Hindawi Discrete Dynamics in Nature and Society Volume 2020, Article ID 1847512, 12 pages https://doi.org/10.1155/2020/1847512



Research Article

Performance Analysis of Queueing Systems with a Particular Service Interruption Discipline

Peng Liu, Tao Jiang, and Xudong Chai²

¹College of Economics and Management, Shandong University of Science and Technology, Qingdao 266590, China ²School of Science, Nanjing University of Science and Technology, Nanjing 210094, China

Correspondence should be addressed to Tao Jiang; jtao0728@163.com

Received 14 July 2020; Revised 20 November 2020; Accepted 25 November 2020; Published 10 December 2020

Academic Editor: Filippo Cacace

Copyright © 2020 Peng Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, queueing models with service interruption have been studied extensively due to their widespread applications. This paper is devoted to the study of an M/M/1 queue with a particular service interruption discipline under two types of the maintenance activities, namely the reactive maintenance and the planned maintenance. In the process of serving customers, when the system receives a negative feedback from customers, and in the next N customers to be served, if a negative feedback from the next N customers is received again, the service is interrupted and a reactive maintenance will be triggered immediately; otherwise, if all of the next certain amount of feedbacks are positive, the system administrator has a reason to believe that the occurrence of the negative feedback is an accidental phenomenon, and the reactive maintenance will not be initiated. By using the matrix analytic approach and spectral expansion method, we derive the steady-state probabilities, which are then used to compute the performance measures of the queueing system. Sensitivity analyses of parameters are included in this work for illustrative purposes, and from the perspective of queueing model, we also show that designing the queue size in terms of the steady-state probabilities is a more reasonable and rational method than the expected queue length.

1. Introduction

Queueing models with service interruption have been studied extensively due to their widespread applications including production systems, transportation systems, complex modern communication systems and service systems. So far, according to the emerged reasons, the literature on the service interruption can be roughly divided into two categories, depending on whether the interruption is caused by external factors (e.g., due to unexpected failures) or if it is caused by internal factors (e.g., due to the gradual deterioration of the servers (facilities)).

The category of service interruption, due to external factors, occurs in queueing systems that are subject to unexpected failure or breakdown. In previous studies, a large number of queueing researchers have investigated the queueing systems with repairable servers, where the maintenance action is carried out once the systems are breakdown. For example, Wartenhorst [1] considered an N parallel queueing system with server breakdown and repair. Gray et al. [2] studied a queueing model with multiple types of server breakdowns. Lam et al. [3] discussed a geometric process model for an M/M/1 queueing system with a repairable service station. By using the supplementary variable technique, the authors obtained some queueing characteristics as well as reliability measures of the service station. Yechiali [4] studied a queueing system with system disasters and impatient customers when the system is down. For more complex queueing systems with breakdown, Li et al. [5] studied a BMAP/G/1 retrial queue with a server subject to breakdowns and maintenances and derived the stationary availability, failure frequency, and queue length. Liu et al. [6] studied an M/G/1 retrial G-queue with preemptive resume and feedback under N-policy, where the server subjects to breakdowns and repairs. Cordeiro and Kharoufeh [7] examined an M/M/1 retrial queue with an unreliable server whose arrival, service, failure, repair, and retrial rates are all modulated by an exogenous random environment. Jain et al. [8] studied a fault-tolerant system with general distributed repair time, server vacation, and server breakdown. Ammar et al. [9] provided the transient analysis of impatient customers in an M/M/1 disasters queue in random environment. Recently, while the server in a queueing system suffers from unexpected failures and becomes defective, the system can be equipped with a substitute server which continues to provide service for the arriving customers, instead of stopping service completely. This phenomenon is also referred to working breakdown. Kalidass and Kasturi [10] first introduced the concept and applied the strategy to the facility replacement problems. Later, a variety of research efforts have contributed to the topic and its applications in manufacturing systems. For example, Kim and Lee [11] investigated an M/G/1 queue with disasters and working breakdown, where the system can provide service by an auxiliary machine while the main machine is in maintenance period. Liu and Song [12], Jiang and Liu [13], respectively, considered a batch arrival M^[X]/M/1 queueing system and a GI/M/1 queueing system with working breakdown. Meanwhile, in [13], the authors gave the applications of working breakdown in an unreliable multiproduct manufacturing network and the wireless sensor networks. Recently, Yang and Wu [14] investigated a finite-capacity Markovian queueing system with working breakdowns, reneging, and retention of impatient customers. Ye and Liu [15] studied an MAP/M/1 queue with working breakdowns. Jiang and Xin [16] investigated a queueing system with working breakdowns and delaying repair under a Bernoulli-schedulecontrolled policy by the matrix analytic method and the spectral expansion method. Ammar and Rajadurai [17] investigated a preemptive priority retrial queueing system with disaster under working breakdown services.

The category of service interruption, due to internal factors, is triggered by the gradual deterioration of the servers. In this category, the so-called condition-based maintenance, which utilizes the finished item quality as the intermediate to monitor the server condition, is often used in the production systems. Under the condition-based maintenance, maintenance action is implemented by setting a control chart to monitor the operating condition of the servers. Once a defective item is detected, the control chart will generate an alert signal; then, the service is interrupted and the server will experience a maintenance period immediately. In recent decades, this type of service interruption with condition-based maintenance has been extensively investigated in the literature. For example, Zhou and Zhu [18] gave an economic design of the integrated model of the control chart and maintenance management. Liu et al. [19] applied the condition-based maintenance to give the economic and economic-statistical designs of a control chart for two-unit series systems. Zhou et al. [20] considered two types of maintenance activities and studied a trade-off between energy consumption and customer delay based on the queueing approach. For more related papers on the maintenance method, the interested readers are referred to Cha

et al. [21], Yang et al. [22], and Alaswad and Xiang [23], and the references therein.

The underlying queueing model is motivated by a problem that involves the maintenance planning in real service systems. For example, in some consulting service centers, customers may be asked for taking a few moments of their time to fill out a Customer Service Feedback Form about their service. If a customer is satisfied with the service, he/she will respond positively; otherwise, he/she will make a negative response. If the consulting service center receives a negative response on the service, in order to ensure whether the negative response is an accidental phenomenon or not, the consulting service center should continue to check the next certain amount of responses from customers. During the next certain amount of responses, if a negative response appears again, the consulting service center will do some adjustments and a reactive maintenance should be scheduled; otherwise, if all of the next certain amount of responses are positive, the consulting service center has a reason to believe that the occurrence of the negative response is an accidental phenomenon, and the adjustments will not be initiated. In addition, the underlying queueing model can be also applied to production systems. For example, in a real production system, equipment may deteriorate gradually due to the accumulation of wear and corrosion. In order to identify the operating condition of the facility, the system usually employs the quality characteristic of items as the intermediate to monitor the facility condition. Then, the system is equipped with a quality control detector and monitored using the Shewhart control charts to check the quality of the finished items. If the quality characteristic of a finished item falls outside the so-called "in-control" region, i.e., a defective item is detected, then, the control chart will generate an alert signal. However, in order to ensure the efficiency of the system, the inspector should estimate the defective item is whether an accidental phenomenon or not, i.e., the inspector should continue to check the next certain amount of finished items. During the next certain amount of finished items, if a defective item is detected again, it is reasonable to believe that the operating condition of the equipment is in a danger state and a reactive maintenance should be scheduled; otherwise, if all of the next certain amount of finished items are acceptable, the inspector has a reason to believe that the occurrence of the defective item is an accidental phenomenon, and the reactive maintenance will not be initiated.

Different from a recent study [24], in which the author considers a clearing queueing model with server maintenance after N negative feedbacks, in this paper, we describe the service process of the queueing system and the maintenance policies as a single-server queue with two types of maintenance activities for the server, i.e., planned maintenance and reactive maintenance. To the best of our knowledge, the underlying queueing model with service interruption has not appeared in the literature. The derived results of the queueing model could be applied to deal with more practical queueing problems and provide important reference value for system administrator to scientifically design the scheduled maintenance policy. For more details

of the common methods on solving the queueing models, the interested readers are referred to [25, 26], which give a detailed introduction on the idea and application of the methods. Next, the main contributions of the paper are as follows:

- (i) Model: We consider a novel and special service interruption discipline with two types of maintenance activities for the underlying queueing system. The model is suitable for reflecting the characteristics of customers' services for the modern service systems and production systems.
- (ii) Methodology and results: The matrix analytic approach and spectral expansion method are applied to analyze the special service interruption discipline in queueing problems. Compared to other methods (e.g., probability generating functions), the two methods could effectively give a detailed computational procedure to derive the steady-state probthe performance measures. and Furthermore, there are always finite number of solution techniques for solving any given problem mathematically; therefore, we feel that the considered model extends the analytical literature of queueing theory by the intricate successful formulation of the problem.
- (iii) Numerical illustrations: Sensitivity analysis of parameters is included in this work for illustrative purpose, and some qualitative take-away messages are also provided and intuitively explained. Moreover, by some numerical examples, we show that designing the queue size in terms of the steady-state probabilities is a more reasonable and rational method than the expected queue length.

The underlying article is organized as follows. Section 2 describes the proposed queueing model. In Section 3, we present the stability analysis of the underlying queueing system. Some numerical examples are conducted to show the impacts of system parameters on the performance measures in Section 4. Section 5 concludes the paper.

2. Model Description

In practice, it is more common to model the queueing system as a queue with multiple servers. To simplify the model, in this paper, we will study the queueing model based on an M/M/1 queue with two types of maintenance activities for the server (the reactive maintenance and the planned maintenance), where the reactive maintenance is triggered by a particular service interruption discipline, which will be described in the following section, and the frequency of planned maintenance is determined by the system administrator, which can be represented by a proportion parameter at the beginning of each idle period. We believe the simplified model could also derive similar results. Below, we use the related parameters of queueing system to describe the underlying queueing model. Our assumptions are driven by

the actual problem that motivates the analysis and our desire for keeping the model as simple as possible.

- (i) Customers arrive according to a Poisson process with rate λ and customers are served based on a first-come, first-served (FCFS) discipline.
- (ii) Uncertain service times at server are exponentially distributed with parameter μ .
- (iii) Once a customer is served, the served customer is required to give a response (feedback) on the quality of customer service. If he/she is satisfied with the service, he/she will respond positively on his/her service; otherwise, he/she will make a negative response for his/her horrible or unpleasant service. To this end, we consider p as the probability that a served customer gives a negative feedback and q =(1-p) as the probability that a served customer gives a positive feedback on the service. Next, we describe the particular service interruption discipline and show that when the reactive maintenance will be triggered. In the process of serving customers, when the system is in the normal working state and a negative feedback is received from the served customers, an alert signal is immediately generated; meanwhile, the system goes into a warning working state and checks the next N customers to be served, so as to ensure the negative feedback is whether an accidental phenomenon or not. In the next N customers to be served, if a negative feedback from the N customers is received again, the reactive maintenance will be scheduled immediately; otherwise, if the next N feedbacks from the N customers are positive, the system administrator believes that the occurrence of the negative feedback in the normal working state is an accidental phenomenon; therefore, the reactive maintenance will not be initiated, and the system returns to the normal working state.
- (iv) Beside the reactive maintenance, planned maintenance is also adopted according to a frequency parameter at the beginning of an idle period to ensure the system in the normal operation. For analytical simplicity, we use a proportion parameter β to denote the frequency parameter (i.e., $\beta=0.2$ represents that the planned maintenance is carried out every 5 idle periods). We also assume that if there is no customer while the system is in the warning working state, the server will also schedule a planned maintenance.
- (v) The reactive maintenance time is assumed to follow an exponential distribution with a rate ξ_1 , the planned maintenance time is assumed to follow an exponential distribution with a rate ξ_2 .
- (vi) In addition, we further assume that the arrival process, the service process, and the maintenance process are mutually independent. Since the checking time is relatively short compared with the

expected service time, then the checking time can be negligible.

3. Model Analysis

Based on the model description, we will first give the stability analysis for the underlying queueing system, and then we use the matrix analytic method and spectral expansion method to analyze the queueing model. Actually, the other possible method of analysis is via probability generating functions which, in this case, will be more complex. Moreover, because of the complicated structure of the queueing model, the steady-state probability vector may be difficult to obtain in a closed form. However, the two methods in consideration could effectively give a detailed computational procedure to derive the steady-state probabilities and the performance measures of the underlying queueing model.

3.1. Stability Condition. Let N(t) denote the number of customers in the system at time t, I(t) denotes the state of the server at time t, which is defined as

$$I(t) = \begin{cases} -1, & \text{the server is in the reactive maintenance state at time } t, \\ 0, & \text{the server is in the planned maintenance state at time } t, \\ 1, & \text{the server is in the normal working state at time } t, \\ 2, & \text{the server is in the warning working state at time } t. \end{cases}$$

(1)

As the server is in the warning state at time t, i.e., I(t) = 2, then, K(t) = k, $0 \le k \le N - 1$ denotes that when the system goes into the warning working state, the next k feedbacks from customers are positive. Obviously, $\{(N(t), I(t), K(t)), t \ge 0\}$ is a continuous-time Markov chain with the state space

$$\{(n,-1), n \ge 0\} \cup \{(n,0), n \ge 0\} \cup \{(n,1), n \ge 0\} \cup \{(n,2,k), n > 0, 0 \le k \le N-1\}. \tag{2}$$

Considering the variety of system states, the transition rate diagram of the continuous-time Markov chain is shown in Figure 1.

Next, according to the transition rate diagram and using the lexicographical sequence for the aforementioned states, the infinitesimal generator matrix of the continuous-time Markov chain in the block-partitioned form can be obtained as follows:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C} & \mathbf{B}_0 & 0 & 0 & \cdots \\ \mathbf{B}_2 & \mathbf{A}_1 & \mathbf{A}_0 & 0 & \cdots \\ 0 & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \cdots \\ 0 & 0 & \mathbf{A}_2 & \mathbf{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \tag{3}$$

 $\mathbf{C} = \begin{pmatrix} -(\lambda + \xi_{1}) & 0 & \xi_{1} \\ 0 & -(\lambda + \xi_{2}) & \xi_{2} \\ 0 & 0 & -\lambda \end{pmatrix}, \mathbf{B}_{0} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},$ $\mathbf{B}_{2} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \mu p + \mu q \beta & \mu q (1 - \beta) \\ \mu p & \mu q & 0 \\ \vdots & \vdots & \vdots \\ \mu p & \mu q & 0 \end{pmatrix}_{(N+3) \times 3},$ (4)

and the other partitioned matrices are square ones with $(N+3)\times(N+3)$ orders,

where

$$\mathbf{A}_{2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mu q & \mu p & 0 & 0 & \cdots & 0 \\ \mu p & 0 & 0 & 0 & \mu q & 0 & \cdots & 0 \\ \mu p & 0 & 0 & 0 & \mu q & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu p & 0 & 0 & 0 & 0 & \cdots & \mu q \\ \mu p & 0 & \mu q & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\mathbf{A}_{0} = \operatorname{diag}(\lambda, \lambda, \dots, \lambda),$$

$$\mathbf{A}_{1} = \begin{pmatrix} -(\lambda + \xi_{1}) & 0 & \xi_{1} & 0 & \cdots & 0 \\ 0 & -(\lambda + \xi_{2}) & \xi_{2} & 0 & \cdots & 0 \\ 0 & 0 & -(\lambda + \mu) & 0 & \cdots & 0 \\ 0 & 0 & 0 & -(\lambda + \mu) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -(\lambda + \mu) \end{pmatrix}.$$

$$(5)$$

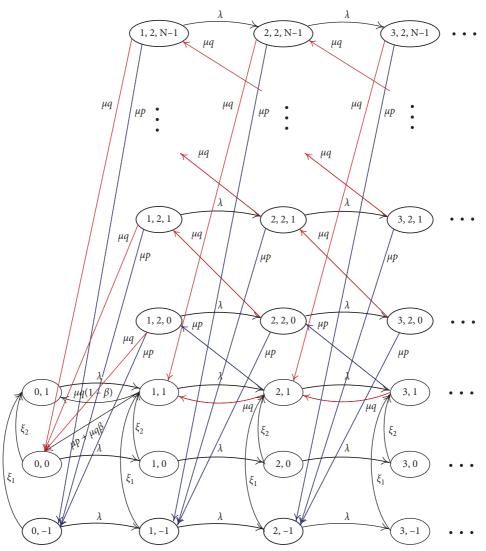


FIGURE 1: Transition rate diagram of the queueing system.

In order to analyze the queueing model, we first study the stability condition. According to the aforementioned state-transition-rate matrix, the stability condition can be derived in the following lemma.

Lemma 1. The underlying queueing system is stable if and only if

$$\lambda < \left(\frac{\left(2\xi_{1} - \xi_{1}q^{N}\right)\mu}{\mu p\left(1 - q^{N}\right) + 2\xi_{1} - \xi_{1}q^{N}}\right). \tag{6}$$

Proof. Based on Theorem 1.7.1 in Neuts [25], the underlying queueing system is stable if and only if

$$\mathbf{x}\mathbf{A}_{0}\mathbf{e} < \mathbf{x}\mathbf{A}_{2}\mathbf{e},\tag{7}$$

where $\mathbf{x} = (x_0, x_1, x_2, x_3, \dots, x_{N+2})$ is the invariant probability vector of $\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2$, which satisfies $\mathbf{x}\mathbf{A} = 0$ and $\mathbf{x}\mathbf{e} = 1$, \mathbf{e} is a column vector with N+3 dimensions and all its elements are equal to one. From the expressions of matrices $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$, we find that

(6)
$$\mathbf{A} = \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2 = \begin{pmatrix} -\xi_1 & 0 & \xi_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -\xi_2 & \xi_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -\mu p & \mu p & 0 & 0 & \cdots & 0 \\ \mu p & 0 & 0 & -\mu & \mu q & 0 & \cdots & 0 \\ \mu p & 0 & 0 & 0 & -\mu & \mu q & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu p & 0 & 0 & 0 & 0 & 0 & \cdots & \mu q \\ \mu p & 0 & \mu q & 0 & 0 & 0 & \cdots & -\mu \end{pmatrix},$$
(7) (8)

is irreducible, then, solving the equations $\mathbf{x}\mathbf{A} = 0$ and $\mathbf{x}\mathbf{e} = 1$, we have

$$x_{0} = \frac{\mu p(1 - q^{N})}{\xi_{1}} x_{2},$$

$$x_{1} = 0,$$

$$x_{2} = \frac{\xi_{1}}{\mu p(1 - q^{N}) + 2\xi_{1} - \xi_{1}q^{N}},$$

$$x_{3} = px_{2}, x_{N+2} = qx_{N+1} = q^{2}x_{N} = \dots = q^{N-1}x_{3}.$$
(9)

Therefore, from $\mathbf{x}\mathbf{A}_0\mathbf{e} < \mathbf{x}\mathbf{A}_2\mathbf{e}$, the sufficient and necessary stability condition can be further derived by

$$\lambda < \frac{\left(2\xi_1 - \xi_1 q^N\right)\mu}{\mu p(1 - q^N) + 2\xi_1 - \xi_1 q^N}.$$
 (10)

Next, in order to analyze the queueing model in stability condition, we first denote the steady-state vector of the queueing model as $\pi = (\pi_0, \pi_1, \pi_2, \dots)$, where

$$\pi_{0} = (\pi_{0,-1}, \pi_{0,0}, \pi_{0,1}), \\
\pi_{n} = (\pi_{n,-1}, \pi_{n,0}, \pi_{n,1}, \pi_{n,2,0}, \dots, \pi_{n,2,N-1}), \quad n \ge 1, \\
\pi_{0,i} = \lim_{t \longrightarrow \infty} P(N(t) = 0, I(t) = i), \quad i = -1, 0, 1, \\
\pi_{n,2,k} = \lim_{t \longrightarrow \infty} P(N(t) = n, I(t) = 2, K(t) = k), \quad k = 0, 1, \dots, N-1, n \ge 1.$$

3.2. Matrix Analytic Method. From the special structure of matrix Q, we find that $\{(N(t), I(t), K(t)), t \ge 0\}$ is a quasibirth-and-death (QBD) process, whose state-transition-rate matrix is block tridiagonal. Therefore, in this subsection, we first use the matrix analytic method to derive the steady-state probabilities for the underlying queueing system. The most important step in applying this method is to obtain the rate matrix \mathbf{R} , which is the minimal nonnegative solution of

$$\mathbf{R}^2 \mathbf{A}_2 + \mathbf{R} \mathbf{A}_1 + \mathbf{A}_0 = 0. \tag{12}$$

According to the structure of matrices \mathbf{A}_0 , \mathbf{A}_1 , \mathbf{A}_2 , we find that it is difficult to obtain the explicit expression of rate matrix \mathbf{R} . However, based on the theory in [25], the matrix equation can be typically solved numerically using the following iteration procedure. Consider the sequence of matrices $\{\mathbf{R}(n), n \geq 0\}$, obtained by successive substitutions, starting with $\mathbf{R}(0) = 0$ and then $\mathbf{R}(n) = -\mathbf{A}_2\mathbf{A}_1^{-1} - \mathbf{R}^2(n-1)\mathbf{A}_2\mathbf{A}_1^{-1}$. If there exist $\varepsilon > 0$, such that $\|\mathbf{R}(n) - \mathbf{R}(n-1)\|_{\infty} < \varepsilon$, then we stop the iterative procedure. Because of the convergence of the iterative algorithm, i.e., $\lim_{n\longrightarrow\infty} \mathbf{R}(n) = R$, then the obtained $\mathbf{R}(n)$ is approximately equal to \mathbf{R} . After \mathbf{R} is obtained, according to the matrix analytic method, we have

$$\pi_{\mathbf{n}} = \pi_1 \mathbf{R}^{n-1}, \quad n \ge 2.$$
(13)

Next, the boundary vector π_0, π_1 can be obtained by solving the equations

$$(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1) \mathbf{B}[\mathbf{R}] = 0,$$

$$\boldsymbol{\pi}_0 \mathbf{e}_1 + \boldsymbol{\pi}_1 (\mathbf{I} - \mathbf{R})^{-1} \mathbf{e} = 1,$$
 (14)

where

$$\mathbf{B}[\mathbf{R}] = \begin{pmatrix} \mathbf{C} & \mathbf{B}_0 \\ \mathbf{B}_2 & \mathbf{A}_1 + \mathbf{R} \mathbf{A}_2 \end{pmatrix}, \mathbf{e}_1 = (1, 1, 1)'. \tag{15}$$

Then, the steady-state probabilities can be obtained by the following theorem.

Theorem 1. If the stability condition is satisfied, then, the steady-state probabilities can be obtained as follows

$$\pi_{0} = (\pi_{0,-1}, \pi_{0,0}, \pi_{0,1}) = -\pi_{1} \mathbf{B}_{2} \mathbf{C}^{-1},
\pi_{\mathbf{n}} = (\pi_{n,-1}, \pi_{n,0}, \pi_{n,1}, \pi_{n,2,0}, \dots, \pi_{n,2,N-1}) = \pi_{1} \mathbf{R}^{\mathbf{n}-1}, \quad n \ge 2,
(16)$$

where π_1 satisfies the following set of equations:

$$\pi_1 \left(\mathbf{A}_1 + \mathbf{R} \mathbf{A}_2 - \mathbf{B}_2 \mathbf{C}^{-1} \mathbf{B}_0 \right) = 0,$$

$$\pi_1 \left(\mathbf{I} - \mathbf{R} \right)^{-1} \mathbf{e} - \pi_1 \mathbf{B}_2 \mathbf{C}^{-1} \mathbf{e}_1 = 1.$$
 (17)

3.3. Spectral Expansion Method. Next, we turn to use an alternate method called spectral expansion method to derive the steady-state probabilities for the underlying queueing system. From $\pi \mathbf{Q} = 0$, we have

$$\pi_{i-1}A_0 + \pi_iA_1 + \pi_{i+1}A_2 = 0, \quad i \ge 2,$$
 (18)

associated with it is the characteristic matrix polynomial, which has the following structure:

$$\mathbf{Q}(x) = \mathbf{A}_0 + \mathbf{A}_1 x + \mathbf{A}_2 x^2 = \begin{pmatrix} f_{01}(x) & 0 & \xi_1 x & 0 & 0 & 0 & \cdots & 0 \\ 0 & f_{11}(x) & \xi_2 x & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & f_2(x) & \mu p x^2 & 0 & 0 & \cdots & 0 \\ \mu p x^2 & 0 & 0 & f_3(x) & \mu q x^2 & 0 & \cdots & 0 \\ \mu p x^2 & 0 & 0 & 0 & f_3(x) & \mu q x^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \mu p x^2 & 0 & 0 & 0 & 0 & 0 & \cdots & \mu q x^2 \\ \mu p x^2 & 0 & \mu q x^2 & 0 & 0 & 0 & \cdots & f_3(x) \end{pmatrix},$$
(19)

where

 $f_{01}(x) = \lambda - (\lambda + \xi_1)x,$ of $\mathbf{Q}(x)$: $f_{11}(x) = \lambda - (\lambda + \xi_2)x,$ (20)

$$\begin{aligned} |\mathbf{Q}(x)| &= f_{01}(x) f_{11}(x) f_2(x) \left(f_3(x) \right)^N + (-1)^N \left(\mu p x^2 \right) \left(\mu q x^2 \right)^N f_{01}(x) f_{11}(x) \\ &+ \left(\xi x \right) \left(\mu p x^2 \right)^2 f_{11}(x) \sum_{i=0}^{N-1} \left(-1 \right)^i \left(\mu q x^2 \right)^i \left(f_3(x) \right)^{N-i-1}. \end{aligned} \tag{21}$$

Referring to Mitrani and Chakka [26], if the underlying queueing system is stable, then the number of eigenvalues in the unit disk of $\mathbf{Q}(x)$ is equal to the degree of $\mathbf{Q}(x)$; that is, N+3 eigenvalues should lie inside the unit disk. Then, the steady-state probabilities have the following form:

 $f_2(x) = \lambda - (\lambda + \mu)x + \mu q x^2$,

 $f_3(x) = \lambda - (\lambda + \mu)x.$

$$\pi_{j} = \sum_{i=0}^{N+2} a_{i} x_{i}^{j-1} \psi_{i}, \quad j \ge 1,$$
 (22)

where x_i (i = 0, 1, ..., N + 2) are the eigenvalues inside the unit disk, ψ_i (i = 0, 1, ..., N + 2) are the corresponding left eigenvectors, coefficients a_i (i = 0, 1, ..., N + 2) can be

determined from the balance equations, and the eigenvalueeigenvector pairs (x_i, ψ_i) of $\mathbf{Q}(x)$ satisfy

From the expression of $\mathbf{Q}(x)$, we have the determinant

$$\psi_i \mathbf{Q}(x_i) = 0, \det(\mathbf{Q}(x_i)) = 0, \quad i = 0, 1, \dots, N+2.$$
 (23)

Solving the two matrix equations, the eigenvectors corresponding to x_i (i = 0, 1, 2, ..., N + 2) can be obtained as

$$\psi_{i} = (\psi_{i,0}, \psi_{i,1}, 1, \psi_{i,3}, \dots, \psi_{i,N+2}), \tag{24}$$

where

$$\psi_{i,0} = -\left(\frac{\mu p x_i^2 T(x_i) (1 - (S(x_i))^N)}{f_{01}(x_i) (1 - S(x_i))}\right) = -\left(\frac{f_2(x_i) + \mu q x_i^2 T(x_i) (S(x_i))^{N-1}}{\xi_1 x_i}\right),$$

$$\psi_{i,1} = 0, \psi_{i,3} = T(x_i),$$

$$\psi_{i,k} = \left(S(x_i))^{k-1} \psi_{i,3}, k = 4, 5, \dots, N+2, S(x) = -\frac{\mu q x^2}{f_3(x)}, T(x) = -\frac{\mu p x^2}{f_3(x)}.$$
(25)

After deriving the eigenvalues and their corresponding left eigenvectors, the coefficients a_i (i = 0, 1, 2, ..., N + 2) can be directly computed by the balance equations and the

normalization condition. From $\pi_0 \mathbf{C} + \pi_1 \mathbf{B}_2 = 0$, π_0 can be obtained in terms of π_1 . Then, combining the expression of π_1 and solving the set of equations

$$\pi_0 \mathbf{B}_0 + \pi_1 \mathbf{A}_1 + \pi_2 \mathbf{A}_2 = 0,$$

$$\pi_0 \mathbf{e}_1 + \sum_{j=1}^{\infty} \pi_j \mathbf{e} = 1,$$
(26)

N+3 unknown coefficients $a_0, a_1, a_2, \ldots, a_{N+2}$ can be derived, and hence the steady-state probabilities.

- 3.4. Performance Measures. From the derived steady-state probabilities, in this subsection, we will obtain some performance measures. Let E[L] and E[W] denote the expected number of customers in the system and the mean sojourn time of a customer needing to be served in the system. Then, some performance measures could be obtained as follows:
 - (i) The probability that the server is idle: $P_I = \pi_{0,1}$.
 - (ii) The probability that the server is busy in the normal working state: $P_B = \sum_{n=1}^{\infty} \pi_{n,1}$.
 - (iii) The probability that the server is in the warning working state: $P_w = \sum_{k=0}^{N-1} \sum_{n=1}^{\infty} \pi_{n,2,k}$.
 - (iv) The probability that the server is in the planned maintenance state: $P_{\rm pm} = \sum_{n=0}^{\infty} \pi_{n,0}$. The probability that the server is in the reactive maintenance state: $P_{\rm rm} = \sum_{n=0}^{\infty} \pi_{n,-1}$.

The expected number of customers in the system is

$$E[L] = \sum_{k=1}^{\infty} k \pi_k \mathbf{e} = \pi_1 (I - R)^{-2} \mathbf{e}, \qquad (27)$$

or

$$E[L] = \sum_{j=1}^{\infty} j \pi_j \mathbf{e} = \sum_{i=0}^{N+2} a_i \left(\frac{1}{(1-x_i)^2} \right) \psi_i \mathbf{e}.$$
 (28)

By using Little's law, the mean sojourn time of a customer needing to be served is

$$E[W] = \frac{E[L]}{\lambda} = \frac{\pi_1 (I - R)^{-2} \mathbf{e}}{\lambda} = \sum_{i=0}^{N+2} \left(\frac{a_i}{\lambda (1 - x_i)^2} \right) \psi_i \mathbf{e}.$$
(29)

4. Numerical Results

In this section, we would provide a set of numerical examples to give a comparison between the underlying queueing model and its special case (N=0). Moreover, sensitivity analysis of parameters is included in this section for illustrative purpose, and some qualitative take-away messages are provided and intuitively explained.

We first derive the steady-state probabilities for the special case of the underlying queueing model (queueing model with the condition-based maintenance). If we assume that N=0, this means that, once the system receives a negative feedback, the reactive maintenance will be triggered immediately. Let $N^*(t)$ denote the number of customers in the system at time t and $I^*(t)$ denote the state of the server at time t; then the state-transition-rate matrix of this model can be obtained as follows:

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{C}^* & \mathbf{A}_0^* & 0 & 0 & \cdots \\ \mathbf{B}_2^* & \mathbf{A}_1^* & \mathbf{A}_0^* & 0 & \cdots \\ 0 & \mathbf{A}_2^* & \mathbf{A}_1^* & \mathbf{A}_0^* & \cdots \\ 0 & 0 & \mathbf{A}_2^* & \mathbf{A}_1^* & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \tag{30}$$

where all partitioned matrices are square ones with 3×3 orders

$$\mathbf{C}^{*} = \begin{pmatrix}
-(\lambda + \xi_{1}) & 0 & \xi_{1} \\
0 & -(\lambda + \xi_{2}) & \xi_{2} \\
0 & 0 & -\lambda
\end{pmatrix}, \mathbf{B}_{2}^{*} = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
\mu p & \mu q \beta & \mu q (1 - \beta)
\end{pmatrix}, \mathbf{A}_{3}^{*} = \begin{pmatrix}
\lambda & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & -(\lambda + \xi_{2}) & \xi_{2} \\
0 & 0 & -(\lambda + \mu)
\end{pmatrix}, \mathbf{A}_{2}^{*} = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
\mu p & 0 & \mu q
\end{pmatrix}. \tag{31}$$

In this case, the sufficient and necessary stability condition is

$$\lambda < \left(\frac{\mu \xi_1}{\mu p + \xi_1}\right). \tag{32}$$

Similar to the analysis in Section 3, if the stability condition is satisfied, we can obtain the steady-state probabilities.

Next, based on the theoretical framework given by the above analysis, we present some figures below to investigate the impact of the parameters on the expected number of customers E[L]. We first assume that $\xi_1 = 0.6$, $\xi_2 = 0.8$, p = 0.1, and $\beta = 0.2$ and then provide a comparison between the cases N = 6 and N = 0.

From Figure 2, we find that E[L] is increasing in λ , while decreasing in μ . It is noteworthy that, for the curves in Figure 2, if values λ and μ are fixed, the value of E[L] in the case N=0 is greater than the case N=6. It is reasonable that, as N=0, once the system receives a negative feedback from customers, the reactive maintenance will be triggered immediately; therefore, the case could increase the reactive maintenance frequency. However, high-frequency

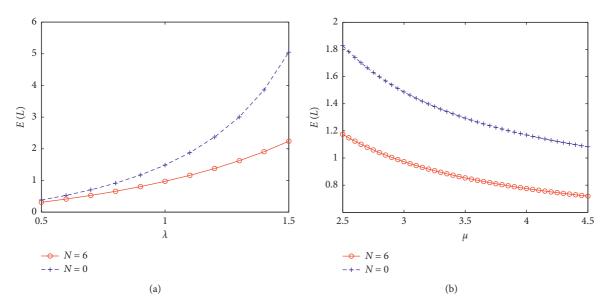


FIGURE 2: E[L] versus λ and μ for N=6 and N=0 (a) $E[L]vs\lambda(\mu=3)$; (b) $E[L]vs\mu(\lambda=1)$.

maintenance will lead to a much longer queue length to the system. On the contrary, as N>0, the system will verify whether the negative feedback is an accidental phenomenon or not. Compared to the case N=0, the case N=6 will lower the reactive maintenance frequency, which could satisfy the short queue length requirement. Thus, from the numerical result, to reduce the holding cost for customers and shorten queue length (or the customer sojourn time), the system administrator should verify whether the negative feedback is an accidental phenomenon or not, so that the reactive maintenance frequency could be properly lowered, and the cost could be reduced.

Next, we assume that $\lambda = 1.5$, $\mu = 3$, $\xi_1 = 0.6$, $\xi_2 = 0.8$, and p = 0.1 and pay attention to the curves of E[L] with the change of N from 1 to 20 for the different values of β . Some interesting managerial insights can be derived from Figure 3. First, E[L] first increases significantly and then tends to be smooth. In the author's opinion, the phenomenon may be caused by the value of p, i.e., the negative feedback percentage (meaning statistically the event may happen every (1/(1-p)) = (1/q) experiments on average). For a fixed p, if N > (1/q), the system will be more likely to receive a negative feedback in the next N feedbacks, which means that the increase of N leads to a high-frequency reactive maintenance and a much longer queue length to the system. However, when $N \gg (1/q)$, the designed inspections/detections do not have chance to be realized, i.e., the reactive maintenance will always be triggered; then, in this case, E[L]tends to be smooth as N increases, and the parameter N has a weak impact on E[L]. According to our model description, we assume that p remains unchanged while the system is in the warning working state. Thus, from the numerical result, the maximum value of the next number of feedbacks needs to be checked in the warning working state is a certain value; when N exceeds this value, the curve becomes smooth. Moreover, E[L] is increasing in β . The reason is that highfrequency planned maintenance could also contribute a

much longer queue length. Thus, according to the negative feedback percentage, the system administrator needs to choose a suitable number of feedbacks from customers to check while the system is in the warning working state. Moreover, the frequency of planned maintenance should be properly lowered, so that the queue length of the system could be reduced, the small sojourn time requirement could be satisfied, and the holding cost for customers could be saved.

Finally, we assume that $\lambda = 1.5$, $\mu = 3$, $\xi_1 = 0.6$, and $\xi_2 =$ 0.8 and explore the impact of the parameters N and β on E[L] for the different values of p. As expected, Figure 4 reveals that for the fixed values N and β , E[L] increases with the increase of p. Thus, from the numerical results, the system administrator should control the negative feedback percentage so as to reduce the queue length of the system and shorten the customer sojourn time. Also, from Figure 4(a), the smaller p is, the smoother of the curve becomes. Moreover, from Figure 4(b), we also find that E[L] increases with the increase of β . The reason is that as β increases, the system has a higher probability to proceed a planned maintenance process when the system is empty, i.e., the system has a higher frequency to carry out the planned maintenance, which could contribute a much longer queue length to the system. Thus, to satisfy the short queue length requirement, the system administrator should lower the frequency of the planned maintenance.

In short, these numerical results can provide important reference value for system administrator to scientifically design the parameter of the maintenance policy to save cost and keep the server staying in an efficient operation state.

From the perspective of the queueing model, it is well known that if the queue capacity design is too large, the construction and operation costs of the system will be too high. On the contrary, if the capacity of the queue is too small, the arriving customers will be lost due to insufficient capacity, which will cause an economic loss and bring a

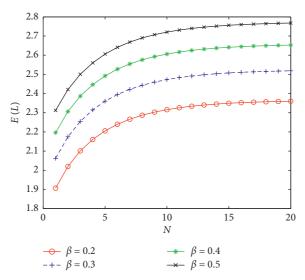


FIGURE 3: E[L] versus N for different values β .

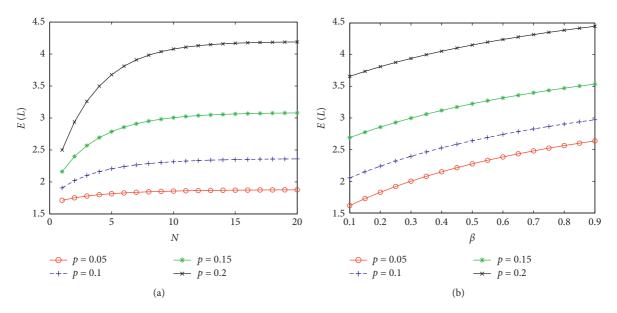


FIGURE 4: E[L] versus N and β for different values p (a) E[L]vs β (N = 0.2); (b) E[L]vs β (N = 6).

negative effect. Therefore, there is a great significance in designing the size of queue capacity. Next, we discuss the capacity setting of the underlying queueing model by numerical examples and analyze the importance of steady-state probabilities in designing the queue capacity. We assume that $N=4,\ \lambda=1.5,\ \mu=3,\ \xi_1=0.6,\ \xi_2=0.8,\ p=0.1,$ and $\beta=0.2.$

We first provide a table to show the steady-state probabilities of the queueing model, i.e., the probability that there are $n(n \ge 0)$ customers in the system. In Table 1, the expected number of customers in the system is E[L] = 2.1605. From Table 1, we find that n is greater than a certain number, $\pi_{\bf n}{\bf e} \longrightarrow 0$, so it is not necessary to design the system capacity to be infinite. According to E[L], we have

$$P\{L > E[L]\} = 1 - \sum_{n=0}^{\lfloor E[L] \rfloor} \pi_{n} \mathbf{e} = 1 - \pi_{0} \mathbf{e}_{1} - \sum_{n=1}^{2} \pi_{n} \mathbf{e} = 0.3094,$$

$$P\{L > E[L] + 1\} = 1 - \sum_{n=0}^{\lfloor E[L] + 1 \rfloor} \pi_{n} \mathbf{e} = 1 - \pi_{0} \mathbf{e}_{1} - \sum_{n=1}^{3} \pi_{n} \mathbf{e} = 0.2138.$$
(33)

n	n = 0	n = 1	n = 2	n = 3	n = 4	n = 5	<i>n</i> = 6
$\pi_{\mathbf{n}}\mathbf{e}$	0.33161	0.21614	0.14284	0.09564	0.06480	0.04437	0.03067
n	n = 7	n = 8	n = 9	n = 10	n = 11	n = 12	n = 13
$\pi_{\mathbf{n}}\mathbf{e}$	0.02137	0.01499	0.01059	0.00752	0.00537	0.00385	0.00278
n	n = 14	n = 15	n = 16	n = 17	n = 18	n = 19	n = 20
$\pi_{\mathbf{n}}\mathbf{e}$	0.00201	0.00146	0.00106	0.00078	0.00057	0.00042	0.00031
n	n = 21	n = 22	n = 23	n = 24	n = 25	n = 26	n = 27
$\pi_{\mathbf{n}}\mathbf{e}$	0.00023	0.00017	0.00012	0.00009	0.00007	0.00005	0.00004
n	n = 28	n = 29	n = 30	n = 31	n = 32	n = 33	n = 34
$\pi_{\mathbf{n}}\mathbf{e}$	0.00003	0.00002	0.00002	0.00001	0.00001	0.00001	0.00000

Table 1: Steady-state probabilities $\pi_n \mathbf{e}, n = 1, 2, \dots, 20$.

That is, if we use the mean value to design the queue size, the loss probability of arriving customers due to the finite capacity reaches 30.94%, even if we expand a unit of mean value, the loss probability is also up to 21.38%. Therefore, at this point, it is not advisable to design the queue size according to the expected number of customers in the system.

Next, we consider the designing the queue size by controlling the loss probability. For example, we assume that the loss probability is less than 0.0001. Then, from $P\{L > L_0\} < 0.0001$, we have $L_0 \ge 28$; that is, we can set the queue size as 28. Therefore, designing the queue capacity based on the steady-state probabilities is a more reasonable and rational method than the expected number of customers in the system.

5. Conclusion and Future Work

This paper is devoted to the study of the performance of the queueing system with service interruptions under a reactive maintenance policy combining the planned maintenance for the server. We dealt with the problem by constructing a queueing model. Using the matrix analytic approach and spectral expansion method, we respectively obtained the steady-state probabilities. Based on the theoretical framework given by the queueing analysis, we derived some performance measures and provided some numerical examples to show the impacts of system parameters on some key performance measures. Moreover, from the perspective of queueing model, we discussed the capacity setting of the underlying queueing model by numerical examples and analyzed the importance of steady-state probabilities in designing the queue capacity. We hope that our results can be applied to more practical queueing problems, and the results can provide important reference value for system administrator to scientifically design the maintenance policy.

Although in this paper, we just considered a single-server queueing system instead of a multiserver queueing system. In the further work, we could also consider the more realistic queueing system with multiple servers and use a Phase-type (PH) distribution for modelling the service times, which needs to develop new methods, and it will be an interesting direction for future research. Another interesting direction for future research is that, based on the steady-state analysis, we could investigate a real-world production system (service system) by using the real data and verify the validity of the

underling queueing model. Finally, we could also consider pricing problem in this system, i.e., we could investigate the strategic behavior of customers, and then obtain the optimal pricing decision on the basis of the equilibrium strategies of customers.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant no. 12001329), Shandong Provincial Natural Science Foundation (China) (Grant no. ZR2019BG014), and Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents (Grant no. 2019RCJJ016).

References

- [1] P. Wartenhorst, "N parallel queueing systems with server breakdown and repair," *European Journal of Operational Research*, vol. 82, no. 2, pp. 302–322, 1995.
- [2] W. J. Gray, P. P. Wang, and M. Scott, "A queueing model with multiple types of server breakdowns," *Quality Technology & Quantitative Management*, vol. 1, no. 2, pp. 245–255, 2004.
- [3] Y. Lam, Y. L. Zhang, and Q. Liu, "A geometric process model for M/M/1 queueing system with a repairable service station," *European Journal of Operational Research*, vol. 168, no. 1, pp. 100–121, 2006.
- [4] U. Yechiali, "Queues with system disasters and impatient customers when system is down," *Queueing Systems*, vol. 56, no. 3-4, pp. 195–202, 2007.
- [5] Q.-L. Li, Y. Ying, and Y. Q. Zhao, "A BMAP/G/1 retrial queue with a server subject to breakdowns and repairs," *Annals of Operations Research*, vol. 141, no. 1, pp. 233–270, 2006.
- [6] Z. Liu, J. Wu, and G. Yang, "An M/G/1 retrial G-queue with preemptive resume and feedback under N-policy subject to the server breakdowns and repairs," *Computers and Mathematics with Applications*, vol. 58, no. 9, pp. 1792–1807, 2009.

- [7] J. D. Cordeiro and J. P. Kharoufeh, "The UnreliableM/M/1 retrial queue in a random environment," *Stochastic Models*, vol. 28, no. 1, pp. 29–48, 2012.
- [8] M. Jain, P. Kumar, and R. K. Meena, "Fuzzy metrics and cost optimization of a fault-tolerant system with vacationing and unreliable server," *Journal of Ambient Intelligence and Hu*manized Computing, vol. 11, pp. 5755–5770, 2020.
- [9] S. I. Ammar, T. Jiang, and Q. Ye, "Transient analysis of impatient customers in an M/M/1 disasters queue in random environment," *Engineering Computations*, vol. 37, no. 6, p. 1945, 2020.
- [10] K. Kalidass and R. Kasturi, "A queue with working break-downs," Computers & Industrial Engineering, vol. 63, no. 4, pp. 779–783, 2012.
- [11] B. K. Kim and D. H. Lee, "The M/G/1 queue with disasters and working breakdowns," *Applied Mathematical Modelling*, vol. 38, no. 5-6, pp. 1788–1798, 2014.
- [12] Z. Liu and Y. Song, "The MX/M/1 queue with working breakdown," RAIRO - Operations Research, vol. 48, no. 3, pp. 399–413, 2014.
- [13] T. Jiang and L. Liu, "The GI/M/1 queue in a multi-phase service environment with disasters and working breakdowns," *International Journal of Computer Mathematics*, vol. 94, no. 4, pp. 707–726, 2017.
- [14] D.-Y. Yang and Y.-Y. Wu, "Analysis of a finite-capacity system with working breakdowns and retention of impatient customers," *Journal of Manufacturing Systems*, vol. 44, pp. 207–216, 2017.
- [15] Q. Ye and L. Liu, "Analysis of MAP/M/1 queue with working breakdowns," Communications in Statistics - Theory and Methods, vol. 47, no. 13, pp. 3073–3084, 2018.
- [16] T. Jiang and B. Xin, "Computational analysis of the queue with working breakdowns and delaying repair under a Bernoulli-schedule-controlled policy," *Communications in Sta*tistics - Theory and Methods, vol. 48, no. 4, pp. 926–941, 2019.
- [17] S. I. Ammar and P. Rajadurai, "Performance analysis of preemptive priority retrial queueing system with disaster under working breakdown services"" *Symmetry*, vol. 11, 2019.
- [18] W.-H. Zhou and G.-L. Zhu, "Economic design of integrated model of control chart and maintenance management," *Mathematical and Computer Modelling*, vol. 47, no. 11-12, pp. 1389–1395, 2008.
- [19] L. Liu, M. Yu, Y. Ma, and Y. Tu, "Economic and economicstatistical designs of an control chart for two-unit series systems with condition-based maintenance," *European Journal of Operational Research*, vol. 226, no. 3, pp. 491–499, 2013
- [20] W. Zhou, Z. Zheng, and W. Xie, "A control-chart-based queueing approach for service facility maintenance with energy-delay tradeoff," *European Journal of Operational Re*search, vol. 261, no. 2, pp. 613–625, 2017.
- [21] J. H. Cha, C. Sanguesa, and I. T. Castro, "Maintenance policy for a system with stochastically dependent failure modes with shock-accumulation effect," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1284–1297, 2016.
- [22] L. Yang, X. Ma, R. Peng, Q. Zhai, and Y. Zhao, "A preventive maintenance policy based on dependent two-stage deterioration and external shocks," *Reliability Engineering & System Safety*, vol. 160, pp. 201–211, 2017.
- [23] S. Alaswad and Y. Xiang, "A review on condition-based maintenance optimization models for stochastically deteriorating system," *Reliability Engineering & System Safety*, vol. 157, pp. 54–63, 2017.

- [24] T. Jiang, "Analysis of a clearing queueing system with server maintenance after N negative feedbacks," *Probability in the Engineering and Informational Sciences*, vol. 33, no. 2, pp. 220–240, 2019.
- [25] M. F. Neuts, Matrix-geometric Solutions in Stochastic Models: Algorithmic Approach, Johns Hopkins University Press, Baltimore, MA, USA, 1981.
- [26] I. Mitrani and R. Chakka, "Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method," *Performance Evaluation*, vol. 23, no. 3, pp. 241–260, 1995.