Probability in the Engineering and Informational Sciences (2024), **38:**2 428–447
doi:10.1017/S0269964823000219

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

On a retrial queue with negative customers, passive breakdown, and delayed repairs

Yunna Han, Ruiling Tian, Xinyu Wu and Liuqing He

School of Science, Yanshan University, Qinhuangdao, China
Corresponding author: Ruiling Tian; Email: tianrl@ysu.edu.cn

Keywords: equilibrium strategy; negative customers; passive breakdowns; performance measures; unreliable retrial queue

Abstract

This paper studies an M/M/1 retrial queue with negative customers, passive breakdown, and delayed repairs. Assume that the breakdown behavior of the server during idle periods is different from that during busy periods. Passive breakdowns may occur when the server is idle, due to the lack of monitoring of the server during idle periods. When the passive breakdown occurs, the server does not get repaired immediately and enters a delayed repair phase. Negative customers arrive during the busy period, which will cause the server to break down and remove the serving customers. Under steady-state conditions, we obtain explicit expressions of the probability generating functions for the steady-state distribution, together with some important performance measures for the system. In addition, we present some numerical examples to illustrate the effects of some system parameters on important performance measures and the cost function. Finally, based on the reward-cost structure, we discuss Nash equilibrium and socially optimal strategy and numerically analyze the influence of system parameters on optimal strategies and optimal social benefits.

1. Introduction

In recent years, there has been an increasing interest in modeling and analyzing retrial queues. Retrial queues have been widely applied in many real-life systems, such as telephone switching systems, mobile communication networks, random access protocols in wireless networks, and call centers. Retrial queues can reflect the customer's service requirements, and arriving customers (users, calls, data, packets, et al.) who find the server unavailable can join a retrial group and request their services later. Jeongsim and Bara [9] studied various retrial queueing models and mainly focused on the analytical results of queue length distribution, waiting time distribution, and tail asymptotics of queue length and the waiting time distributions. Falin [4] analyzed a retrial queueing system with batch arrival and used a generating function approach to derive the distribution of the orbit length, which yields some important performance measures. Sherman and Kharoufeh [13] analyzed an unreliable M/M/1 retrial queue with infinite capacity orbit and a normal queue. Retrial customers do not rejoin the normal queue and try to access the server directly at random intervals independently of arrivals or other retrial customers until they find the server in operation and idle. Due to unpredictable factors in reality, such as limited server lifetime, external disturbances, startup failures, etc., the servers may break down during idle or busy periods. Kulkarni and Choi [11] introduced a retrial queue with breakdown and repairs and derived stability conditions and the steady-state distribution of the system. Since then, the retrial queues with breakdowns and repairs have been studied extensively from stability and reliability perspectives. Krishna et al. [10] introduced a Markovian retrial queue with two types of breakdowns in which the server breaks down at different Poisson rates during an idle or busy period. They derived some performance measures and analyzed the orbit characteristics. Based on this, Gao et al. [5] studied the M/G/1 retrial queues with

two types of breakdowns and delayed repairs and obtained some important performance measures and reliability measures by the supplementary variable method.

In recent decades, a new trend has emerged in the study of queueing systems from an economic perspective, where the most fundamental problem is determining individual equilibrium and socially optimal strategies. The study of customer behavior strategy can be traced back to Naor [12], who imposed a simple linear reward-cost structure of the observable M/M/1 model. This model has subsequently been extended and supplemented by many researchers. See Hassin and Haviv [8] and Hassin [7]. Zhu et al. [16] considered equilibrium joining strategy for the almost observable case of an unreliable Mn/G/1 queue, where the arrival rate depends on the number of customers in the system. Bontali and Economou [1] considered equilibrium joining strategy for batch service queueing systems in unobservable and observable cases. Gao et al. [6] dealt with an M/M/1 retrial queue with unreliable servers from an economic point of view. Do et al. [2] studied M/M/1 retrial queues with working vacation and constant retrial rates and obtained customer equilibrium and socially optimal strategies for different information levels. Economou and Kanta [3] considered equilibrium balking strategy in a single-server repairable queueing system under two observable information levels. Zhang et al. [15] discussed equilibrium strategies in repairable M/M/1 constant retrial queues. Wang et al. [14] discussed an M/M/1 constant retrial queue with balking customers and set-up times, and they studied equilibrium strategies in the almost unobservable queue.

Motivated by the aforementioned studies, we consider an M/M/1 retrial queueing model with two different breakdowns: passive breakdown with delayed repairs and active breakdown caused by negative customers. This model has many applications. For example, this retrial queue has potential applications in packet-switched networks. If a source host wishes to send the packets to a destination host, it first sends packets to the router it is connected to and then sends packets to the destination host. If that router is available, those packets are accepted and transmitted immediately. Otherwise, due to transmission control protocol/IP network path maximum transmission unit limitations, packets are blocked by the router, in which case the blocked packets are stored in the source host's buffer and retransmitted after some time. The router may suffer from viruses, while transferring data, causing those packets to be lost and unable to continue transmission. If it suffers from the virus breakdown, the repair takes immediately. The system breakdown may occur, while the router is idle, then the repair will delay until the next incoming packet arrives.

As another application of the retrial model, we take a telephone consultation system as an example. In such an advice scenario, the telephone operator is responsible for establishing communication between the server and the customers; the operator needs to record the call information in a registration form (corresponding to an "orbit"). When a customer calls, if the server is idle, the operator takes down the information and the advisor serves the customer immediately. On the other hand, if the server is busy, the operator tells the customer to call again after a certain period of time (called a retrial) and the customer can choose to leave a message waiting for service or leave the system. The telephone consultation system may break down. If a breakdown occurs, while a customer is being received, it will cause the customer being consulted to end the call. If the breakdown occurs, while no one is receiving a consultation, the operator cannot immediately detect the fault and leaves it in a delayed state of repair. The server continues to serve the customer after the repair is completed.

The proposed paper aims to study the stationary performance analysis and customers' strategic behavior in the repairable retrial queues. Our contributions are as follows: under the stability condition, we construct the balance equations to obtain the steady-state probabilities of the server in different states and derive the system performance measures. Numerical examples to analyze the effects of parameters on the performance measures. We propose a cost function and find the minimum operating cost with the numerical example. Finally, we extensively analyze the customers' equilibrium joining behavior and socially optimal strategies.

The rest of the paper is organized as follows. Section 2 describes the model in detail. Section 3 obtains the steady-state probability distribution by using the probability generating function method for

steady-state analysis. Section 4 shows some important performance measures of the system. Section 5 shows the impact of system parameters on performance measures through numerical examples. And, a cost function is proposed and optimized. Section 6 analyzes the optimal strategies of the customers through the individual utility function and social benefit function. The conclusion is given in Section 7.

2. Model description

We consider an unreliable retrial queue with two types of breakdowns: passive breakdown with delayed repairs and active breakdown caused by negative customers. Customers arrive at the system according to a Poisson process with rate λ. If a customer arrives and finds the server idle, he will immediately receive the service. We assume that the customer's service time follows an exponential distribution with parameter μ. Otherwise, arriving customers who find the server unavailable join the orbit with probability q or balk with complementary probability $1 - q$. The retrial time obeys an exponential distribution with parameter ν. The server may encounter two types of breakdowns, namely passive breakdown and active breakdown. Passive breakdown means that when the server is idle, the server breaks down according to a Poisson process with rate η. However, due to the lack of monitoring of the server during idle periods, when a passive breakdown occurs, the server is not repaired immediately and stays there until the customer arrives at the system from outside or in orbit (if available). The repair time of passive breakdown is exponentially distributed with the parameter θ, and delayed time obeys an exponential distribution with the parameter δ. After the passive breakdown occurs, the server starts to be repaired. During this repair time, the customers do not leave the system and will receive their service immediately after completing the repair. Active breakdown means that when the server is busy, the arrival of negative customers affects the system, causing the server to break down and simultaneously forcing the customer being served to leave the system. The server is repaired immediately. The negative customers arrive according to a Poisson process with rate φ. The repair time of the active breakdown obeys an exponential distribution with the parameter β.

Finally, we assume that the inter-arrival time, the service time, the breakdown time, the repair time, the delayed time, and the retrial time are independent of each other.

Let $N(t)$ denote the number of customers in orbit at time t and $I(t)$ denote the state of the server at time t as defined below:

$$I(t) = \begin{cases} 0, & \text{The server is idle,} \\ 1, & \text{The server is busy,} \\ 2, & \text{The server is being repaired due to passive breakdown,} \\ 3, & \text{The server is being repaired due to negative customers,} \\ 4, & \text{The server is in delayed repair status.} \end{cases}$$

Thus, the state of the system at time t can be described by the pair $(N(t), I(t))$. From Markov process theory, we know that $\{(N(t), I(t)), t > 0\}$ constitutes a two-dimensional Markov chain with the state space $\Omega = \{(j, i), j \geq 0, i = 0, 1, 2, 3, 4\}$. The transition rate diagram of the Markov chain is shown in Figure 1.

3. Steady-state analysis

In this section, we focus on the steady-state analysis of the system.

Let $\pi(j, i)$ denote the stationary joint probability at state (j, i).

$$\pi(j, i) = \lim_{t \to \infty} P\{N(t) = j, I(t) = i\}, (j, i) \in \Omega.$$

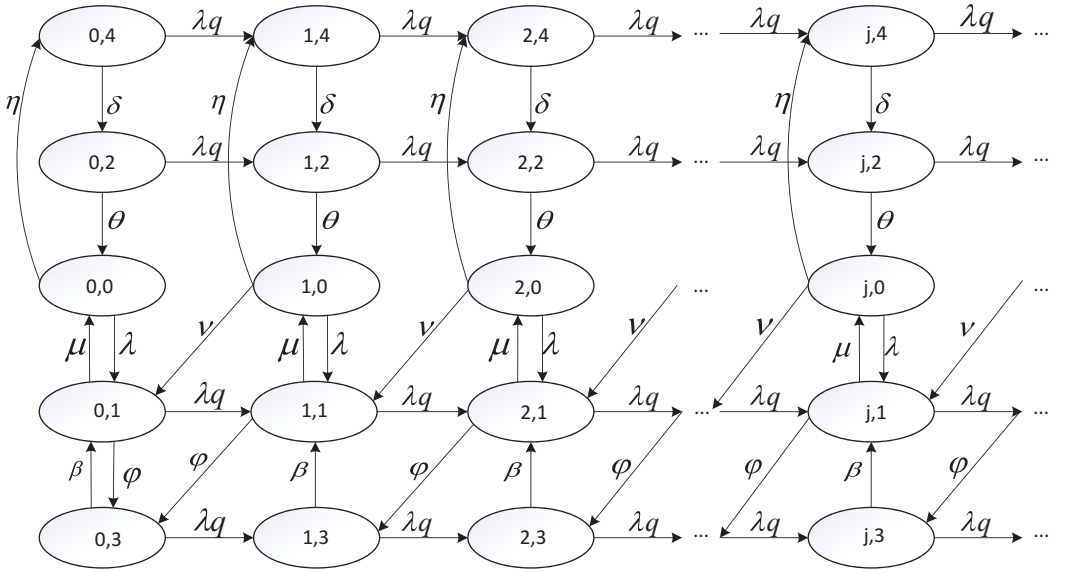

Figure 1. *Transition rate diagram of the Markov chain.*

The probability $\pi(j,i)$ satisfies the following balance equations:

$$(\lambda+\eta)\pi(0,0)=\mu\pi(0,1)+\theta\pi(0,2), \tag{1}$$

$$(\lambda+\eta+\nu)\pi(j,0)=\mu\pi(j,1)+\theta\pi(j,2),\ \ j\geq 1, \tag{2}$$

$$(\lambda q+\mu+\varphi)\pi(0,1)=\lambda\pi(0,0)+\beta\pi(0,3)+\nu\pi(1,0), \tag{3}$$

$$(\lambda q+\mu+\varphi)\pi(j,1)=\lambda\pi(j,0)+\beta\pi(j,3)+\nu\pi(j+1,0)+\lambda q\pi(j-1,1),\ \ j\geq 1, \tag{4}$$

$$(\lambda q+\theta)\pi(0,2)=\delta\pi(0,4), \tag{5}$$

$$(\lambda q+\theta)\pi(j,2)=\delta\pi(j,4)+\lambda q\pi(j-1,2),\ \ j\geq 1, \tag{6}$$

$$(\lambda q+\beta)\pi(0,3)=\varphi\pi(0,1)+\varphi\pi(1,1), \tag{7}$$

$$(\lambda q+\beta)\pi(j,3)=\varphi\pi(j+1,1)+\lambda q\pi(j-1,3),\ \ j\geq 1, \tag{8}$$

$$(\lambda q+\delta)\pi(0,4)=\eta\pi(0,0), \tag{9}$$

$$(\lambda q+\delta)\pi(j,4)=\eta\pi(j,0)+\lambda q\pi(j-1,4),\ \ j\geq 1. \tag{10}$$

In order to solve the above equations for obtaining the system steady-state distribution, we define the partial generating functions:

$$G_i(z)=\sum_{j=0}^{\infty}\pi(j,i)z^j,\quad |z|<1,\quad i\in\{0,1,2,3,4\}.$$

Multiplying (1)–(10) by z^j and summing over j ($j \geq 0$), we obtain:

$$(\lambda + \eta + \nu)G_0(z) = \mu G_1(z) + \theta G_2(z) + \nu\pi(0,0), \tag{11}$$

$$(\lambda q + \mu + \varphi)G_1(z) = \lambda G_0(z) + \beta G_3(z) + \lambda qz G_1(z) + \frac{\nu}{z}G_0(z) - \frac{\nu}{z}\pi(0,0), \tag{12}$$

$$(\lambda q + \theta)G_2(z) = \delta G_4(z) + \lambda qz G_2(z), \tag{13}$$

$$(\lambda q + \beta)G_3(z) = \frac{\varphi}{z}G_1(z) - \frac{\varphi(1-z)}{z}\pi(0,1) + \lambda qz G_3(z), \tag{14}$$

$$(\lambda q + \delta)G_4(z) = \eta G_0(z) + \lambda qz G_4(z). \tag{15}$$

Theorem 3.1. *The probabilities of the server being in each state are as follows:*
(1) The probability that the server is in a normal idle state:

$$P_0 = G_0(1) = \frac{\mu\nu\theta\delta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\delta\mu\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}\pi(0,1). \tag{16}$$

(2) The probability that the server is busy:

$$P_1 = G_1(1) = \frac{\beta\mu\lambda\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\delta\mu\kappa_2(\lambda + \nu)}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}\pi(0,1). \tag{17}$$

(3) The probability that the server is in a repaired state due to a passive breakdown:

$$P_2 = G_2(1) = \frac{\mu\nu\delta\eta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\delta\mu\eta\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}\pi(0,1). \tag{18}$$

(4) The probability that the server is in a repaired state due to an active breakdown:

$$P_3 = G_3(1) = \frac{\varphi\mu\lambda\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \varphi^2\theta\delta\kappa_2(\lambda + \nu)}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}\pi(0,1). \tag{19}$$

(5) The probability that the server is in a delayed repair state due to delayed repair:

$$P_4 = G_4(1) = \frac{\mu\nu\theta\eta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\eta\mu\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}\pi(0,1), \tag{20}$$

where

$$\kappa_1 = \beta(\varphi + \mu) - \lambda q(\beta + \varphi),$$

$$\kappa_2 = \lambda\left[\lambda q(\lambda q + \eta q + \theta + \delta) + \eta q(\theta + \delta) + \theta\delta\right],$$

$$\kappa_3 = \eta q(\theta + \delta) + \theta\delta,$$

$$A_1 = \nu\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\kappa_2,$$

$$A_2 = \lambda\mu\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \varphi\theta\delta\kappa_2(\lambda + \nu),$$

$$\pi(0,1) = \frac{\kappa_2\left[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3\right]}{\mu\left[\theta\delta + \eta(\theta + \delta)\right]A_1 + (\beta + \varphi)A_2}. \tag{21}$$

Proof. Combining from (11) and (12), we obtain

$$[\lambda(1-z)+\eta]\,G_0(z) = [(1-z)(\mu-\lambda qz)-\varphi z]\,G_1(z) + \theta G_0(z) + \beta z G_3(z). \tag{22}$$

Combining (13)–(15) and (22) after some algebraic operations, we have

$$G_1(z) = \Theta_1 \times \frac{\lambda q(1-z)+\beta}{\beta\varphi}\Theta_2 G_0(z) + \Theta_2\pi(0,1), \tag{23}$$

where

$$\Theta_1 = \frac{\lambda\,[\lambda q(1-z)\,[\lambda q(1-z)+\eta q+\theta+\delta]+\kappa_3]}{[\lambda q(1-z)+\theta]\,[\lambda q(1-z)+\delta]}, \tag{24}$$

$$\Theta_2 = \frac{\beta\varphi}{\lambda q\,[(1-z)(\mu-\lambda qz)-z(\beta+\varphi)]+\beta(\varphi+\mu)}. \tag{25}$$

Using the above method in (15), we have

$$G_4(z) = \frac{\eta}{\lambda q(1-z)+\delta} G_0(z). \tag{26}$$

Substituting (26) into (13), we obtain

$$G_2(z) = \frac{\delta\eta}{[\lambda q(1-z)+\delta]\,[\lambda q(1-z)+\theta]} G_0(z). \tag{27}$$

Organizing (14), we gain

$$G_3(z) = \frac{\varphi}{z\,[\lambda q(1-z)+\beta]} G_1(z) - \frac{\varphi(1-z)}{z\,[\lambda q(1-z)+\beta]}\pi(0,1). \tag{28}$$

After substituting the above equations into (11) with some algebraic operations, we have

$$\begin{aligned} G_0(z) = {} & \frac{\nu}{\lambda+\eta+\nu-\mu\Theta_1\times\frac{\lambda q(1-z)+\beta}{\beta\varphi}\Theta_2-\frac{\theta\delta\eta}{[\lambda q(1-z)+\theta][\lambda q(1-z)+\delta]}}\pi(0,0) \\ & + \frac{\mu\Theta_2}{\lambda+\eta+\nu-\mu\Theta_1\times\frac{\lambda q(1-z)+\beta}{\beta\varphi}\Theta_2-\frac{\theta\delta\eta}{[\lambda q(1-z)+\theta][\lambda q(1-z)+\delta]}}\pi(0,1). \end{aligned} \tag{29}$$

□

So far, we have obtained the requested results (23)–(29), and only $\pi(0,1)$ and $\pi(0,0)$ remains to be determined, from (1), (5), and (9), we have

$$\left[\lambda+\eta-\frac{\theta\delta\eta}{(\lambda q+\theta)(\lambda q+\delta)}\right]\pi(0,0) = \mu\pi(0,1). \tag{30}$$

Substituting $z=1$ into (24)–(28), and using the normalization condition:

$$G_0(1)+G_1(1)+G_2(1)+G_3(1)+G_4(1) = 1, \tag{31}$$

we can get $\pi(0,1)$. By calculations, the above theorem can be obtained.

From (16)–(20), the system steady-state condition is

$$\theta\delta(\lambda+\nu)\kappa_1 > \beta\mu\lambda\kappa_3. \tag{32}$$

4. Performance measures

In this section, we study some important performance measures of the retrial queueing system under the stability condition (32).

(1) The average orbit sizes N_i $(i = 0, 1, 2, 3, 4)$ when the server is idle, busy, passive breakdown, active breakdown, and delayed time, respectively, are

$$N_0 = G_0'(1) = \frac{\omega_1}{\theta\delta\kappa_1\left[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3\right]} G_0(1) + \frac{\lambda q\beta\varphi\mu\theta\delta\alpha_1}{\kappa_1\left[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3\right]}\pi(0,1), \tag{33}$$

$$\begin{aligned} N_1 = G'_1(1) = &\left(\frac{\omega_1(\lambda+\nu)}{\theta\delta\mu\kappa_1[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3]} + \frac{\beta\mu\lambda - \kappa_1(\lambda+\nu)}{\beta\mu^2}\right) G_0(1) \\ &+ \left(\frac{\lambda q\beta\varphi\theta\delta\alpha_1(\lambda+\nu)}{\mu\kappa_1\kappa_2[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3]} + \frac{\beta\varphi\kappa_2 - \nu\kappa_1(\lambda q+\theta)(\lambda q+\delta)}{\beta\mu\kappa_2}\right)\pi(0,1), \end{aligned} \tag{34}$$

$$\begin{aligned} N_2 = G'_2(1) = &\left(\frac{\eta\omega_1}{\theta^2\delta\kappa_1[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3]} + \frac{\lambda q\eta(\theta+\delta)}{\theta^2\delta}\right) G_0(1) \\ &+ \frac{\lambda q\beta\varphi\mu\theta\delta\alpha_1}{\kappa_1[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3]}\pi(0,1), \end{aligned} \tag{35}$$

$$\begin{aligned} N_3 = G'_3(1) = &\left(\frac{\varphi\omega_1(\lambda+\nu)}{\beta\theta\delta\mu\kappa_1\left[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3\right]} + \frac{\beta\varphi\lambda\mu - \varphi\kappa_1(\lambda+\nu)}{\beta^2\mu^2}\right) G_0(1) \\ &+ \left(\frac{\mu\omega_2 + \lambda q\beta\varphi^2\theta\delta\alpha_1(\lambda+\nu)}{\mu\beta\kappa_1\kappa_2\left[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3\right]} + \frac{\beta\mu\varphi\kappa_2(\varphi+\mu) - \varphi\mu\nu\kappa_1(\lambda q+\theta)(\lambda q+\delta)}{\beta^2\mu^2\kappa_2}\right)\pi(0,1), \end{aligned} \tag{36}$$

$$N_4 = G_4'(1) = \left(\frac{\eta\omega_1}{\delta^2\theta\kappa_1[\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3]} + \frac{\lambda q\eta}{\delta^2}\right) G_0(1) + \frac{\lambda q\beta\varphi\mu\theta\delta\alpha_1}{\kappa_1\left(\theta\delta\kappa_1(\lambda+\nu) - \beta\mu\lambda\kappa_3\right)}\pi(0,1), \tag{37}$$

where $G_0(1)$ and $\pi(0,1)$ are given in (16) and (21), and

$$\alpha_1 = \varphi + \mu + \beta - \lambda q,$$

$$\omega_1 = \lambda q\eta\theta\delta\kappa_1^2(\theta+\delta) - \lambda^2 q\theta\delta\mu\kappa_1\left[\kappa_3 + \beta(\theta+\delta+q\eta)\right] + \lambda^2 q\beta\mu\kappa_3\left[\kappa_1(\theta+\delta) + \theta\delta\alpha_1\right],$$

$$\omega_2 = \lambda\mu\nu\varphi\kappa_3(\lambda q+\theta)(\lambda q+\delta)(\lambda q-\beta) + \theta\delta\varphi^2\kappa_2(\lambda+\nu)(\lambda q-\beta).$$

(2) Let $E(N)$ and $E(L)$ denote the average number of customers in the orbit and in the system, respectively. Thus,

$$E(N) = N_0 + N_1 + N_2 + N_3 + N_4. \tag{38}$$

The average number of customers in the system $E(L)$, is the average number of customers in orbit plus the probability that a customer is being for service. So

$$E(L) = E(N) + G_1(1) + G_2(1) + G_4(1). \tag{39}$$

(3) Assuming that a tagged customer finds that the server is unavailable on arrival and decides to join the retrial orbit, his expected (conditional) waiting time in the orbit is given by

$$\begin{aligned} T(q) &= \frac{E(N)}{\lambda_{ret}} = \frac{E(N)}{\lambda q\left(G_1(1)+G_2(1)+G_3(1)+G_4(1)\right)} \\ &= \frac{\tau_a}{\tau_b}\left(\frac{\eta(\theta^2+\delta^2+\theta\delta)}{\theta^2\delta^2}+\frac{(\beta+\varphi)\left(\beta\mu\lambda-\kappa_1(\lambda+\nu)\right)}{\lambda q\beta^2\mu^2}\right) \\ &+\left(\frac{\eta(\theta+\delta)+\theta\delta}{\theta\delta}+\frac{(\beta+\varphi)(\lambda+\nu)}{\beta\mu}\right)\left(\frac{\beta\mu\varphi\theta\delta\alpha_1\kappa_2}{\kappa_1\tau_a}+\frac{\omega_1}{\lambda q\theta\delta\kappa_1\left(\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right)}\right) \\ &+\frac{1}{\lambda q\beta\tau_a}\left(\omega_2+\frac{\tau_b(\beta+\varphi)\left(\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right)}{\beta\theta\delta\mu^2}\right), \end{aligned} \tag{40}$$

where

$$\tau_a = \mu\nu(\lambda q+\theta)(\lambda q+\delta)\left[\lambda\kappa_3(\beta+\varphi)+\eta\kappa_1(\theta+\delta)\right]+\varphi\kappa_2\left[\theta\delta(\beta+\varphi)(\lambda+\nu)+\beta\mu\eta(\theta+\delta)\right],$$

$$\tau_b = \mu\nu\theta\delta\kappa_1(\lambda q+\theta)(\lambda q+\delta)+\beta\varphi\theta\delta\mu\kappa_2.$$

(4) The busy cycle T is defined as the length of time starting when the server completes a service and the orbit is empty and ending when the server becomes idle and the orbit is empty again. Therefore, we have

$$E(T) = \frac{\frac{1}{\lambda}}{\pi(0,0)} = \frac{\mu\left[\theta\delta+\eta(\theta+\delta)\right]\mathrm{A}_1+(\beta+\varphi)\mathrm{A}_2}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}, \tag{41}$$

where $\mathrm{A}_1, \mathrm{A}_2$ are given in (21).

(5) In the busy cycle, the expected length of idle period, $E(T_0)$, is computed as:

$$E(T_0) = E(T)G_0(1) = \frac{\mu\nu\theta\delta\kappa_1(\lambda q+\theta)(\lambda q+\delta)+\beta\varphi\theta\delta\mu\kappa_2}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}. \tag{42}$$

(6) In the busy cycle, the expected length of busy period, $E(T_1)$, is determined as:

$$E(T_1) = E(T)G_1(1) = \frac{\beta\mu\lambda\nu\kappa_3(\lambda q+\theta)(\lambda q+\delta)+\beta\varphi\theta\delta\kappa_2(\lambda+\nu)}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}. \tag{43}$$

(7) In the busy cycle, the expected length of the repair period due to passive breakdown, $E(T_2)$, is calculated as:

$$E(T_2) = E(T)G_2(1) = \frac{\mu\nu\delta\eta\kappa_1(\lambda q+\theta)(\lambda q+\delta)+\beta\varphi\delta\mu\eta\kappa_2}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}. \tag{44}$$

(8) In the busy cycle, the expected length of the repair period due to active breakdown, $E(T_3)$, is derived as:

$$E(T_3) = E(T)G_3(1) = \frac{\varphi\mu\lambda\nu\kappa_3(\lambda q+\theta)(\lambda q+\delta)+\varphi^2\theta\delta\kappa_2(\lambda+\nu)}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}. \tag{45}$$

(9) In the busy cycle, the expected length of delayed period, $E(T_4)$, is obtained as:

$$E(T_4) = E(T)G_4(1) = \frac{\mu\nu\theta\eta\kappa_1(\lambda q+\theta)(\lambda q+\delta)+\beta\varphi\theta\mu\eta\kappa_2}{\lambda\mu(\lambda q+\theta)(\lambda q+\delta)\left[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3\right]}. \tag{46}$$

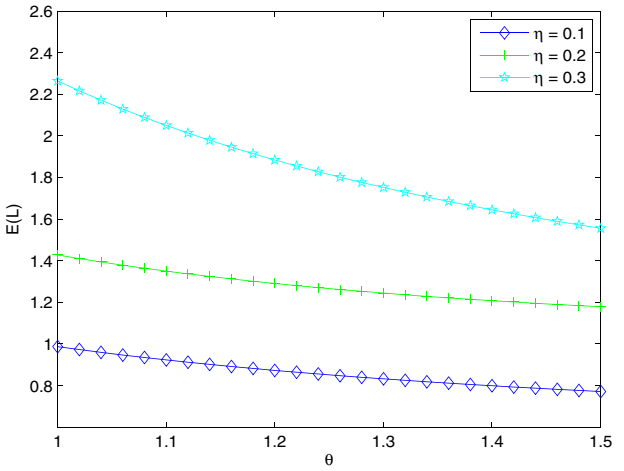

Figure 2. *Average system length E(L) versus θ for different values of η* ($q = 0.65, \lambda = 2.2, \mu = 8, \beta = 1.4, \varphi = 0.5, \delta = 3, \nu = 3$).

(10) The probability that the server is under breakdown/repaired is given by:

$$P_R = G_2(1) + G_3(1) + G_4(1) = \frac{\mu\nu(\lambda q+\theta)(\lambda q+\delta)[\eta\kappa_1(\theta+\delta)+\lambda\varphi\kappa_3]+\varphi\kappa_2[\beta\mu\eta(\theta+\delta)+\varphi\theta\delta(\lambda+\nu)]}{\kappa_2[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3]}\pi(0,1). \tag{47}$$

(11) The probability of the server under the operative/working state is:

$$P_W = G_0(1) + G_1(1) = \frac{\mu\nu(\lambda q+\theta)(\lambda q+\delta)(\theta\delta\kappa_1+\beta\lambda\kappa_3)+\beta\varphi\theta\delta\kappa_2(\lambda+\mu+\nu)}{\kappa_2[\theta\delta\kappa_1(\lambda+\nu)-\beta\mu\lambda\kappa_3]}\pi(0,1). \tag{48}$$

5. Numerical illustrations

5.1. Impact of parameters on performance measures

In this subsection, under the system stability condition (32), we show numerical examples to explain the effects of some parameters on the performance measures.

From Figure 2, we can observe that the number of customers in the system decreases with θ and increases with η. The repair time becomes short, resulting in the server getting repaired faster and can serve more customers. As the breakdown interval becomes short, the server is more likely to cause congestion in the system queue, and the average length increases.

Figure 3 shows that, as intuitively expected, the expected waiting time $T(q)$ decreases as ν increases. Furthermore, for a fixed ν, $T(q)$ decreases with δ. This is because as the waiting time for repair becomes short, the expected waiting time for customers in the system decreases.

In Figure 4, the average orbit length $E(N)$ decreases with μ and ν. The service time and retrial time become fast, which leads to a short queue length of the orbit. Figure 5 shows the busy cycle $E(T)$ increases with λ and decreases with μ. Because the increase in arrival rates will lead to more customers joining the system, thus increasing the busy cycle. With the increase in service rate, the server serves the customers faster, leading to a decrease in the busy cycle.

Figure 6 depicts how the server breakdown/repair probability P_R increases with φ and decreases with β. Because the breakdown rate increases, the server tends to break down more frequently, resulting in an increased probability of the server being in a breakdown/repair state. Conversely, as the repair rate

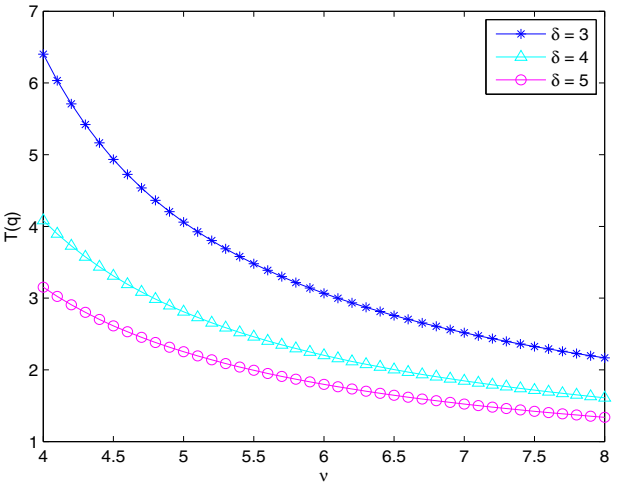

Figure 3. *Expected waiting time T(q) versus ν for different values of δ* ($q = 0.6, \lambda = 3, \mu = 8, \beta = 3.5, \varphi = 1.5, \theta = 6, \eta = 2$).

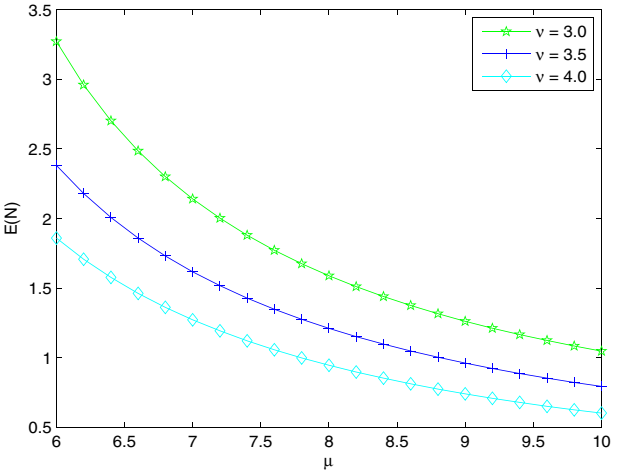

Figure 4. *Average orbit length E(N) versus μ for different values of ν* ($q = 0.4, \lambda = 4, \beta = 1.4, \varphi = 0.5, \delta = 3, \theta = 1.5, \eta = 0.1$).

increases, the server is repaired more quickly, resulting in a lower probability of the server being in a breakdown/repair state decrease.

Figure 7 describes the relationship between the time occupation rates P_i with the arrival rate λ. As the arrival rate increases, the system becomes busy. Therefore, the probability of the system is idle, passive breakdown, and delayed repair decreases, while the probability of being busy and active breakdown increases.

5.2. Optimization

In this section, we present an optimization analysis of the operating cost through the constructed cost function.

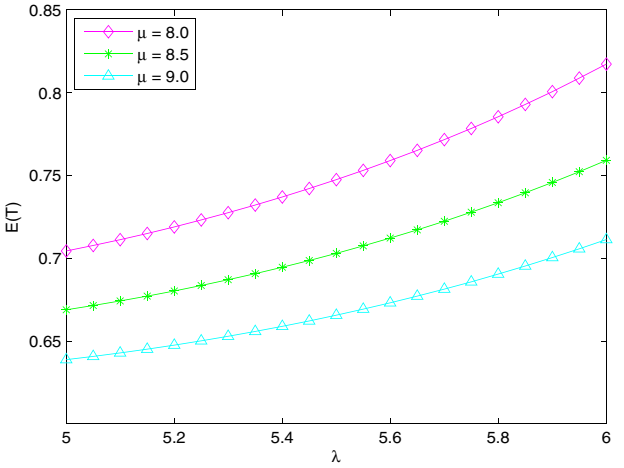

Figure 5. *The busy cycle E(T) versus λ for different values of μ* ($q = 0.2, \beta = 2, \varphi = 0.7, \theta = 1.8, \delta = 3, \eta = 0.3, \nu = 3$).

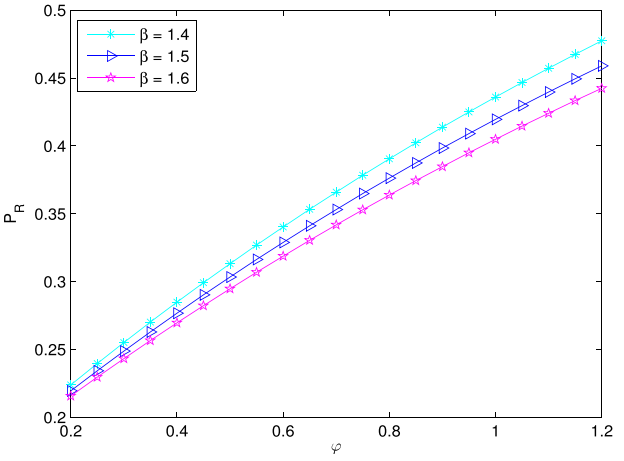

Figure 6. *Server breakdown/repair probability P_R versus φ for different values of β* ($q = 0.6, \lambda = 4.3, \mu = 8, \theta = 1.8, \delta = 4, \eta = 0.1, \nu = 5.5$).

Define

C_h = Waiting cost per unit time of the customer in the system,
C_s = Cost per unit time of providing services,
C_e = Cost per unit time of delayed repair time,
C_o = Cost per unit time of keeping the system running,
C_f = Cost per unit time of server in a passive breakdown state,
C_a = Cost per unit time of server being down due to negative customers,
C_i = Cost per unit time of server in a delayed repair state.

Thus, the cost function per unit time is

$$F(\delta, \mu) = C_h E(N) + C_s \mu + C_e \delta + C_o P_1 + C_f P_2 + C_a P_3 + C_i P_4. \tag{49}$$

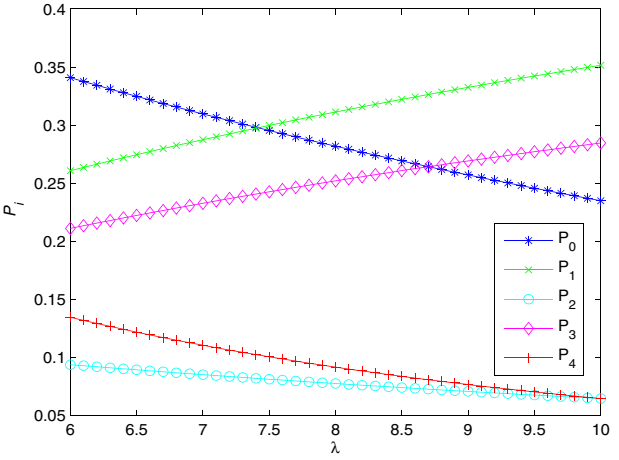

Figure 7. *The probabilities of each state P_i versus λ* ($q = 0.4, \beta = 4.2, \varphi = 3.4, \theta = 4, \eta = 1.1, \mu = 12, \delta = 4, \nu = 8$).

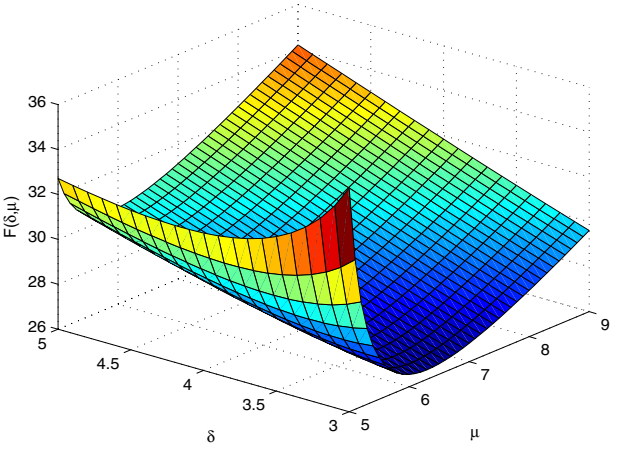

Figure 8. *The cost variation with δ and μ* ($q = 0.42, \varphi = 0.5, \eta = 0.4, \lambda = 3, \beta = 2, \theta = 1.7, \nu = 1.5$).

5.2.1. Single-objective optimization

In the following subsection, we discuss the single-objective optimization problem, which is the expected cost function per unit time illustrated in Equation (49). However, it is difficult to obtain the optimal value because the cost function is nonlinear and very complex. Therefore, we propose some numerical calculations to optimize the cost. In the following numerical experiments, we set the parameters for

$$C_h = 0.5, C_s = 2, C_e = 2.5, C_o = 3, C_f = 4, C_a = 4, C_e = 4.5.$$

In Figure 8, we find that $F(\delta, \mu)$ first decreases and then increases as δ and μ increase, so we can get the minimum value. As the service rate μ increases, customers' waiting time in the system is reduced, thus reducing the cost. The constant improvement of the service rate leads to more customers joining the system, so the cost increases. The reduction in delayed repair time reduces the time it takes for the server to enter a working state, which leads to a decrease in the cost. However, as the delayed repair rate continues to increase, more customers join the system, leading to an increase in the cost.

Table 1. *Optimal solutions* (δ^*, μ^*) *and the corresponding costs* $(q = 0.42, \eta = 0.4, \nu = 1.5, \theta = 1.7)$.

$(\lambda, \beta, \varphi)$	δ^*	μ^*	$F(\delta^*, \mu^*)$
(2.5,2.5,1.0)	1.2	5.3	18.5262
(2.5,3.0,1.0)	1.1	4.9	18.2471
(2.5,2.5,1.5)	1.0	7.0	21.6897
(2.5,3.0,1.5)	0.9	6.6	20.7809
(3.0,2.5,1.0)	1.9	5.4	23.0954
(3.0,3.0,1.0)	1.7	5.0	22.0600
(3.0,2.5,1.5)	1.4	7.2	23.8755
(3.0,3.0,1.5)	1.3	7.0	23.4105

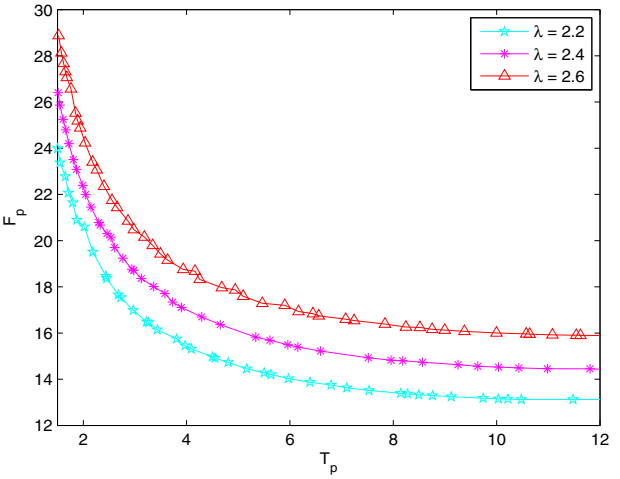

Figure 9. *Pareto-front solution found by genetic algorithm.*

From Table 1, we can see that δ^* and μ^* increase as the arrival rate λ increases, and decrease as the active repair rate β increases. Moreover, δ^* decreases as the active breakdown rate φ increases, while μ^* increases with respect to φ. According to the optimal cost, $F(\delta^*, \mu^*)$ increases with λ and φ. The reason is that as λ increases, more customers join the system and thus the cost increases. And with the increase of φ, the frequency of breakdowns increases, leading to an increase in customers' waiting time in the system and an increase in the cost. In addition, It can be seen that $F(\delta^*, \mu^*)$ decreases as β increases. This is because as the repair time decreases, the customers can get service faster, which leads to the decrease of the cost.

5.2.2. Bi-objective optimization

Most research on the optimal design of queueing models has focused on single-objective problems where a cost or profit function is the optimization objective. However, in the real world, there are many optimization problems where multiple objective functions need to be simultaneously optimized and where multiple considerations are necessary for decision-making. Expected waiting time is the most important factor for determining customer satisfaction with service quality. In queueing systems, cost is often in conflict with service quality. In this subsection, we use the non-dominated sorting genetic algorithm (NSGA-II) to find the Pareto optimal solution set satisfying both the minimum cost and the minimum waiting time. We build a bi-objective optimization model to minimize both the expected cost $F(\delta, \mu)$ and the expected waiting time $T(\delta, \mu)$, and consider the regression relationships that exist between them.

Table 2. *The Pareto optimal solutions for various values of* λ.

λ	δ^*	μ^*	T_p	F_p
2.2	1.013685239	3.047898338	10.47939618	13.1248014
	1.013685345	3.472140762	7.531527811	13.51567069
	1.265884653	4.156402737	4.548649438	14.91442655
	1.563049853	5.43597263	2.68099876	17.65367789
	2.738025415	7.427174976	1.50336452	23.98950303
	...	...	...	...
2.4	1.013685239	3.376344086	12.77151603	14.41235041
	1.095796676	3.786901271	8.566595525	14.73423676
	1.430107527	4.813294233	4.291037585	16.69683244
	1.449657869	6.831867058	2.535452456	20.15018117
	2.174975562	8.474095797	1.661489684	24.79362586
	...	...	...	...
2.6	1.001955034	3.978494624	12.65447072	15.88314227
	1.078201369	4.395894428	8.751180879	16.16272361
	1.258064516	5.107526882	5.465581314	17.29136247
	1.926686217	6.044965787	3.183612616	20.15862665
	2.251221896	8.467253177	1.890007053	25.18239912
	...	...	...	...

The bi-objective optimization problem is formulated as:

$$\min \ [F(\delta, \mu)\, T(\delta, \mu)]. \tag{50}$$

Multi-objective genetic algorithm is an evolutionary algorithm used to analyze and solve multi-objective optimization problems that is based on the genetic algorithm and Pareto optimal concept. Its core is to coordinate the relationship between each objective function and find the optimal solution set that makes each objective function as small as possible (or relatively large). We select the following system parameters: $q = 0.4, \varphi = 0.5, \beta = 1.4, \eta = 0.2, \theta = 1.5, \nu = 1.8$. The non-dominated solutions are obtained when using the multi-objective genetic algorithm are given in Figure 9, and the Pareto optimal solutions for various values are summarized in Table 2.

From Figure 9 and Table 1, the three curves show that the increase in cost leads to a decrease in the waiting time of customers. F_p increases as the average arrival rate increases. When T_p approaches infinity, the limit value of the minimum expected cost F_p can be regarded as the minimum cost.

Moreover, Figure 9 shows that the relationship between F_p and T_p is inversely related, approximating an exponential function with a negative exponent, and F_p is positively related to λ. The original and adjusted coefficients of determination are $R^2 = 99.6\%$ and $R^2(adj) = 99.6\%$, respectively, and the F-test of the regression equation has a p-value less than 0.01. Therefore, the regression model is appropriate. Figure 10 shows that the histogram of the regression residuals here is symmetric between the left and right sides, and the K-S normality test plot shows the scatter is essentially close to a straight line. The expected cumulative probabilities match well with the measured cumulative probabilities, indicating that the normal distribution is obeyed, the assumption of the regression is satisfied, and the regression relationship can be established. This regression model will help managers determine the minimum cost required to achieve a satisfactory level of service for a given estimate.

The least-square regression equation is established as follows:

$$F_P = -8.039 + 8.846\lambda + 13.694T_P^{-1} + 18.712e^{-T_p}. \tag{51}$$

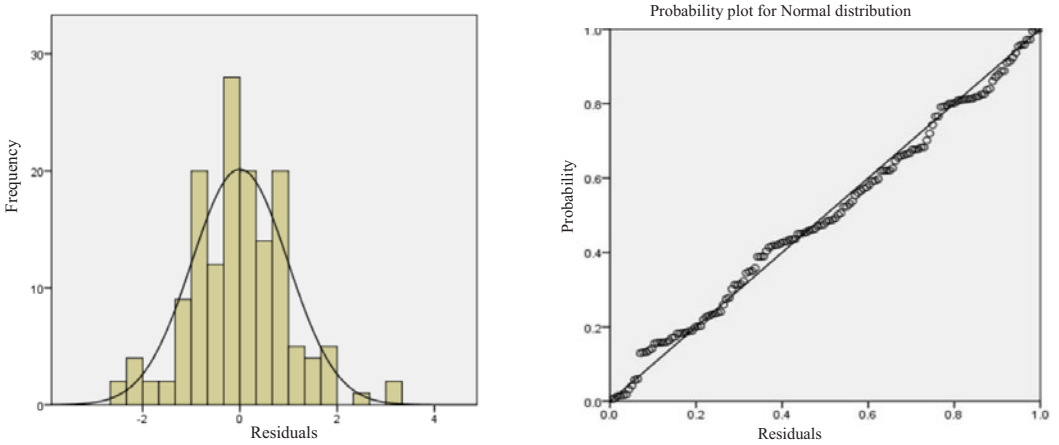

Figure 10. *Histogram and K-S normality test for the regression residuals.*

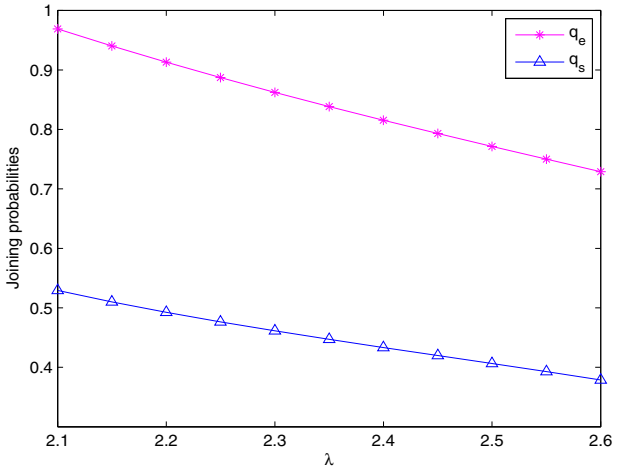

Figure 11. q_e *and* q_s *versus* λ ($\mu = 4.5, \beta = 0.3, \varphi = 0.1, \theta = 0.8, \eta = 0.1, \delta = 5, \nu = 10, R = 9, C = 1$).

6. Optimal strategy analysis

In this section, we give the reward-cost structure and focus on customers' equilibrium joining strategies and the social benefit maximization problem. After service completion, each customer receives a reward R, and every customer pays the cost of remaining in the system, where the cost per unit of waiting in the system is C. All customers are risk-neutral and behave rationally to maximize their benefits.

Under the reward-cost structure as given above, the expected individual utility for the tagged customer who finds the server unavailable and decides to enter the orbit is

$$U(q) = R - CT(q), \tag{52}$$

where $T(q)$ is given in (39).

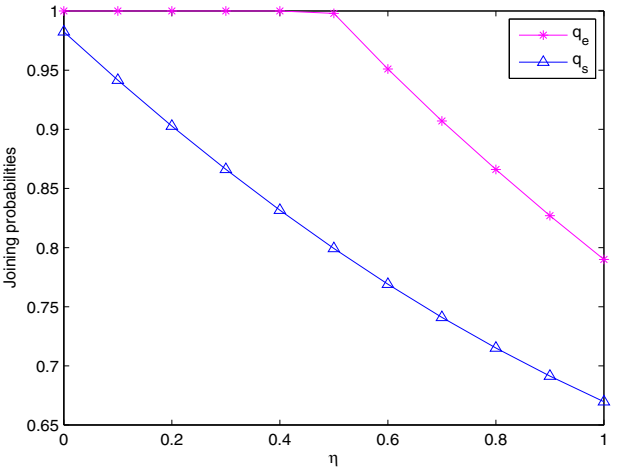

Figure 12. *q_e and q_s versus η ($\lambda = 1.8, \mu = 5, \beta = 1.2, \varphi = 0.8, \theta = 1.5, \delta = 4, \nu = 5, R = 9, C = 1$).*

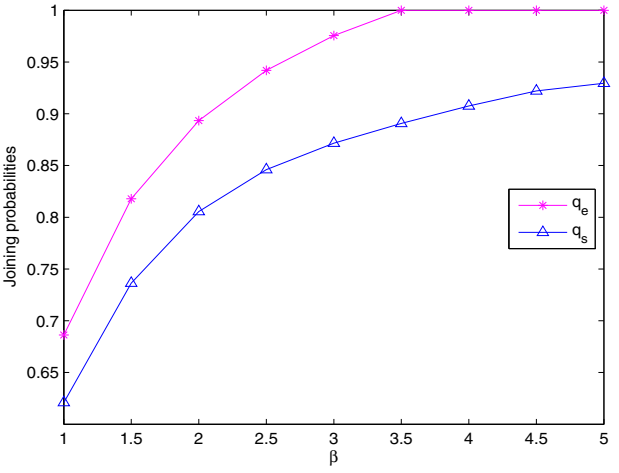

Figure 13. *q_e and q_s versus β ($\lambda = 1.8, \mu = 4.5, \varphi = 0.8, \theta = 1.5, \delta = 2, \eta = 1, \nu = 7, R = 9, C = 1$).*

When customers arrive at the system, they judge whether to join based on what they gain or lose. From the individual utility function, we can find the customers' equilibrium joining probability q_e as follows.

(i) If $U(1) > 0$, then the equilibrium strategy is $q_e = 1$, this means that when customers choose to join the system, their expected individual utility is positive.
(ii) If $U(0) < 0$, then the equilibrium strategy is $q_e = 0$, which means that when customers decide to join the system, their expected individual utility is negative.
(iii) In addition, the necessary and sufficient condition for $q_e \in (0, 1)$ to be an equilibrium joining probability is that $U(q_e) = 0$, this means that when customers choose to join the system, their expected individual utility is zero.

We continue with the problem of maximizing the social welfare per time unit.
The social welfare per time unit is

$$S(q) = \lambda^* R - CE(N), \tag{53}$$

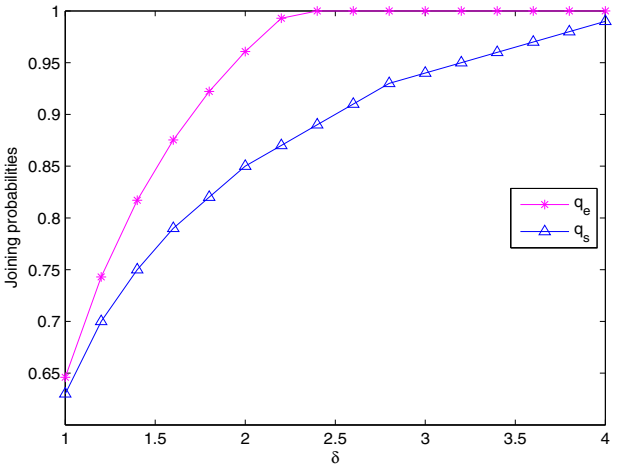

Figure 14. *q_e and q_s versus δ* ($\lambda = 1.8, \mu = 4.4, \beta = 2, \varphi = 0.5, \theta = 4.5, \eta = 1.3, \nu = 6, R = 9, C = 1$).

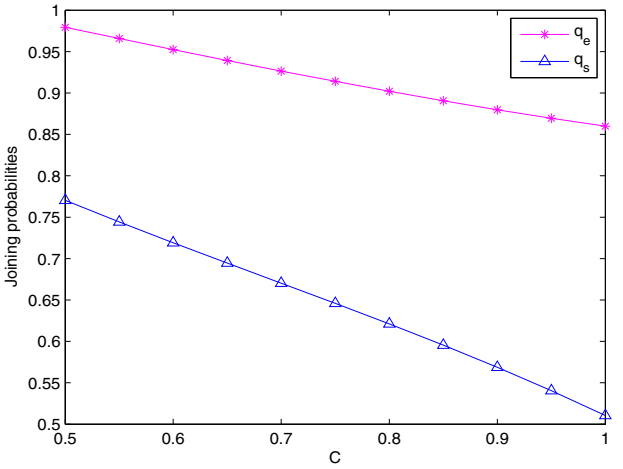

Figure 15. *q_e and q_s versus* C ($\lambda = 2.5, \mu = 5, \beta = 0.3, , \varphi = 0.1, \theta = 0.7, , \eta = 0.1, \delta = 5, \nu = 10, R = 9$).

where λ^* is the effective arrival rate of customers, and

$$\begin{aligned}\lambda^* &= \lambda G_0(1) + \lambda q[G_1(1) + G_2(1) + G_3(1) + G_4(1)] \\ &= \frac{\lambda\mu\kappa_3 \mathrm{A}_1 + \lambda q(\beta + \varphi)\mathrm{A}_2}{\kappa_2[\theta\delta\kappa_1 - \beta\mu\lambda\kappa_3]}\pi(0,1).\end{aligned} \tag{54}$$

From the above equation $S(q)$, a socially optimal strategy q_s is determined, which maximizes the social benefit per time unit.

Because the expressions obtained for individual and social benefits are very complex, it is hard to get specific results through traditional calculations. In the following analysis, we can use the particle swarm algorithm (PSO algorithm) to find the numerical solve this problem. When it comes to PSO algorithm, the most significant advantage is that it does not require too much analytic property of the objective function. It is an optimization algorithm based on swarm intelligence theory. During each iterative search, the particles in the swarm can dynamically adjust their position and velocity by tracking the two

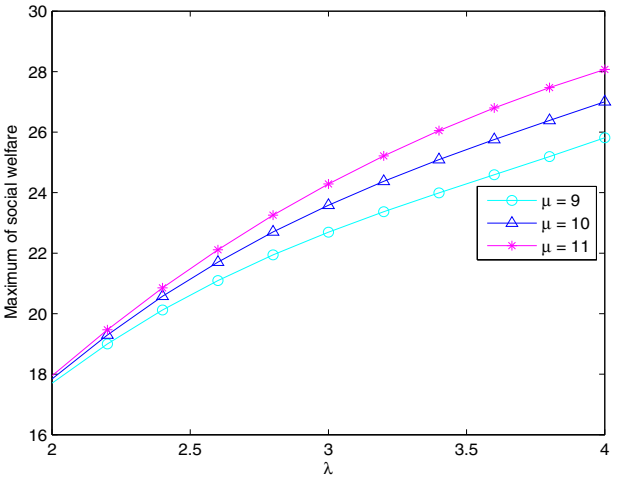

Figure 16. *Maximum of social welfare versus* λ *for different values of* μ $(\beta = 0.3, \varphi = 0.1, \theta = 0.8, \eta = 0.1, \delta = 5, \nu = 10, R = 9, C = 1)$.

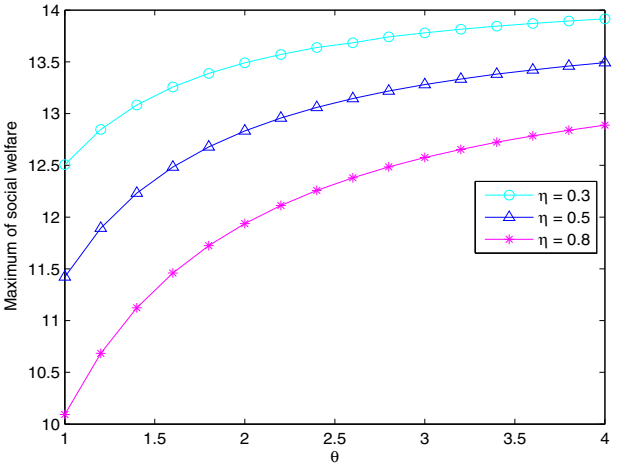

Figure 17. *Maximum of social welfare versus* θ *for different values of* η $(\lambda = 1.8, \mu = 5, \beta = 1.2, \varphi = 0.8, \delta = 4, \nu = 5, R = 9, C = 1)$.

extremes of the swarm: the optimal solution P-best found by the particle itself and the optimal solution G-best found by the swarm. Through many iterations, the global optimal solution can be obtained.

In Figure 11, both q_e and q_s are monotonically decreasing concerning λ. Because more customers join the system, the orbital load will increase, which reduces the probability of customers joining the system. From Figure 12, both q_e and q_s decrease with respect to η. As η increases, customers' waiting time on the orbit increases, which leads to a low probability of customers entering the system.

Figure 13 shows that the probabilities q_e and q_s increase with β. The repair rate increases, in this case, the server is quick to serve customers. In Figure 14, q_e and q_s increase with δ, delayed repair time becomes shorter, and the system performs repair services faster, so customers tend to join the system. As we expected, in Figure 15 q_e and q_s are monotonically decreasing concerning C. When the cost of waiting on the orbit increases, customers are unwilling to enter the system.

Figure 16 shows that the maximum social benefit increases as the arrival rate λ and service rate μ increase. The reason is that the number of customers in the system increases, which contributes to

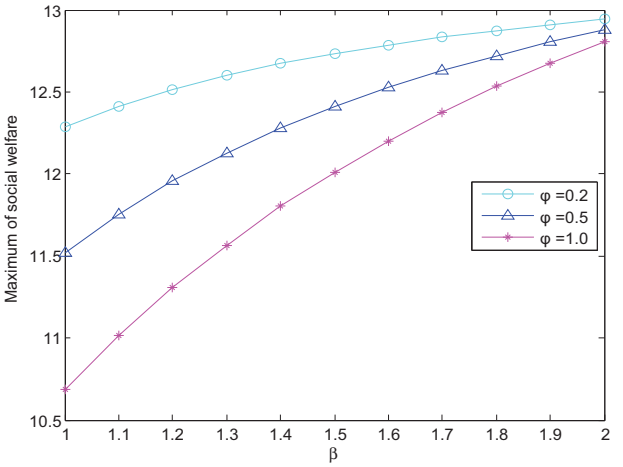

Figure 18. *Maximum of social welfare versus β for different values of φ* ($\lambda = 1.8, \mu = 4.5, \theta = 1.5, \eta = 1, \delta = 5, \nu = 7, R = 9, C = 1$) .

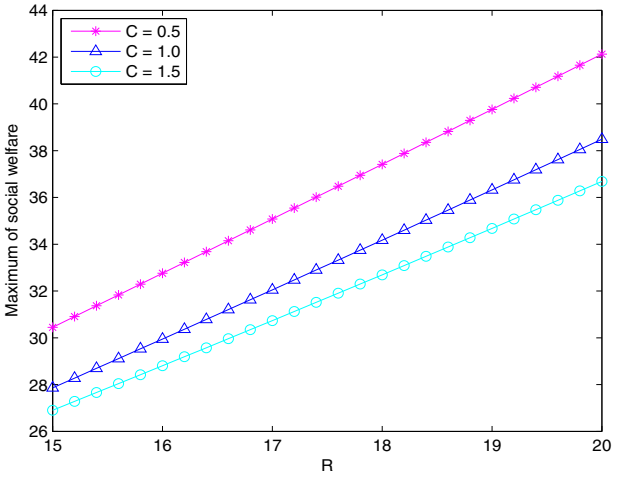

Figure 19. *Maximum of social welfare versus R for different values of C* ($\lambda = 2.5, \mu = 5, \beta = 0.3, \varphi = 0.1, \theta = 0.7, \eta = 0.1, \delta = 4, \nu = 10$) .

increasing the social benefits. On the other hand, as the service rate increases, customers' waiting times on the orbit decrease, so customers tend to join the system.

In Figures 17 and 18, the maximum social benefit increases as θ and β increase. Server availability increases due to the repair rate increased. However, the social benefit decreases as η and φ increase. As the breakdown rate increases, the server availability decreases. Thus, the social benefit decreases. It is shown in Figure 19 that the maximum social welfare increases as the reward R increases and decreases as the waiting cost C increases, which is consistent with our intuitive idea.

7. Conclusion

In this paper, we studied the M/M/1 retrial queueing system, which has passive breakdown with delayed repairs and active breakdown caused by negative customers during the busy period. Using the probability generating functions, we obtained the important performance measures, and we studied the effects of some parameters on the important performance measures of the model by numerical examples. We

proposed a cost function to determine the optimal parameter settings for the system under stationary conditions. Moreover, we analyzed the equilibrium joining strategy and the socially optimal joining probability of customers and numerically analyzed the impact of some parameters on the maximum social welfare. A possible future research direction is to consider equilibrium strategies for unreliable M/M/C retrial queueing systems with negative customers, passive breakdown and delayed repairs as well as finding the number of servers such that the system incurs minimum costs.

Funding statement. This research was supported by the National Natural Science Foundation of China under the grant no. 71971189.

References

[1] Bountali, O. & Economou, A. (2017). Equilibrium joining strategies in batch service queueing systems. *European Journal of Operational Research* 260(3): 1142–1151.
[2] Do, N.H., Van Do, T., & Melikov, A. (2020). Equilibrium customer behavior in the M/M/1 retrial queue with working vacations and a constant retrial rate. *Operational Research* 20(2): 627–646.
[3] Economou, A. & Kanta, S. (2008). Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters* 36(6): 696–699.
[4] Falin, G.I. (2010). A single-server batch arrival queue with returning customers. *European Journal of Operational Research* 201(3): 786–790.
[5] Gao, H., Zhang, J., & Wang, X. (2020). Analysis of a retrial queue with two-type breakdowns and delayed repairs. *IEEE Access* 8: 172428–172442.
[6] Gao, S., Dong, H., & Wang, X. (2021). Equilibrium and pricing analysis for an unreliable retrial queue with limited idle period and single vacation. *Operational Research* 21(1): 621–643.
[7] Hassin, R. (2016). *Rational Queueing*, Boca Raton: CRC Press.
[8] Hassin, R. & Haviv, M. (2003). *To queue or not to queue: equilibrium behavior in queueing systems*, Berlin: Springer Science & Business Media.
[9] Kim, J. & Kim, B. (2015). A survey of retrial queueing systems. *Annals of Operations Research* 247(1): 3–36.
[10] Krishna, K.B., Rukmani, R., Thanikachalam, A., & Kanakasabapathi, V. (2016). Performance analysis of retrial queue with server subject to two types of breakdowns and repairs. *Operational Research* 18(2): 521–559.
[11] Kulkarni, V.G. & Choi, B.D. (1990). Retrial queues with server subject to breakdowns and repairs. *Queueing systems* 7(2): 191–208.
[12] Naor, P. (1969). The regulation of queue size by levying tolls. *Econometric Society* 37: 15–24.
[13] Sherman, N.P. & Kharoufeh, J.P. (2006). An M/M/1 retrial queue with unreliable server. *Operations Research Letters* 34(6): 697–705.
[14] Wang, L., Liu, L., Wang, Z., & Chai, X. (2020). Strategic behavior and optimization in an unobservable constant retrial queue with balking and set-up time. *Journal of Systems Science and Information* 8(3): 273–290.
[15] Zhang, Z., Wang, J., & Zhang, F. (2014). Equilibrium customer strategies in the single-server constant retrial queue with breakdowns and repairs. *Mathematical Problems in Engineering* 2014: 1–14.
[16] Zhu, S., Wang, J., & Liu, B. (2020). Equilibrium joining strategies in the Mn/G/1 queue with server breakdowns and repairs. *Operational Research* 20(4): 2163–2187.

Cite this article: Han Y., Tian R., Wu X. and He L. (2024). On a retrial queue with negative customers, passive breakdown, and delayed repairs. *Probability in the Engineering and Informational Sciences* 38(2): 428–447. https://doi.org/10.1017/S0269964823000219