found at state i , then we have

$$E(N^-|0) = \frac{\sum_{n=0}^{+\infty} np_{po}(0, n)}{p(i = 0)} = \frac{\sigma_0}{1 - \sigma_0}, \tag{40}$$

$$E(N^-|1) = \frac{\sum_{n=1}^{+\infty} np_{po}(1, n)}{p(i = 1)} = \frac{1}{1 - \sigma_0} + \frac{\rho_1}{1 - \rho_1}. \tag{41}$$

Thus we can get a customer who finds the server at state i upon arrival, his mean sojourn time is

$$S(i) = \frac{(\xi + \eta)[E(N^-|i) + 1]}{\mu\eta} + \frac{1 - i}{\theta}, \tag{42}$$

the expected net benefit of such customer who decides to enter is

$$B(i) = R - C \cdot S(i) = R - C \left\{ \frac{(\xi + \eta)[E(N^-|i) + 1]}{\mu\eta} + \frac{1 - i}{\theta} \right\}. \tag{43}$$

Then the social benefit per time unit when all customers follow the mixed policy (q_0, q_1) can be eventually computed as:

$$\begin{aligned} S_{po} &= \lambda_0 \cdot p(i = 0) \cdot B(0) + \lambda_1 \cdot p(i = 1) \cdot B(1) \\ &= \lambda_0 \cdot \frac{\eta(1 - \rho_1)}{\eta(1 - \rho_1) + (\eta + \xi)\rho_0} \cdot \left\{ R - \frac{C(\xi + \eta)[E(N^-|0) + 1]}{\mu\eta} - \frac{C}{\theta} \right\} \\ &\quad + \lambda_1 \cdot \frac{\eta\rho_0}{\eta(1 - \rho_1) + (\eta + \xi)\rho_0} \cdot \left\{ R - \frac{C(\xi + \eta)[E(N^-|1) + 1]}{\mu\eta} \right\} \end{aligned} \tag{44}$$

Theorem 2 *In the fully observable M/M/1 queue with setup times, server breakdowns and repairs, there exists a unique mixed equilibrium strategy $(q_e(0), q_e(1))$ ‘observe $I(t)$ and enter with probability $q_e(I(t))$ ’ when condition*

$\lambda < \mu$ holds, where the vector $(q_e(0), q_e(1))$ is given as follows.

Case I: $\frac{1}{\theta} < \frac{\xi + \eta}{\mu\eta}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, 0 \right) \right) & \text{if } R \in \left(\frac{C(\xi + \eta)}{\mu\eta} + \frac{C}{\theta}, \frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta} \right); \\ (1, 0) & \text{if } R \in \left[\frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta}, \frac{C(\xi + \eta)}{\mu\eta} \left(1 + \frac{\lambda + \theta}{\theta} \right) \right); \\ \left(1, \frac{\mu[R\mu\eta\theta - C(\lambda + 2\theta)(\eta + \xi)]}{\lambda[R\mu\eta\theta - C(\lambda + \theta)(\eta + \xi)]} \right); \\ \text{if } R \in \left[\frac{C(\xi + \eta)}{\mu\eta} \left(1 + \frac{\lambda + \theta}{\theta} \right), \frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\mu}{\mu - \lambda} + \frac{\lambda + \theta}{\theta} \right) \right); \\ (1, 1) & \text{if } R \in \left[\frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\mu}{\mu - \lambda} + \frac{\lambda + \theta}{\theta} \right), +\infty \right). \end{cases}$$

Case II: $\frac{\xi + \eta}{\mu\eta} \leq \frac{1}{\theta} \leq \frac{\xi + \eta}{(\mu - \lambda)\eta}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, \frac{\eta\mu - \theta(\xi + \eta)}{\lambda\eta} \right) \right) & \text{if } R \in \left(\frac{C(\xi + \eta)}{\mu\eta} + \frac{C}{\theta}, \frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta} \right); \\ \left(1, \frac{\mu[R\mu\eta\theta - C(\lambda + 2\theta)(\eta + \xi)]}{\lambda[R\mu\eta\theta - C(\lambda + \theta)(\eta + \xi)]} \right) & \text{if } R \in \left[\frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta}, \frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\mu}{\mu - \lambda} + \frac{\lambda + \theta}{\theta} \right) \right); \\ (1, 1) & \text{if } R \in \left[\frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\mu}{\mu - \lambda} + \frac{\lambda + \theta}{\theta} \right), +\infty \right). \end{cases}$$

Case III: $\frac{1}{\theta} > \frac{\xi + \eta}{(\mu - \lambda)\eta}$,

$$(q_e(0), q_e(1)) = \begin{cases} \left(\left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, 1 \right) \right) & \text{if } R \in \left(\frac{C(\xi + \eta)}{\mu\eta} + \frac{C}{\theta}, \frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta} \right); \\ (1, 1) & \text{if } R \in \left[\frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta}, +\infty \right). \end{cases}$$

Proof Consider a tagged customer who finds the server at state 0 upon arrival. From equations (40) and (43), if he decides to enter, his expected net benefit is

$$B(0) = R - \frac{C}{\mu\eta\theta} (\xi + \eta)(\lambda q(0) + \theta) - \frac{C}{\theta},$$

Then we discuss the values of $B(0)$ in two cases below:

Case 1: $B(0) \leq 0$, i.e., $\frac{C(\xi + \eta)}{\mu\eta} + \frac{C}{\theta} < R \leq \frac{C(\lambda + \theta)(\xi + \eta)}{\mu\eta\theta} + \frac{C}{\theta}$.

In this case, if all customers who find the system empty enter with probability $q_e(0) = 1$, then the tagged customer suffers a negative expected benefit if he decides to enter. Hence $q_e(0) = 1$ does not lead to equilibrium. Similarly, if all customers use $q_e(0) = 0$, then the tagged customer receives a positive benefit from entering, thus $q_e(0) = 0$ also cannot be part of an equilibrium mixed strategy. Therefore, there exists a unique $q_e(0)$ satisfying

$$R - \frac{C}{\mu\eta\theta} (\xi + \eta)(\lambda q(0) + \theta) - \frac{C}{\theta} = 0,$$

for which customers are indifferent between entering and balking. Solving the above linear equation with respect to $q(0)$, we obtain

$$q_e(0) = \frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}.$$

Case 2: $B(0) > 0$, i.e. $R > \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$.

In this case, for every strategy of the other customers, the tagged customer has a positive expected net benefit if he decides to enter, so the best response is $q_e(0) = 1$.

We next consider $q_e(1)$ and tag a customer who finds the server at state 1 upon arrival. From (41) and (43), if he decides to enter his expected net benefit is equal to

$$B(1) = R - \frac{C(\xi + \eta)}{\mu\eta} \left(\frac{1}{1 - \sigma_0} + \frac{\rho_1}{1 - \rho_1} + 1 \right) = R - \frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\lambda_0 + \theta}{\theta} + \frac{\mu}{\mu - \lambda_1} \right),$$

so

$$B(1) = \begin{cases} \frac{C}{\theta} - \frac{C(\xi + \eta)}{\eta(\mu - \lambda q(1))} & \text{in Case 1,} \\ R - \frac{C(\xi + \eta)}{\mu\eta} \left(\frac{\lambda + \theta}{\theta} + \frac{\mu}{\mu - \lambda q(1)} \right) & \text{in Case 2.} \end{cases}$$

Therefore, to find $q_e(1)$ in equilibrium, we must examine Cases 1 and 2 separately and consider the following subcases in each.

Case 1a: $\frac{C(\xi+\eta)}{\mu\eta} + \frac{C}{\theta} < R \leq \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $\frac{C}{\theta} < \frac{C(\xi+\eta)}{\mu\eta}$, we have

$$(q_e(0), q_e(1)) = \left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, 0 \right).$$

Case 1b: $\frac{C(\xi+\eta)}{\mu\eta} + \frac{C}{\theta} < R \leq \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $\frac{C(\xi+\eta)}{\mu\eta} \leq \frac{C}{\theta} \leq \frac{C(\xi+\eta)}{(\mu-\lambda)\eta}$, we have

$$q_e(0), q_e(1) = \left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, \frac{\eta\mu - \theta(\xi + \eta)}{\lambda\eta} \right).$$

Case 1c: $\frac{C(\xi+\eta)}{\mu\eta} + \frac{C}{\theta} < R \leq \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $\frac{C}{\theta} > \frac{C(\xi+\eta)}{(\mu-\lambda)\eta}$, we have

$$(q_e(0), q_e(1)) = \left(\frac{R\eta\theta\mu - C[\eta(\theta + \mu) + \theta\xi]}{\lambda(\eta + \xi)C}, 1 \right).$$

Case 2a: $R > \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $R < \frac{C(\xi+\eta)}{\mu\eta} (1 + \frac{\lambda+\theta}{\theta})$, we have $(q_e(0), q_e(1)) = (1, 0)$.

Case 2b: $R > \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $\frac{C(\xi+\eta)}{\mu\eta} (\frac{\lambda+\theta}{\theta} + 1) \leq R \leq \frac{C(\xi+\eta)}{\mu\eta} (\frac{\lambda+\theta}{\theta} + \frac{\mu}{\mu-\lambda})$, we have

$$(q_e(0), q_e(1)) = \left(1, \frac{\mu[R\mu\eta\theta - C(\lambda + 2\theta)(\eta + \xi)]}{\lambda[R\mu\eta\theta - C(\lambda + \theta)(\eta + \xi)]} \right).$$

Case 2c: $R > \frac{C(\lambda+\theta)(\xi+\eta)}{\mu\eta\theta} + \frac{C}{\theta}$ and $R > \frac{C(\xi+\eta)}{\mu\eta} (\frac{\lambda+\theta}{\theta} + \frac{\mu}{\mu-\lambda})$, we have $(q_e(0), q_e(1)) = (1, 1)$.

By rearranging Cases 1a–2c as R varies from $\frac{C(\xi+\eta)}{\mu\eta} + \frac{C}{\theta}$ to infinity, keeping the operating parameters $\lambda, \mu, \theta, \xi, \eta$ and the waiting cost rate C fixed, we obtain Cases I–III in the theorem statement. □

Theorem 2 quantifies several expected situations about the behavior of the customers when they are informed about the server’s state upon arrival. The key factor in the theorem is the mean setup time $\frac{1}{\theta}$ in comparison with the mean generalized service time (the mean service time plus the mean overall repair time due to failures during the service time $\frac{\xi+\eta}{\mu\eta}$) and $\frac{\xi+\eta}{(\mu-\lambda)\eta}$. As Theorem 2 shows: it is always true that $q_e(0) \geq q_e(1)$ in case I, $q_e(0) \leq q_e(1)$ in case III and the situation varies in case II. If a tagged customer is given the information that the server is on setup and the mean setup time is small, it is optimal for the tagged customer to enter if few customers ahead of him, which belongs to case I. On the contrary, if a tagged customer is given the information that the server is on setup, the mean setup time is large and many customers ahead of him, this would mean that he may wait for a long time, then his optimal response is balking, which should be in accordance with case III. When the mean setup time is relatively moderate, the inequality relations of $q_e(0)$ and $q_e(1)$ is various, which should be in accordance with case II.

5 Equilibrium threshold strategies for the fully unobservable case

We finally study the fully unobservable case where the newly arriving customers are not admitted to join the system when the server is broken. Here a mixed strategy for a customer is specified by the probability q of entering, where q is the probability of joining the queue whenever the server is in a regular busy period or a setup period. The stationary distribution of the system state is given by Proposition 2 by taking $q_0 = q_1 = q$. The equilibrium behavior of the customers is described in the following.

Proposition 3 *In the fully unobservable M/M/1 queue with setup times, server breakdowns and repairs, where the newly arriving customers are not admitted to join the system when the server is broken. The expected mean sojourn time of a customer who decides to enter is given by*

$$E(W) = \frac{\eta + \xi}{\eta(\mu - \lambda q)} + \frac{1}{\theta} + \frac{\lambda q \xi}{\eta \theta \mu}. \tag{45}$$

Proof By taking $q_0 = q_1 = q$ in (23)–(25), then we have the mean number of customers in the system as follows:

$$\begin{aligned} E(N) &= \sum_{n=0}^{+\infty} np(0, n) + \sum_{n=1}^{+\infty} np(1, n) + \sum_{n=1}^{+\infty} np(2, n) \\ &= \lambda q \left(\frac{1}{\theta} - \frac{\mu(\eta + \xi)}{(\lambda q - \mu)(\eta\mu + \lambda q \xi)} \right), \end{aligned}$$

and the server is broken with probability

$$p(i = 2) = \sum_{n=1}^{+\infty} p(2, n) = \frac{\lambda q \xi}{\eta \mu + \lambda q \xi},$$

since the newly arriving customers are not admitted to join the system when the server is broken, then the effective arrival rate is $\lambda q(1 - p(i = 2))$. By using Little’s law, we have the expected mean sojourn time of a customer who decides to enter the system is

$$E(W) = \frac{E(N)}{\lambda q(1 - p(i = 2))} = \frac{\eta + \xi}{\eta(\mu - \lambda q)} + \frac{1}{\theta} + \frac{\lambda q \xi}{\eta \theta \mu}.$$

□

The social benefit per time when all allowed joining customers follow a mixed strategy q is given by

$$\begin{aligned} S_{un} &= \lambda q(1 - p(i = 2))(R - CE(W)) \\ &= \frac{\lambda q \eta \mu}{\eta \mu + \lambda q \xi} (R - CE(W)). \end{aligned} \tag{46}$$

Theorem 3 *In the fully unobservable M/M/1 queue with setup times, server breakdowns and repairs, where the newly arriving customers are not admitted to join the system when the server is under repair. there exists a mixed equilibrium strategy ‘enter with probability q_e ’, where q_e is given by*

$$q_e = \begin{cases} q_e^*, & \text{if } R \in \left(C\left(\frac{\eta + \xi}{\eta \mu} + \frac{1}{\theta}\right), C\left(\frac{\lambda \xi}{\eta \theta \mu} + \frac{\eta + \xi}{\eta(\mu - \lambda)} + \frac{1}{\theta}\right) \right); \\ 1, & \text{if } R \in \left[C\left(\frac{\lambda \xi}{\eta \theta \mu} + \frac{\eta + \xi}{\eta(\mu - \lambda)} + \frac{1}{\theta}\right), +\infty \right). \end{cases} \tag{47}$$

where

$$q_e^* = \frac{U + CF\mu - \sqrt{4C^2FV + (U - CF\mu)^2}}{2CF\lambda}, \quad U = R - \frac{C}{\theta}, \quad V = \frac{\eta + \xi}{\eta}, \quad F = \frac{\xi}{\eta \theta \mu}.$$

Proof We consider a tagged customer at his arrival instant, if he is allowed to join the system and decides to enter with probability q , his expected net benefit is

$$B(q) = R - CE(W) = R - C\left(\frac{\eta + \xi}{\eta(\mu - \lambda q)} + \frac{1}{\theta} + \frac{\lambda q \xi}{\eta \theta \mu}\right), \tag{48}$$

where $q \in [0, 1]$. It is easy to see that $B(q)$ is strictly decreasing for $q \in [0, 1]$. We can get

$$B(0) = R - C\left(\frac{\eta + \xi}{\eta\mu} + \frac{1}{\theta}\right),$$

$$B(1) = R - C\left(\frac{\lambda\xi}{\eta\theta\mu} + \frac{\eta + \xi}{\eta(\mu - \lambda)} + \frac{1}{\theta}\right).$$

When $R \in \left(C\left(\frac{\eta + \xi}{\eta\mu} + \frac{1}{\theta}\right), C\left(\frac{\lambda\xi}{\eta\theta\mu} + \frac{\eta + \xi}{\eta(\mu - \lambda)} + \frac{1}{\theta}\right)\right)$, we find that (48) has a unique root in $(0, 1)$, which is given by the first branch of (47). When $R \geq C\left(\frac{\lambda\xi}{\eta\theta\mu} + \frac{\eta + \xi}{\eta(\mu - \lambda)} + \frac{1}{\theta}\right)$, $B(q)$ is positive for every q , then the unique equilibrium point is $q_e = 1$, which gives the second branch of (47). \square

6 Numerical results

In this section, we will present a set of numerical experiments that are based on the analytical results of this article. In particular, we concern about the values of the equilibrium thresholds for the fully observable case, the values of the equilibrium entrance probabilities for the partially observable and fully unobservable cases as well as the equilibrium social benefit per unit time when customers follow equilibrium strategies.

We first consider the fully observable case and explore the sensitivity of the equilibrium pure thresholds policies $(n_e(0), n_e(1))$ with respect to the service reward R , setup rate θ , and the ratio of parameters ξ and η . The results are presented in Fig. 4a–c. We find that when adding parameter R , the thresholds $(n_e(0), n_e(1))$ increase in a linear fashion, up to the integrality requirement; when the setup rate θ varies, the threshold $n_e(0)$ increases, while $n_e(1)$ remains constant. Finally, when ξ/η increases, it means that the lifetime of the system becomes shorter, and the customers in the system need to wait longer for the server to restore, which leads to new customers leave the system and the thresholds decrease.

We then turn to the partially observable and fully unobservable cases and investigate the sensitivity of the equilibrium entrance probabilities $q_e(0), q_e(1)$ and q_e . From the results of Fig. 5a–c, we find that the entrance probability q_e in the fully unobservable case is always inside the interval formed by $q_e(0)$ and $q_e(1)$ in the partial observable case. The entrance probabilities $q_e(0), q_e(1)$ and q_e are expected to increase with respect to reward R . The equilibrium entrance probabilities $q_e(0), q_e(1)$ and q_e are nondecreasing with θ for the most part, and there is a range of small values of θ in $q_e(1)$ is decreasing. The entrance probabilities $q_e(0), q_e(1)$ and q_e are expected to decrease with respect to ξ/η .

The last three numerical experiments are concerned with the social benefit under the equilibrium strategy for fully observable, partially observable and fully unobservable cases. Generally speaking, the equilibrium social benefit is increasing with respect to the reward R in Fig. 6a. The two curves of fully observable and partial observable cases are getting closer with the increasing of R . The equilibrium social benefit is decreasing with respect to the breakdown rate ξ in Fig. 6b and increasing with respect to the repair rate η in Fig. 6c. In our system, we find that the

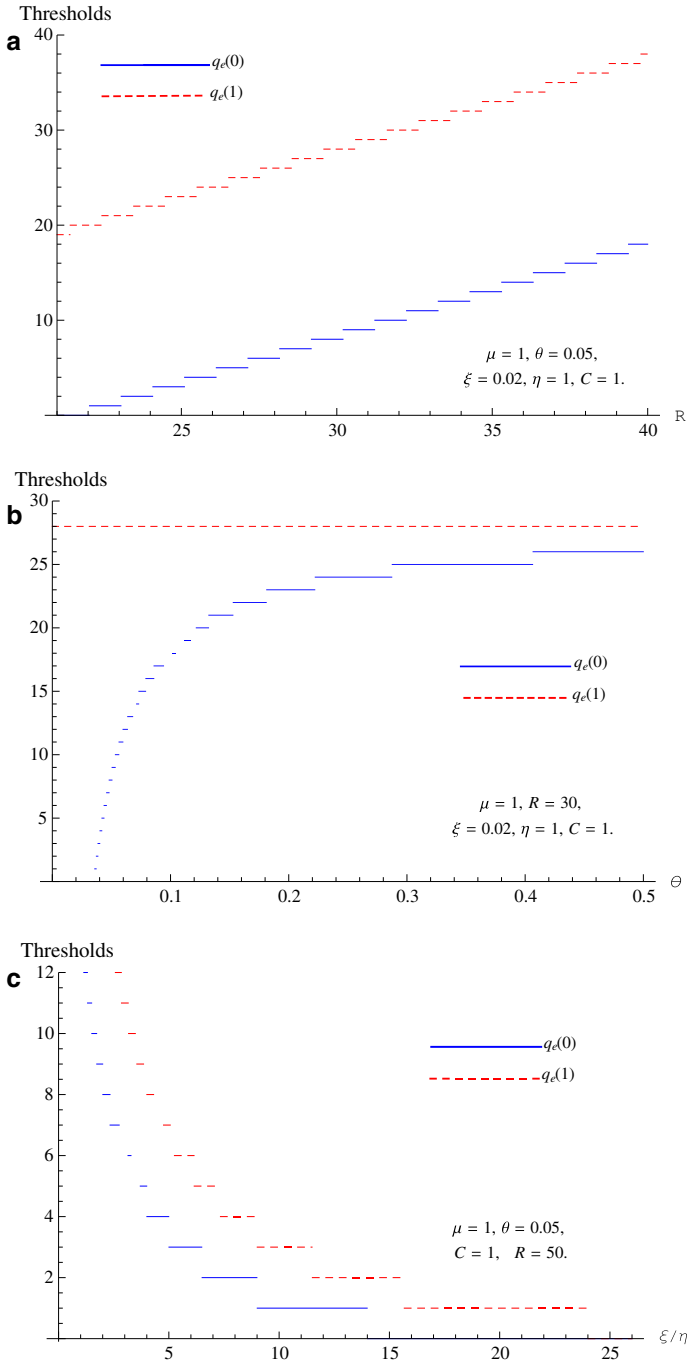


Fig. 4 Equilibrium thresholds for observable system. **a** Sensitivity with respect to R , for $\mu = 1, \theta = 0.05, \xi = 0.02, \eta = 1, C = 1$. **b** Sensitivity with respect to θ , for $\mu = 1, R = 30, \xi = 0.02, \eta = 1, C = 1$. **c** Sensitivity with respect to ξ/η , for $\mu = 1, \theta = 0.05, C = 1, R = 50$

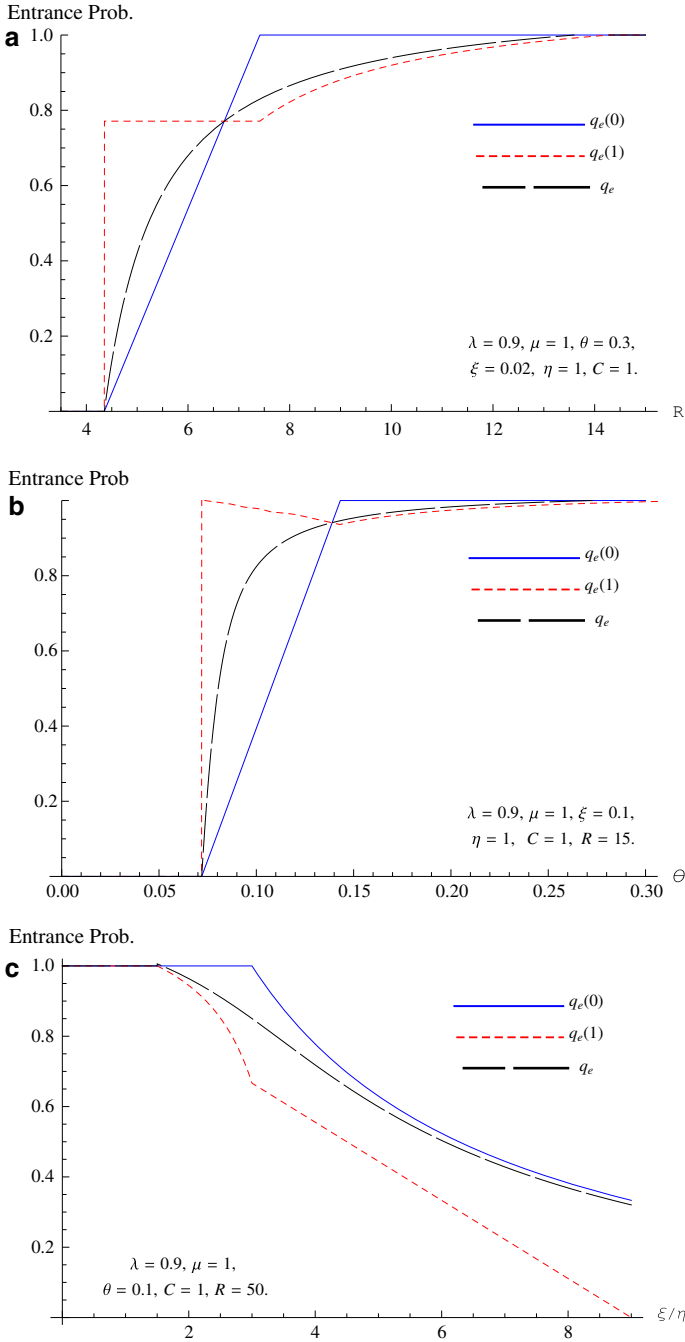


Fig. 5 Equilibrium entrance probabilities $q_e(0), q_e(1)$ and q_e , **a** sensitivity with respect to R , for $\lambda = 0.9, \mu = 1, \theta = 0.3, \xi = 0.02, \eta = 1, C = 1$. **b** Sensitivity with respect to θ , for $\lambda = 0.9, \mu = 1, \xi = 0.1, \eta = 1, C = 1, R = 15$. **c** Sensitivity with respect to ξ/η , for $\lambda = 0.9, \mu = 1, \theta = 0.1, C = 1, R = 50$

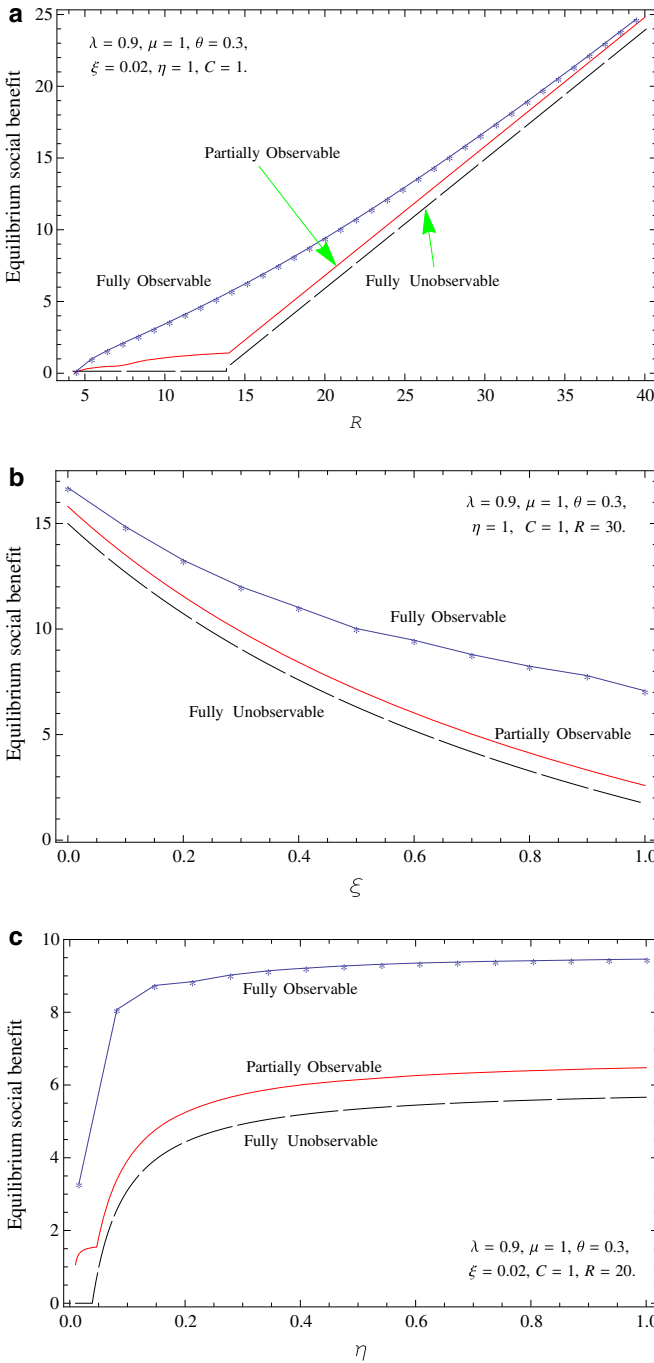


Fig. 6 Equilibrium social benefit for three different cases, **a** sensitivity with respect to R , for $\lambda = 0.9, \mu = 1, \theta = 0.3, \xi = 0.02, \eta = 1, C = 1$. **b** Sensitivity with respect to ξ , for $\lambda = 0.9, \mu = 1, \theta = 0.3, \eta = 1, C = 1, R = 30$. **c** Sensitivity with respect to η , for $\lambda = 0.9, \mu = 1, \theta = 0.3, \xi = 0.02, C = 1, R = 20$

equilibrium social benefit in the partial observable case is always inside the equilibrium social benefit formed by fully observable case and fully unobservable case.

7 Conclusions

In this paper, we have considered customer's equilibrium behavior in the $M/M/1$ queue with setup times, breakdowns and repairs. We have analyzed the equilibrium thresholds in the fully observable case and the counterpart equilibrium entrance probabilities in the partially observable case and fully unobservable case. We also discussed the sensitivity with respect to various parameters in the three different cases.

We should pay attention to the effect of the server unreliability on the system performances. It is clear that customers' equilibrium thresholds and optimal entrance probabilities decrease as ξ/η increases, which means that many customers unwillingness to join the system and the social welfare decreases at the same time. So, the analysis on the Nash equilibrium of the present model with server breakdowns is more important and practical than the model with a reliable server.

Acknowledgments The authors would like to thank the anonymous referees for their valuable comments and suggestions which are very helpful for us to improve the presentation of the paper. This research is partially supported by NSFC (71371075, 71090403, 71271089), NSFA (KJ2013B162, KJ2014A174), FSYT (2011195). It is also supported by the ISCISI of South China University of Technology and Chaoju University Scientific Research Fund.

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- Allahverdi A, Ng CT, Cheng TE, Kovalyov MY (2008) A survey of scheduling problems with setup times or costs. *Eur J Oper Res* 187(3):985–1032
- Bischof W (2001) Analysis of M/G/1 queues with setup times and vacations under six different service disciplines. *Queueing Syst* 39:265–301
- Boudali O, Economou A (2012) Optimal and equilibrium balking strategies in the single server Markovian queue with catastrophes. *Eur J Oper Res* 218:708–715
- Burnetas A, Economou A (2007) Equilibrium customer strategies in a single server Markovian queue with setup times. *Queueing Syst* 56:213–228
- Chen H, Frank M (2004) Monopoly pricing when customers queue. *IIE Trans* 36:569–581
- Choudhury G (2000) An $M^X/G/1$ queueing system with a setup period and a vacation period. *Queueing Syst* 36:23–38
- Economou A, Kanta S (2008) Equilibrium balking strategies in the observable single server queue with breakdowns and repairs. *Oper Res Lett* 36:696–699
- Economou A, Kanta S (2011) Equilibrium customer strategies and social-profit maximization in the single-server constant retrial queue. *Naval Res Logist* 58(2):107–122
- Edelson NM, Hildebrand K (1975) Congestion tolls for Poisson queueing processes. *Econometrica* 43:81–92

- Hassin R, Haviv M (2003) To queue or not to queue: equilibrium behavior in queueing systems. Kluwer Academic, Boston
- Naor P (1969) The regulation of queue size by levying tolls. *Econometrica* 37:15–24
- Reddy GV, Nadarajan R, Arumuganathan R (1998) Analysis of a bulk queue with N-policy multiple vacations and setup times. *Comput Oper Res* 25(8):957–967
- Wang J, Cao J, Li Q (2001) Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Syst* 38(2):363–380
- Zhang F, Zhu Z (2013) A discrete-time unreliable *Geo/G/1* retrial queue with balking customers, second optional service, and general retrial times. *Math Probl Eng* 2013:1–12, Art ID 832318
- Zhang Z, Wang J, Zhang F (2014) Equilibrium customer strategies in the single server constant retrial queue with breakdowns and repairs. *Math Probl Eng* 2014:1–14, Art ID 379572