Probability in the Engineering and Informational Sciences, **33**, 2019, 220–240.
doi:10.1017/S026996481800013X

ANALYSIS OF A CLEARING QUEUEING SYSTEM WITH SERVER MAINTENANCE AFTER *N* NEGATIVE FEEDBACKS

TAO JIANG
College of Economics and Management,
Shandong University of Science and Technology,
Qingdao 266590,
People's Republic of China
E-mails: jtao0728@163.com

This paper is devoted to the study of a clearing queueing system with a special discipline. As soon as the server receives N negative feedbacks from customers, all present customers are forced to leave the system and the server undergoes a maintenance procedure. After an exponential maintenance time, the system resumes its service immediately. Using the matrix analytic method, we derive the steady-state distributions, which are then used for the computation of other performance measures. Furthermore, using first step analysis, we obtain the Laplace–Stieltjes transform of the sojourn time of an arbitrary customer. We also study the busy period of the system and derive the generating function of the total number of lost customers in a busy period. Finally, we investigate a long-run rate of cost and explore the optimal N value that minimizes the total cost per unit time. We also present some numerical examples to illustrate the impact of several model parameters to the performance measures.

Keywords: busy period, clearing queue, matrix analytic approach, negative feedback, sojourn time

1. INTRODUCTION

Queueing models with removals of customers before being served have been studied extensively due to their wide area of applications, such as complex modern communication systems, telecommunication networks, computer networks, and manufacturing systems. Since the introduction of this topic, there has been considerable attention to this theme. So far, the articles on this topic can be roughly divided into two categories, depending on whether the removal of customers (without receiving service) is the result of an internal factor (e.g., due to customer impatience) or if it is the result of an external factor (e.g., due to system's breakdowns). The category in which the removals of customers before being served are triggered by the customers themselves appears in queueing systems with reneging, where customers become impatient when their patience times expire and they leave the system without receiving service. The category of customers being removed from the

© Cambridge University Press 2018 0269-9648/18 $25.00

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

system, without receiving service, due to external factors, occurs in queueing systems that are subject to disasters/catastrophes/breakdowns. In this category, the occurrence of disasters forces all present customers to leave the system and the server abandons the service. Such systems are referred to as *stochastic clearing systems*.

Since the introduction of stochastic clearing systems, there is an ever growing interest in their performance analysis, see, for example, Stidham [18], Artalejo and Gómez-Corral [1] and Yang, Kim, and Chae [22], etc. Recently, some researchers have studied stochastic clearing systems in the framework of queueing systems subject to disasters/catastrophes. For example, Kim and Lee [13] have considered an M/G/1 queue with disastrous interruptions and working breakdowns, in which the system is equipped with a substitute server, who provides service during the repair period instead of stopping the service completely. Paz and Yechiali [16] studied an M/M/1 queue in a random environment with disasters and gave an explicit analysis of steady-state distribution by combining the probability generating function (PGF) approach and the matrix analytic approach. Jiang, Liu, and Li [11] investigated an M/G/1 queue in a multi-phase random environment with disasters, and derived the stationary queue length distribution as well as some important performance measures. Jiang and Liu [9,10] extended the model in Jiang et al. [11] to a GI/M/1 queue, in which the system operates in a multi-phase service environment and it may suffer from disastrous breakdowns in operative service environments. Mytalas and Zazanis [14] discussed a queueing model with batch Poisson arrival subject to disasters under a multiple adapted vacation policy, and provided a detailed analysis of the queueing system. Giorno, Nobile, and Spina [8] considered a non-stationary queueing system subject to catastrophes occurring according to some time-varying intensity. Sudhesh, Sebasthi Priya, and Lenin [20] analyzed an N-policy queue subject to disasters. Kapodistria, Phung-Duc, and Resing [12] investigated a birth/immigration-death processes under binomial catastrophes and gave an elaborate analysis on both the transient and equilibrium distribution of the population size. For some related papers on impatient customers and diasters, the interested readers are referred to Yechiali [23], Chakravarthy [4], Sudhesh [19] and Dimou and Economou [5], and the references therein. Lately, the analysis of such queueing systems with disasters/catastrophes focuses on an economic viewpoint, see, for example, Economou and Manou [7], Boudali and Economou [2,3]. In these papers, based on a natural reward-cost structure, the authors discuss the queueing models under the assumption that arriving customers choose whether to join the system or balk.

In all aforementioned papers, all customers are forced to leave the system due to the occurrence of disasters/catastrophes. However, in practice, we may encounter the following situation: all present customers are forced to abandon the system due to the server having accumulated a predetermined number of negative feedbacks. For example, in the context of a production system, it is quite custom to check the quality of the produced items. The production system is equipped with a quality control detector and monitored using the so called Shewhart charts, which are modeling undeniably well the process and can accurately identify out-of-control situations. In the setting of a Shewhart chart (see e.g., Zhou, Zheng, and Xie [24]), the quality characteristic measurements of the items are independent and identically distributed (i.i.d.) random variables with a normal distribution and an item is faulty if the quality characteristic falls outside the so-called 'in-control' region, meanwhile, as an out of control item is detected, the Shewhart chart will generate an alert signal. Hence, it is reasonable to consider p to be the probability that the quality characteristic of an item is outside the in-control region and $q = 1 - p$ to be the probability that the quality characteristic of an item is inside the in-control region. During the production process, once the number of faulty items reaches a predetermined threshold N, that is, the Shewhart chart records N alert signals, the machine of the production line experiences a maintenance

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

period immediately. So, the number of produced items before the machine experiencing a maintenance period follows a Pascal distribution with parameters (N, p). After being maintained, the machine resumes its work. Another example, in the context of a service system, at a service completion instant, a customer is asked for giving a feedback (evaluation) for his service. Specifically, in some well-known brand restaurant chain restaurants of China, when customers pay for their meal, they are asked for taking a few moments to complete out a customer response form about the service and the taste of food. If the customer is satisfied with the service and the taste of food, he will make a definite positive feedback, otherwise, he will give a negative feedback. When the restaurant receives a certain number of negative feedbacks N on the service and the taste of food, the restaurant will do some adjustments and take some steps (e.g., provide technical training for its chefs and service staffs or hire new staffs). After experiencing the period of adjustment, the restaurant resumes business and continues to provide service for arriving customers. In these two examples, we may encounter the following questions: what is the threshold value N (i.e., the number of faulty items appeared before the machine experiencing a maintenance period in the context of a production system and the number of negative feedbacks appeared before the restaurant experiencing an adjustment period in the context of a service system), how to determine the optimal value N to minimize the total cost. That indeed is true, take a production system as example, one is faced with the question of when to intervene (e.g., if the production line is maintained after one single item of an out of control measurement) then the cost may be too high, on the other end of the spectrum, if one waits for many out of control items then again the cost may be too high. This question motivate the research performed here. First, we construct a clearing queueing system with server maintenance after N negative feedbacks, where the produced items represent the served customers, and the out of control items (faulty items) represent the served customers who make negative feedback. Then, taking into account the cost associated with the accumulation of N negative feedbacks, the maintenance cost, as well as the cost for the customers that are lost, we establish a cost function and investigate the optimal number of negative feedbacks for minimizing the total cost.

The rest of this paper is organized as follows. Section 2 describes the proposed queueing model. In Section 3, using the matrix analytic method, we obtain the steady-state distributions. In Section 4, we derive some important performance measures of the system. Sections 5 and 6 are devoted to the derivation of the Laplace–Stieltjes transform (LST) of the sojourn time of an arbitrary customer, the LST of the duration of a busy period and the generating function of the total number of lost customers in a busy period. Numerical examples are conducted to find the optimal value N and check the impacts of system parameters on some key performance measures in Section 7. In Section 8, we conclude the paper and point to some interesting future directions for further investigation.

2. MODEL DESCRIPTION

In this paper, we consider a single-server clearing queueing system with server maintenance after N negative (dissatisfied) feedbacks. The queueing model is described in detail below.

Customers arrive according to a Poisson process with rate λ.

At a service completion instant, the server will receive a positive feedback if the customer is satisfied with the quality of service, otherwise the server will receive a negative feedback if the customer is not satisfied with the quality of service. To this end, we consider p as the probability that the server receives a negative feedback after serving a customer and consider $q = 1 - p$ as the probability that the server receives a positive feedback after serving

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

a customer. Once the number of negative feedbacks reaches a predetermined number N , all present customers are forced to leave the system, and the server undergoes a maintenance process immediately.

Service times and maintenance times are assumed to be exponentially distributed with parameters μ and ξ, respectively. After being maintained, the server resumes the normal service immediately.

We further assume that customers are served in accordance to the first-come, first-served (FCFS) discipline. The arrival process, the service times and the maintenance times are mutually independent.

3. STATIONARY QUEUE LENGTH DISTRIBUTION

In this section, we use the matrix analytic approach to analyze the queueing model. It is worth noting that as long as $0 < p < 1$ and $\xi > 0$, the number of customers in the system never goes to infinity, then the system in consideration can be analyzed in steady state. Then, at an arbitrary time t, the system can be described by a stochastic process $\{(L(t), J(t)), t \geq 0\}$, where $L(t)$ denotes the number of customers present in the system at time t, $J(t) = -1$ if the server being maintained at time t, $J(t) = i, i = 0, 1, 2, \ldots, N-1$ if the server has accumulated i negative feedbacks before undergoing a maintenance period at time t. Obviously, $\{(L(t), J(t)), t \geq 0\}$ is a continuous-time Markov chain with state space

$$\Omega = \{(n, -1), n \geq 0\} \cup \{(n, j), n \geq 0, j = 0, 1, 2, \ldots, N-1\}.$$

Next, we consider the continuous-time Markov chain, with generator matrix is given by

$$Q = \begin{bmatrix} C & A_0 & 0 & 0 & \cdots \\ B + A_2 & A_1 & A_0 & 0 & \cdots \\ B & A_2 & A_1 & A_0 & \cdots \\ B & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix},$$

where all block matrices are square ones with $(N+1) \times (N+1)$ orders and

$$A_1 = \begin{bmatrix} -(\lambda+\xi) & \xi & & & \\ & -(\lambda+\mu) & & & \\ & & \ddots & & \\ & & & & -(\lambda+\mu) \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \mu q & \mu p & \cdots & 0 \\ 0 & 0 & \mu q & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \mu q \end{bmatrix},$$

$$A_0 = \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu p & 0 & \cdots & 0 \end{bmatrix}, \quad C = \begin{bmatrix} -(\lambda+\xi) & \xi & & \\ & -\lambda & & \\ & & \ddots & \\ & & & -\lambda \end{bmatrix}.$$

Note that Q is a special case of GI/M/1 type. For more details about this special structure Markov chain, the interested readers are referred to Paz and Yechiali [16], Phung-Duc [17] and Van Houdt and van Leeuwaarden [21]. According to Neuts [15], in order

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

to analyze the Markov chain effectively, an important matrix is rate matrix R, which is obtained as the minimal non-negative solution to the matrix-quadratic equation

$$A_0 + RA_1 + R^2 A_2 = 0. \tag{1}$$

Next, under the stability condition $0 < p < 1$ and $\xi > 0$, we will derive the steady-state distributions of the system. First, let (L, J) be the stationary limit of the Markov chain $(L(t), J(t))$. Denote the steady-state probability vector of Q by $P = (P_0, P_1, P_2, P_3, \ldots)$, with

$$P_m = (P_{m,-1}, P_{m,0}, P_{m,1}, \ldots, P_{m,N-1}), \quad m \geq 0,$$

where

$$P_{m,-1} = \lim_{t\to\infty} \Pr(L(t) = m, J(t) = -1), \quad m \geq 0,$$
$$P_{m,j} = \lim_{t\to\infty} \Pr(L(t) = m, J(t) = j), \quad m \geq 0, \qquad j = 0, 1, \ldots N-1.$$

Then, from the results in Neuts [15], we know that the steady-state distributions satisfy the following set of equations:

$$P_m = P_0 R^m, \quad m \geq 0,$$

$$P_0 C + P_0 R(B + A_2) + \left(\sum_{m=2}^{\infty} P_m\right) B = \mathbf{0},$$

$$\sum_{m=0}^{\infty} P_m \mathbf{e} = P_0 \left(\sum_{m=0}^{\infty} R^m\right) \mathbf{e} = 1.$$

Finally, we can obtain the steady-state distributions as follows.

Proposition 1: *If $0 < p < 1$ and $\xi > 0$, the steady-state probability vector P_0 satisfies the following set of equations:*

$$P_0[C + R(B + A_2) + R^2(I - R)^{-1}B] = \mathbf{0}, \tag{2}$$
$$P_0(I - R)^{-1}\mathbf{e} = 1, \tag{3}$$

and P_m satisfies

$$P_m = P_0 R^m, \quad m \geq 1, \tag{4}$$

where I is the identity matrix, $\mathbf{0}$ is a row vector with $N+1$ dimensions and all elements equal to zero, and $\mathbf{e}$ is a column vector with $N+1$ dimensions and all elements equal to one, the elements of R satisfy

$$r_{0,0} = \frac{\lambda}{\lambda + \xi}, \quad r_{0,1} = \frac{\xi r_{0,0}}{(\lambda + \mu) - \mu q(r_{0,0} + r_{1,1})},$$

$$r_{i,i} = \frac{(\lambda + \mu) - \sqrt{(\lambda + \mu)^2 - 4\lambda\mu q}}{2\mu q}, \quad i = 1, 2, \ldots, N,$$

$$r_{i,j} = \frac{\mu p \sum_{k=i}^{j-1} r_{i,k} r_{k,j-1} + \mu q \sum_{k=i+1}^{j-1} r_{i,k} r_{k,j-1}}{(\lambda + \mu) - \mu q(r_{i,i} + r_{j,j})}, \quad i < j, \quad i = 0, 1, \ldots, N, \quad j = 2, \ldots, N,$$

with $\sum_{k=i}^{i-1}(\cdot) = 0$.

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

PROOF: In the proofs, we mainly give the procedure to obtain the rate matrix R. First, due to the special upper triangular of matrices $A_k, k = 0, 1, 2$, it follows from Eq. (1) that R is also an upper triangular matrix having the following form:

$$R = \begin{bmatrix} r_{0,0} & r_{0,1} & \cdots & r_{0,N} \\ 0 & r_{1,1} & \cdots & r_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{N,N} \end{bmatrix}.$$

In the left-hand side of Eq. (1), we have

$$R^2 A_2 = \begin{bmatrix} 0 & \beta_{0,1} & \beta_{0,2} & \cdots & \beta_{0,N} \\ & \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,N} \\ & & \beta_{2,2} & \cdots & \beta_{2,N} \\ & & & \ddots & \vdots \\ & & & & \beta_{N,N} \end{bmatrix},$$

$$RA_1 = \begin{bmatrix} -(\lambda+\xi)r_{0,0} & \xi r_{0,0} - (\lambda+\mu)r_{0,1} & \cdots & -(\lambda+\mu)r_{0,N} \\ 0 & -(\lambda+\mu)r_{1,1} & \cdots & -(\lambda+\mu)r_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -(\lambda+\mu)r_{N.N} \end{bmatrix},$$

with

$$\begin{aligned} &\beta_{i,i} = \alpha_{i,i}\mu q, \quad i = 1, \ldots, N, \\ &\beta_{0,1} = \mu q \alpha_{0,1}, \\ &\beta_{i,j} = \alpha_{i,j-1}\mu p + \alpha_{i,j}\mu q, \quad i < j, \quad i = 0, 1, \ldots, N-1, \quad j = 2, \ldots, N, \end{aligned}$$

where

$$\begin{aligned} &\alpha_{i,i} = r_{i,i}^2, \quad i = 0, 1, \ldots, N, \\ &\alpha_{i,j} = r_{i,i}r_{i,j} + r_{i,i+1}r_{i+1,j} + \cdots + r_{i,j}r_{j,j}, \quad i < j, \quad i = 0, 1, \ldots, N-1, \quad j = 1, 2, \ldots, N. \end{aligned}$$

Then, substituting R^2A_2, RA_1 and A_0 into Eq. (1), we can obtain the following set of equations:

$$\lambda - (\lambda + \xi)r_{0,0} = 0, \tag{5}$$

$$\lambda - (\lambda + \mu)r_{i,i} + \beta_{i,i} = 0, \quad i = 1, 2, \ldots, N, \tag{6}$$

$$\beta_{0,1} + \xi r_{0,0} - (\lambda + \mu)r_{0,1} = 0, \tag{7}$$

$$\beta_{i,j} - (\lambda + \mu)r_{i,j} = 0, i < j, \quad i = 0, 1, \ldots, N, \quad j = 2, \ldots, N. \tag{8}$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

Solving the set of equations, we have

$$r_{0,0} = \frac{\lambda}{\lambda+\xi}, \quad r_{0,1} = \frac{\xi r_{0,0}}{(\lambda+\mu)-\mu q(r_{0,0}+r_{1,1})},$$

$$r_{i,i} = \frac{(\lambda+\mu)-\sqrt{(\lambda+\mu)^2-4\lambda\mu q}}{2\mu q}, \quad i=1,2,\ldots,N,$$

$$r_{i,j} = \frac{\mu p\sum_{k=i}^{j-1} r_{i,k}r_{k,j-1} + \mu q\sum_{k=i+1}^{j-1} r_{i,k}r_{k,j-1}}{(\lambda+\mu)-\mu q(r_{i,i}+r_{j,j})}, \quad i<j, \quad i=0,1,\ldots,N, \quad j=2,\ldots,N,$$

where $\sum_{k=i}^{i-1}(\cdot)=0$. ■

Actually, from these recursive formulae, we can recursively calculate the elements of R from the main diagonal part and then above the main diagonal parts in the following order.

We first compute the elements on the first diagonal above the main diagonal. More precisely, let $j=i+1$ in Eq. (8), we have

$$r_{i,i+1} = \frac{\mu p r_{i,i}^2}{(\lambda+\mu)-\mu q(r_{i,i}+r_{i+1,i+1})}, \quad i=1,\ldots,N-1.$$

We continue to calculate the entries on the next diagonal, that is, let $j=i+2$ in Eq. (8), we have

$$r_{i,i+2} = \frac{\mu p r_{i,i+1}(r_{i,i}+r_{i+1,i+1}) + \mu q r_{i,i+1}r_{i+1,i+1}}{(\lambda+\mu)-\mu q(r_{i,i}+r_{i+2,i+2})}, \quad i=0,1,\ldots,N-2.$$

Using the same recursion with $j=i+3,\ldots,N$, we can compute all entries of R. In general, the $(i,i+l)$-th element of R is given by

$$r_{i,i+l} = \frac{\mu p\sum_{k=i}^{i+l-1} r_{i,k}r_{k,i+l-1} + \mu q\sum_{k=i+1}^{i+l-1} r_{i,k}r_{k,i+l-1}}{(\lambda+\mu)-\mu q(r_{i,i}+r_{i+l,i+l})},$$
$$i=0,1,\ldots,N-l, \quad l=3,\ldots,N-i.$$

4. PERFORMANCE MEASURES

In this section, we give some important performance measures of the system. The expected number of customers in the system can be obtained by

$$E[L] = \sum_{n=0}^{\infty} nP_n\mathbf{e} = \sum_{n=1}^{\infty} nP_0R^n\mathbf{e} = P_0R(I-R)^{-2}\mathbf{e}.$$

The probability that the server is in maintenance period (the percentage of the time spent in maintenance period) is given by $P_{\mathrm{mp}} = \sum_{n=0}^{\infty} P_{n,-1}$.

The server is available with probability $P_{\mathrm{a}} = \sum_{n=0}^{\infty}\sum_{k=0}^{N-1} P_{n,k}$.

The system is empty with probability $P_{\mathrm{e}} = \sum_{k=0}^{N-1} P_{0,k} + P_{0,-1}$.

Next, we consider the number of customers being served in a cycle. To avoid confusion, a cycle is defined as the time interval between two consecutive instants at which a maintenance

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

period commences. Let C denote the duration of a cycle, L_s denote the number of customers being served in C. Then, we can easily obtain the distribution and the mean value of L_s

$$P(L_s = n) = \binom{n-1}{N-1} p^N q^{n-N}, \quad n \geq N,$$

$$E[L_s] = Np^N + \sum_{n=N+1}^{\infty} n\binom{n-1}{N-1} p^N q^{n-N}.$$

From the distribution of L_s, we observe that L_s follows a Pascal distribution with parameters (N, p), that is, $L_s \sim Pa(N, p)$. In other words, $L_s - N$ has the negative binomial distribution with parameters (N, p), that is, $L_s - N \sim NB(N, p)$. Therefore, the mean value of L_s can be simplified as $E[L_s] = N/p$. Take the production system as example, we find that the value of N is connected with the number of runs used in the Shewhart chart. Let L_s denote the number of produced items in a production cycle and N denote the number of faulty items in a production cycle, then the long-run proportion of faulty items during a production cycle N/L_s is a constant value p.

Let B denote the duration of time that the server is active (working) in C. Then its LST $B^*(s)$ and mean value $E[B]$ can be obtained by

$$B^*(s) = p^N \left(\frac{\mu}{\mu + s}\right)^N + \sum_{n=N+1}^{\infty} \binom{n-1}{N-1} p^N q^{n-N} \left(\frac{\mu}{\mu + s}\right)^n,$$

$$E[B] = \frac{1}{\mu}\left[Np^N + \sum_{n=N+1}^{\infty} n\binom{n-1}{N-1} p^N q^{n-N}\right] = \frac{N}{\mu p}.$$

5. SOJOURN TIMES

In this section, we turn our attention to the sojourn time of an arbitrary customer, which is defined to be the overall time from the arrival till the departure from the system, due to either a service completion or the occurrence of server maintenance. Let W and $W^*(s)$ respectively denote the stationary sojourn time of an arbitrary customer and its corresponding LST. In order to compute $W^*(s)$, we define $W_{n,k}$ as the conditional sojourn time of a customer given that he is the nth older in the system and that the system is in state $(n, k), n \geq 1, k = -1, 0, 1, \ldots, N-1$. Then, $W^*(s)$ can be obtained by the following proposition.

PROPOSITION 2: *If $0 < p < 1$ and $\xi > 0$, the LST of the sojourn time of an arbitrary customer $W^*(s)$ can be obtained by*

$$W^*(s) = \sum_{k=0}^{N-1} \sum_{n=1}^{\infty} P_{n-1,k} W^*_{n,k}(s) + \sum_{n=1}^{\infty} P_{n-1,-1} W^*_{n,-1}(s), \tag{9}$$

where

$$W^*_{n,k}(s) = \left(\frac{\mu}{\mu + s}\right)^n, \quad 1 \leq n \leq N-k, \quad 0 \leq k \leq N-1, \tag{10}$$

$$W^*_{n,N-1}(s) = \frac{\mu p[(\mu+s)^{n-1} - (\mu q)^{n-1}]}{(\mu+s)^{n-1}(\mu p + s)} + \frac{\mu^n q^{n-1}}{(\mu+s)^n}, \quad n \geq 1, \tag{11}$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

$$W^*_{n,k}(s) = \frac{\mu q}{\mu+s} W^*_{n-1,k}(s) + \frac{\mu p}{\mu+s} W^*_{n-1,k+1}(s), \quad 0 \le k \le N-2, \quad n \ge N-k, \tag{12}$$

$$W^*_{n,-1}(s) = \frac{\xi}{\xi+s} W^*_{n,0}(s), \quad n \ge 1. \tag{13}$$

PROOF: First, it is easy to find that only if the server accumulates N negative feedbacks, the system undergoes a maintenance period, causing all present customers to leave the system, so we can easily have

$$W^*_{n,k}(s) = \left(\frac{\mu}{\mu+s}\right)^n, \quad 1 \le n \le N-k, \quad 0 \le k \le N-1,$$

which is equation (10).

Next, we consider the case for $n \ge 1, k = N-1$. By conditioning on the next future event (first-step analysis) and using the strong Markov property, we can obtain

$$W^*_{n,N-1}(s) = \frac{\lambda}{\lambda+\mu+s} W^*_{n,N-1}(s) + \frac{\mu q}{\lambda+\mu+s} W^*_{n-1,N-1}(s) + \frac{\mu p}{\lambda+\mu+s}, \quad n \ge 1,$$

where $W^*_{0,N-1}(s) = 1$. Solving for $W^*_{n,N-1}(s)$ recursively, we have

$$\begin{aligned} W^*_{n,N-1}(s) &= \frac{\mu q}{\mu+s} W^*_{n-1,N-1}(s) + \frac{\mu p}{\mu+s} \\ &= \frac{\mu p}{\mu+s} + \left(\frac{\mu}{\mu+s}\right)^2 qp + \cdots + \left(\frac{\mu}{\mu+s}\right)^{n-1} q^{n-2}p + \left(\frac{\mu}{\mu+s}\right)^n q^{n-1} \\ &= \frac{\mu p[(\mu+s)^{n-1} - (\mu q)^{n-1}]}{(\mu+s)^{n-1}(\mu p+s)} + \frac{\mu^n q^{n-1}}{(\mu+s)^n}, \quad n \ge 1, \end{aligned}$$

which is Eq. (11). Consider the case for $k = N-2, n \ge 2$. Analogously, we have

$$W^*_{n,N-2}(s) = \frac{\lambda}{\lambda+\mu+s} W^*_{n,N-2}(s) + \frac{\mu q}{\lambda+\mu+s} W^*_{n-1,N-2}(s) + \frac{\mu p}{\lambda+\mu+s} W^*_{n-1,N-1}(s),$$

which leads to

$$W^*_{n,N-2}(s) = \frac{\mu q}{\mu+s} W^*_{n-1,N-2}(s) + \frac{\mu p}{\mu+s} W^*_{n-1,N-1}(s), \quad n \ge 2.$$

Substituting Eq. (11) into above equation, and recursively calculating it from $W^*_{1,N-2}(s) = W^*_{1,N-1}(s) = \mu/(\mu+s)$, then, we can obtain the expression $W^*_{n,N-2}(s)$.

Generally, for $0 \le k \le N-3$ and $n \ge N-k$, we have

$$W^*_{n,k}(s) = \frac{\lambda}{\lambda+\mu+s} W^*_{n,k}(s) + \frac{\mu q}{\lambda+\mu+s} W^*_{n-1,k}(s) + \frac{\mu p}{\lambda+\mu+s} W^*_{n-1,k+1}(s),$$

that is,

$$W^*_{n,k}(s) = \frac{\mu q}{\mu+s} W^*_{n-1,k}(s) + \frac{\mu p}{\mu+s} W^*_{n-1,k+1}(s),$$

which is Eq. (12). So, from the expression $W^*_{n,N-1}(s)$ in Eq. (11), we can recursively calculate the closed-form formula of $W^*_{n,k}(s)$ for the cases $0 \le k \le N-2, n \ge N-k$.

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

Finally, we consider that a customer is the nth older in the system and that the system is in state $(n,-1)$, $n \geq 1$. Similar to aforementioned analysis, we have

$$W^*_{n,-1}(s) = \frac{\lambda}{\lambda+\xi+s} W^*_{n,-1}(s) + \frac{\xi}{\lambda+\xi+s} W^*_{n,0}(s), \quad n \geq 1.$$

Simplifying $W^*_{n,-1}(s)$, we have

$$W^*_{n,-1}(s) = \frac{\xi}{\xi+s} W^*_{n,0}(s), \quad n \geq 1,$$

which is Eq. (13). Once $W^*_{n,-1}(s)$ and $W^*_{n,k}(s)$ are obtained, $W^*(s)$ can be derived by conditioning on the state of the system at an arrival epoch and using PASTA property

$$W^*(s) = \sum_{k=0}^{N-1} \sum_{n=1}^{\infty} P_{n-1,k} W^*_{n,k}(s) + \sum_{n=1}^{\infty} P_{n-1,-1} W^*_{n,-1}(s).$$

■

6. BUSY PERIOD

In this section, similarly to the method used in Dimou and Economou [5] and Dimou, Economou, and Fakinos [6], we study the busy period of the system, which is defined as the time from the epoch at which an arriving customer finds the system empty to the next epoch that the system is empty again (due either the server having accumulated N negative feedbacks or due to the server having served all present customers). First, we derive the LST $T^*(s)$ of the duration T of a busy period. Without loss of generality, define $T_{n,-1}$ as a first passage time from state $(n,-1)$ to one of the states in set $\{(0,-1),(0,k), k = 0,1,\ldots,N-1\}$, define $T_{n,j}$ as a first passage time from state (n,j) to one of the states in set $\{(0,-1),(0,k), j \leq k \leq N-1, 0 \leq j \leq N-1\}$. Then, the LST $T^*(s)$ of the duration of a busy period can be obtained by the following proposition.

PROPOSITION 3: *If $0 < p < 1$ and $\xi > 0$, the LST $T^*(s)$ of the duration of a busy period can be obtained by*

$$T^*(s) = \pi_{0,-1} T^*_{1,-1}(s) + \sum_{k=0}^{N-1} \pi_{0,k} T^*_{1,k}(s),$$

with

$$\pi_{0,-1} = \frac{P_{0,-1}}{P_{0,-1} + \sum_{j=0}^{N-1} P_{0,j}}, \quad \pi_{0,k} = \frac{P_{0,k}}{P_{0,-1} + \sum_{j=0}^{N-1} P_{0,j}}, \quad 0 \leq k \leq N-1,$$

$$T^*_{1,-1}(s) = \frac{\xi[R_0(s,a(s))-1]}{\lambda},$$

$$T^*_{1,N-1}(s) = \frac{z_1(s)[\mu - \mu q z_1(s)]}{\lambda(1-z_1(s))},$$

$$T^*_{1,k}(s) = \frac{z_1(s)(\mu q + \mu p R_{k+1}(s, z_1(s)))}{\lambda}, \quad 0 \leq k \leq N-2,$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

where

$$R_{N-1}(s,z)=\frac{-\lambda z T^*_{1,N-1}(s)+\mu q z^2+\frac{\mu p z^2}{1-z}+f_1(s,z)}{f_1(s,z)},$$

$$R_k(s,z)=\frac{-\lambda z T^*_{1,k}(s)+\mu q z^2+\mu p z^2 R_{k+1}(s,z)+f_1(s,z)}{f_1(s,z)},\quad 0\le k\le N-2,$$

$$f_1(s,z)=-\mu q z^2+(\lambda+\mu+s)z-\lambda,$$

$$z_1(s)=\frac{(\lambda+\mu+s)-\sqrt{(\lambda+\mu+s)^2-4\lambda\mu q}}{2\mu q},\quad a(s)=\frac{\lambda}{\lambda+\xi+s}.$$

PROOF: For obtaining the LST $T^*(s)$ of the duration of a busy period, we employ first-step analysis. First, we have

$$\begin{cases}
T^*_{n,N-1}(s)=\dfrac{\lambda}{\lambda+\mu+s}T^*_{n+1,N-1}(s)+\dfrac{\mu q}{\lambda+\mu+s}T^*_{n-1,N-1}(s)+\dfrac{\mu p}{\lambda+\mu+s},n\ge 1,\\
T^*_{n,N-2}(s)=\dfrac{\lambda}{\lambda+\mu+s}T^*_{n+1,N-2}(s)\\
\qquad\qquad+\dfrac{\mu q}{\lambda+\mu+s}T^*_{n-1,N-2}(s)+\dfrac{\mu p}{\lambda+\mu+s}T^*_{n-1,N-1}(s),n\ge 1,\\
\qquad\vdots\\
T^*_{n,0}(s)=\dfrac{\lambda}{\lambda+\mu+s}T^*_{n+1,0}(s)+\dfrac{\mu q}{\lambda+\mu+s}T^*_{n-1,0}(s)+\dfrac{\mu p}{\lambda+\mu+s}T^*_{n-1,1}(s),n\ge 1,\\
T^*_{n,-1}(s)=\dfrac{\lambda}{\lambda+\xi+s}T^*_{n+1,-1}(s)+\dfrac{\xi}{\lambda+\xi+s}T^*_{n,0}(s),n\ge 1,
\end{cases}\tag{14}$$

where $T^*_{0,-1}(s)=1, T^*_{0,k}(s)=1, k=0,1,2,\ldots,N-1$. To find an explicit solution of the recursive schemes, we define the mixed transforms

$$R_{-1}(s,z)=\sum_{n=0}^{\infty}T^*_{n,-1}(s)z^n,\quad R_k(s,z)=\sum_{n=0}^{\infty}T^*_{n,k}(s)z^n,$$
$$k=0,1,\ldots,N-1,\quad |z|<1,\quad s\ge 0.$$

Multiplying Eqs. (14) by z^n and summing for $n\ge 1$, we have

$$\begin{cases}
f_1(s,z)[R_{N-1}(s,z)-1]=-\lambda z T^*_{1,N-1}(s)+\mu q z^2+\dfrac{\mu p z^2}{1-z},\\
f_1(s,z)[R_{N-2}(s,z)-1]=-\lambda z T^*_{1,N-2}(s)+\mu q z^2+\mu p z^2 R_{N-1}(s,z),\\
\qquad\vdots\\
f_1(s,z)[R_0(s,z)-1]=-\lambda z T^*_{1,0}(s)+\mu q z^2+\mu p z^2 R_1(s,z),\\
f_2(s,z)[R_{-1}(s,z)-1]=-\lambda z T^*_{1,-1}(s)+\xi z(R_0(s,z)-1),
\end{cases}\tag{15}$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

where $f_1(s,z) = -\mu q z^2 + (\lambda+\mu+s)z - \lambda$, $f_2(s,z) = (\lambda+\xi+s)z - \lambda$. We note that $f_1(s,z) = 0$ has two distinct roots

$$z_1(s) = \frac{(\lambda+\mu+s) - \sqrt{(\lambda+\mu+s)^2 - 4\lambda\mu q}}{2\mu q}$$

lies in $(0,1)$ and

$$z_2(s) = \frac{(\lambda+\mu+s) + \sqrt{(\lambda+\mu+s)^2 - 4\lambda\mu q}}{2\mu q}$$

lies in $(1,\infty)$. However, $f_1(s, z_1(s)) = 0$ and $R_k(s,z) < \infty$, therefore, substituting $z_1(s)$ into the first equation of (15), we have

$$T^*_{1,N-1}(s) = \frac{z_1(s)[\mu - \mu q z_1(s)]}{\lambda(1 - z_1(s))},$$

and

$$R_{N-1}(s,z) = \frac{-\lambda z T^*_{1,N-1}(s) + \mu q z^2 + \frac{\mu p z^2}{1-z} + f_1(s,z)}{f_1(s,z)}.$$

From the second equation of (15), we obtain

$$T^*_{1,N-2}(s) = \frac{z_1(s)[\mu q + \mu p R_{N-1}(s, z_1(s))]}{\lambda},$$

and

$$R_{N-2}(s,z) = \frac{-\lambda z T^*_{1,N-2}(s) + \mu q z^2 + \mu p z^2 R_{N-1}(s,z) + f_1(s,z)}{f_1(s,z)}.$$

Repeating this process, we can determine

$$T^*_{1,k}(s) = \frac{z_1(s)(\mu q + \mu p R_{k+1}(s, z_1(s)))}{\lambda}, \quad 0 \le k \le N-2,$$

and

$$R_k(s,z) = \frac{-\lambda z T^*_{1,k}(s) + \mu q z^2 + \mu p z^2 R_{k+1}(s,z) + f_1(s,z)}{f_1(s,z)}, \quad 0 \le k \le N-2.$$

Next, for any fixed $s \ge 0$, $a(s) = (\lambda/(\lambda+\xi+s))$ is the unique solution of $f_2(s,z) = 0$ in the open unit disk $\{z : |z| < 1\}$. Substituting $a(s)$ and $R_0(s,z) - 1$ into the last equation of (15), using $f_2(s, a(s)) = 0$ and the convergence of $R_{-1}(s,z)$ in the open unit disk $\{z : |z| < 1\}$, we have

$$T^*_{1,-1}(s) = \frac{\xi[R_0(s, a(s)) - 1]}{\lambda}.$$

Finally, we show the conditional probability that a new arrival finds the system in state $(0,-1)$ and $(0,k)$, given that the arriving customer finds no customers in the system at the

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

arriving epoch. It is easy to derive that

$$\pi_{0,-1} = \frac{P_{0,-1}}{P_{0,-1} + \sum_{j=0}^{N-1} P_{0,j}}, \quad \pi_{0,k} = \frac{P_{0,k}}{P_{0,-1} + \sum_{j=0}^{N-1} P_{0,j}}, \quad 0 \le k \le N-1.$$

Then, the LST $T^*(s)$ of the duration of a busy period is

$$T^*(s) = \pi_{0,-1} T^*_{1,-1}(s) + \sum_{k=0}^{N-1} \pi_{0,k} T^*_{1,k}(s).$$

■

A similar method can be used to compute the total number of lost customers Q during a busy period. Define $Q_{n,-1}$ as the total number of customers that abandon the system without receiving service during a first passage time from state $(n,-1)$ to one of the state in set $\{(0,-1),(0,k),k=0,1,\ldots,N-1\}$, define $Q_{n,j}$ as the total number of customers that leave the system without receiving service during a first passage time from state (n,j) to one of the state in set $\{(0,-1),(0,k),j \le k \le N-1, 0 \le j \le N-1\}$. The PGFs corresponding to Q, $Q_{n,-1}$ and $Q_{n,j}$ are $Q(z)$, $Q_{n,-1}(z)$, and $Q_{n,j}(z)$, respectively. Then, The PGF of the total number of lost customers during a busy period can be obtained by the following proposition.

PROPOSITION 4: *If $0 < p < 1$ and $\xi > 0$, the PGF of the total number of lost customers during a busy period $Q(z)$ can be obtained by*

$$Q(z) = \pi_{0,-1} Q_{1,-1}(z) + \sum_{k=0}^{N-1} \pi_{0,k} Q_{1,k}(z),$$

with

$$Q_{1,-1}(z) = \frac{\xi(M_0(z,b)-1)}{\lambda},$$

$$Q_{1,N-1}(z) = \frac{x_1(\mu q(1-zx_1)+\mu p)}{\lambda(1-zx_1)},$$

$$Q_{1,k}(z) = \frac{x_1(\mu q + \mu p M_{k+1}(z,x_1))}{\lambda}, \quad 0 \le k \le N-2,$$

where

$$M_{N-1}(z,x) = \frac{-\lambda x Q_{1,N-1}(z) + \mu q x^2 + \frac{\mu p x^2}{1-zx} + g_1(x)}{g_1(x)},$$

$$M_k(z,x) = \frac{-\lambda x Q_{1,k}(z) + \mu q x^2 + \mu p x^2 M_{k+1}(z,x) + g_1(x)}{g_1(x)}, \quad 0 \le k \le N-2,$$

$$g_1(x) = -\mu q x^2 + (\lambda+\mu)x - \lambda,$$

$$x_1 = \frac{(\lambda+\mu) - \sqrt{(\lambda+\mu)^2 - 4\lambda\mu q}}{2\mu q}, \quad b = \frac{\lambda}{\lambda+\xi}.$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

PROOF: First, from first-step analysis, we have

$$\begin{cases} Q_{n,N-1}(z) = \dfrac{\lambda}{\lambda+\mu}Q_{n+1,N-1}(z) + \dfrac{\mu q}{\lambda+\mu}Q_{n-1,N-1}(z) + \dfrac{\mu p}{\lambda+\mu}z^{n-1}, \quad n \geq 1, \\ Q_{n,N-2}(z) = \dfrac{\lambda}{\lambda+\mu}Q_{n+1,N-2}(z) + \dfrac{\mu q}{\lambda+\mu}Q_{n-1,N-2}(z) + \dfrac{\mu p}{\lambda+\mu}Q_{n-1,N-1}(z), \quad n \geq 1, \\ \qquad \vdots \\ Q_{n,0}(z) = \dfrac{\lambda}{\lambda+\mu}Q_{n+1,0}(z) + \dfrac{\mu q}{\lambda+\mu}Q_{n-1,N-2}(z) + \dfrac{\mu p}{\lambda+\mu}Q_{n-1,1}(z), \quad n \geq 1, \\ Q_{n,-1}(z) = \dfrac{\lambda}{\lambda+\xi}Q_{n+1,-1}(z) + \dfrac{\xi}{\lambda+\xi}Q_{n,0}(z), \quad n \geq 1, \end{cases} \tag{16}$$

where $Q_{0,-1}(z) = 1, Q_{0,k}(z) = 1, k = 1, 2, \ldots, N-1$. Similarly, define the following generating functions:

$$M_{-1}(z,x) = \sum_{n=0}^{\infty} Q_{n,-1}(z)x^n, M_k(z,x) = \sum_{n=0}^{\infty} Q_{n,k}(z)x^n, k = 0, 1, \ldots, N-1,$$
$$|z| < 1, \quad s \geq 0.$$

Multiplying Eqs. (16) by x^n and summing for $n \geq 1$, we have

$$\begin{cases} g_1(x)[M_{N-1}(z,x) - 1] = -\lambda x Q_{1,N-1}(z) + \mu q x^2 + \dfrac{\mu p x^2}{1-zx}, \\ g_1(x)[M_{N-2}(z,x) - 1] = -\lambda x Q_{1,N-2}(z) + \mu q x^2 + \mu p x^2 M_{N-1}(z,x), \\ \qquad \vdots \\ g_1(x)[M_0(z,x) - 1] = -\lambda x Q_{1,0}(z) + \mu q x^2 + \mu p x^2 M_1(z,x), \\ g_2(x)[M_{-1}(z,x) - 1] = -\lambda x Q_{1,-1}(z) + \xi x(M_0(z,x) - 1), \end{cases} \tag{17}$$

where $g_1(x) = -\mu q x^2 + (\lambda+\mu)x - \lambda, g_2(x) = (\lambda+\xi)x - \lambda$. From $g_1(x) = 0$, the two distinct roots

$$x_1 = \frac{(\lambda+\mu) - \sqrt{(\lambda+\mu)^2 - 4\lambda\mu q}}{2\mu q}$$

lies in $(0,1)$ and

$$x_2 = \frac{(\lambda+\mu) + \sqrt{(\lambda+\mu)^2 - 4\lambda\mu q}}{2\mu q}$$

lies in $(1,\infty)$. From $g_2(x) = 0$, the unique root $b = \frac{\lambda}{\lambda+\xi}$ lies in $(0,1)$. Then, from the first equation of (17), we have

$$Q_{1,N-1}(z) = \frac{x_1(\mu q(1-zx_1) + \mu p)}{\lambda(1-zx_1)},$$

and

$$M_{N-1}(z,x) = \frac{-\lambda x Q_{1,N-1}(z) + \mu q x^2 + \frac{\mu p x^2}{1-zx} + g_1(x)}{g_1(x)}.$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

Continue to do the process recursively, we can obtain

$$Q_{1,k}(z) = \frac{x_1(\mu q + \mu p M_{k+1}(z, x_1))}{\lambda}, \quad 0 \le k \le N-2,$$

and

$$M_k(z,x) = \frac{-\lambda x Q_{1,k}(z) + \mu q x^2 + \mu p x^2 M_{k+1}(z,x) + g_1(x)}{g_1(x)}, \quad 0 \le k \le N-2.$$

From the last equation of (17), we have

$$Q_{1,-1}(z) = \frac{\xi(M_0(z,b)-1)}{\lambda}.$$

Finally, $Q(z)$ can be derived by

$$Q(z) = \pi_{0,-1} Q_{1,-1}(z) + \sum_{k=0}^{N-1} \pi_{0,k} Q_{1,k}(z).$$

■

7. NUMERICAL EXAMPLES

In this section, based on the obtained results, we present some numerical examples to illustrate the qualitative behavior of the key performance measures of the queueing system.

First, we investigate a renewal–reward cost structure (long-run rate of cost) by considering a cost associated with the accumulation of N negative feedbacks, a cost for the maintenance, as well as a cost for the customers that are lost, and explore the optimal N value. Then the objective function for minimizing the total cost per unit time is as follows:

$$\begin{aligned} \min_{N\ge 1} TC &= \min_{N\ge 1} \frac{N \times c_{\mathrm{f}}(N) + c_m}{E[C]} + c_{\mathrm{l}} \mu p \sum_{n=0}^{\infty} n P_{n,N-1} \\ &= \min_{N\ge 1} \frac{N \times c_{\mathrm{f}}(N) + c_{\mathrm{m}}}{E[V] + E[M]} + c_{\mathrm{l}} \mu p \sum_{n=0}^{\infty} n P_{n,N-1}, \end{aligned} \tag{18}$$

where c_{m} is the cost maintenance, $c_{\mathrm{f}}(N)$ is the cost of the server receiving a negative feedback, which is a function depending on the variable N, c_{l} is the cost for a customer that is lost, C is the duration of a cycle, M is the maintenance period in a cycle, V is the amount of time in a cycle during which the server is available. (In this renewal–reward cost structure, we mainly study the impact of the cost of the server receiving a negative feedback on the long-run rate of cost.) Then, under the stability condition, similarly to the argument of alternative renewal process and the renewal–reward theorem, the probability that the server in maintenance period can be rewritten as

$$P_{\mathrm{mp}} = \frac{E[M]}{E[M] + E[V]}. \tag{19}$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

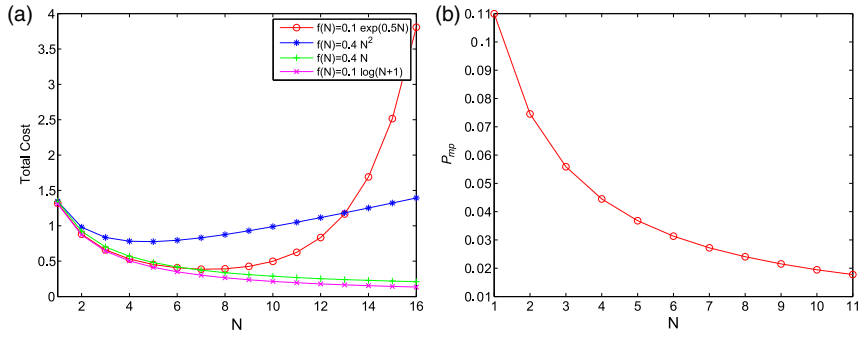


FIGURE **1.** TC and P_{mp} versus N, (a) Total cost in per unit time over $f(N)$ and N, (b) The percentage of the time spent in maintenance period P_{mp} versus N

Substituting $E[M] = 1/\xi$ and $P_{\mathrm{mp}} = \sum_{n=0}^{\infty} P_{n,-1}$ into (19), we have

$$E[M] + E[V] = \frac{1}{\xi \sum_{n=0}^{\infty} P_{n,-1}}.$$

Then, the objective function for minimizing the total cost per unit time can be simplified as

$$\begin{aligned} \min_{N\geq 1} TC &= \min_{N\geq 1} \left[\xi(N \times c_{\mathrm{f}}(N) + c_{\mathrm{m}}) \sum_{n=0}^{\infty} P_{n,-1} + c_{\mathrm{l}}\mu p \sum\nolimits_{n=0}^{\infty} n P_{n,N-1} \right] \\ &= \min_{N\geq 1} \left[\xi(N \times c_{\mathrm{f}}(N) + c_m) P_{mp} + c_{\mathrm{l}}\mu p \sum\nolimits_{n=0}^{\infty} n P_{n,N-1} \right]. \end{aligned} \tag{20}$$

Next, we examine the impact of N on the total cost per unit time. Without loss of generality, we assume $\mu = 3$, $\lambda = 2$, $p = 0.1$, $\xi = 0.2$, $c_{\mathrm{m}} = 2$ and $c_{\mathrm{l}} = 0.2$. Actually, the cost associated with the accumulation of N negative feedbacks increases with the increase of the number of negative feedbacks. However, in real life, especially in the context of a service system, a negative feedback incurs a goodwill cost as well as reputation damage for the system, due to anecdotal reasoning influences customers' decisions, customers may obtain several anecdotes about the system from their social network or public platforms such as third party review; therefore, bigger N may lead to more reputation damage, which increases the potential economic loss. Then, the cost associated with the accumulation of N negative feedbacks may not increase linearly with the increase of the number of negative feedbacks. So, we define $f(N) = N \times c_{\mathrm{f}}(N)$, and consider the following increasing functions $f(N) = 0.1e^{0.5N}$, $f(N) = 0.4N^2$, $f(N) = 0.4N$, $f(N) = 0.1\ln(N+1)$.

In Figure 1(a), we show the impact of N on the the total cost for different $f(N)$. From Figure 1(a), we find that, for $f(N) = 0.4N$ and $f(N) = 0.1\ln(N+1)$, the total cost decreases with the increase of N, for $f(N) = 0.1e^{0.5N}$ and $f(N) = 0.4N^2$, the total cost first decreases to a minimum value and then increases with the increase of N. Next, we give the interpretation of the phenomenon depicted in Figure 1(a) based on Eq. (20). First, from Figure 1(b), we observe that the percentage of the time spent in maintenance P_{mp} is a decreasing function of N, which is expected to decrease as N increases. Therefore, if we choose the cost function $f(N) = 0.4N$ and $f(N) = 0.1\ln(N+1)$, then the long-run rate of cost (total cost per unit time) TC is a decreasing function of N, which means that the percentage of the time spent in maintenance P_{mp} makes a main contribution to decrease the

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

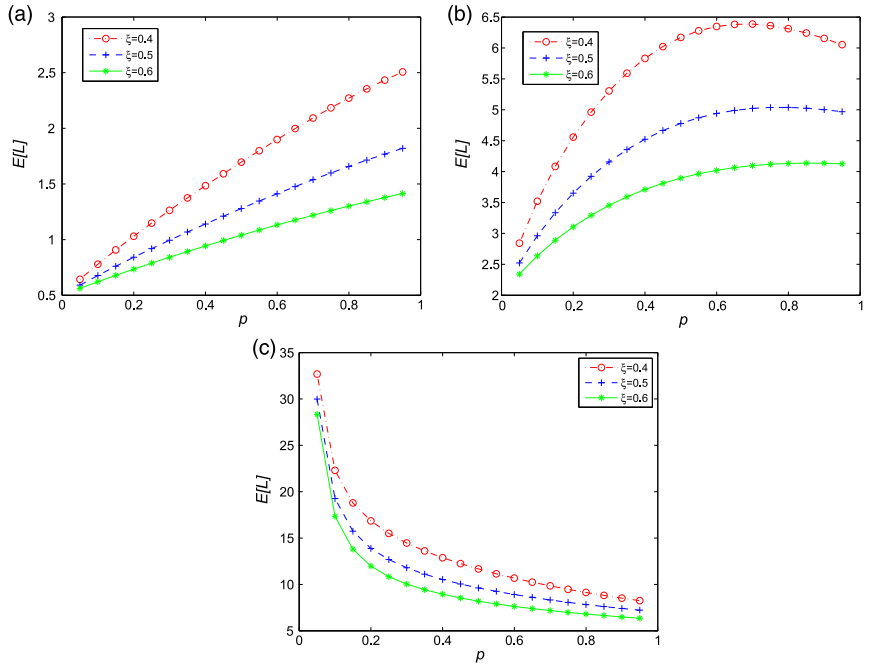


FIGURE **2.** $E[L]$ versus p for different ξ, (a) $E[L]$ versus p ($\mu = 3, \lambda = 1$), (b) $E[L]$ versus p ($\mu = 3, \lambda = 2$), (c) $E[L]$ versuss p ($\mu = 3, \lambda = 3$)

total cost. If we choose the cost function $f(N) = 0.1e^{0.5N}$ and $f(N) = 0.4N^2$, the long-run rate of cost TC is a convex function of N. For the two cost function $f(N) = 0.1e^{0.5N}$ and $f(N) = 0.4N^2$, when N is small, P_{mp} plays a dominant role in decreasing the total cost, then an increasing N will decrease the total cost. When N is large, the cost associated with the accumulation of N negative feedbacks plays a dominant role in increasing the total cost, then increasing N has a significant impact on the increase of total cost. Furthermore, from Figure 1(a), for $f(N) = 0.1e^{0.5N}$ and $f(N) = 0.4N^2$, we can obtain an optimal value N to minimize the total cost, that is, when $N = 7$ and $N = 5$, the total cost has the minimum value $TC = 0.3867$ and $TC = 0.7754$.

Next, we mainly study the impact of the system parameters on some performance measures.

In Figure 2, we respectively assume $\mu = 3, \lambda = 2$, $\mu = 3, \lambda = 1$ and $\mu = 3, \lambda = 3$. Then, Figure 2 shows the impact of p on the expected number of customers $E[L]$ for different values of ξ. Interestingly, from Figure 2(a), we see that $E[L]$ increases with the increase of p, from Figure 2(b), we see that $E[L]$ first increases with the increase of p and then decreases with the increasing of p, from Figure 2(c), we see that $E[L]$ decreases with the increase of p. It may be because, as λ approaches to μ, the value of p makes a key contribution to reducing the number of customers in the system. When $\mu - \lambda$ is larger, the system has a tendency to be empty; hence, as p increases, the server tends to be in maintenance period, which leads to the increase of the number of customers. Obviously, if p is fixed, the smaller ξ is, the larger $E[L]$ becomes.

Figure 3 depicts the impact of the parameter p on $E[L_{\text{s}}]$ for different values of N. Clearly, from Figure 3, we can see that $E[L_{\text{s}}]$ decreases as p increases from 0.05 to 0.95. It is reasonable that, as p increases, the server may receive a negative feedback with a higher probability, that is, the server may accumulate N negative feedbacks with serving a fewer

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

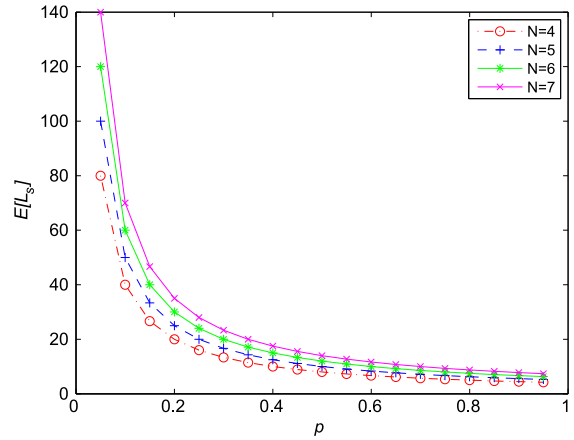


FIGURE **3.** The impact of p on $E[L_s]$

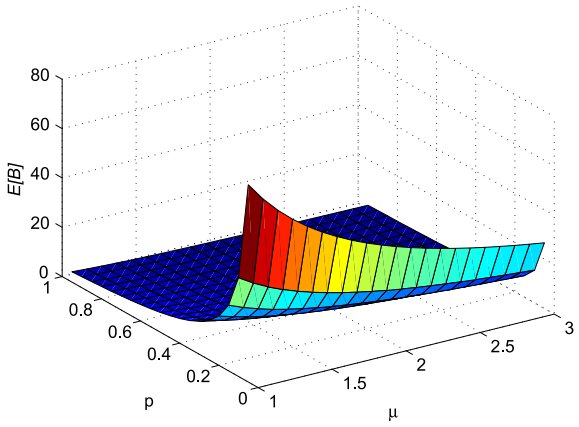


FIGURE **4.** The effect of p and μ on $E[B]$ $(N = 4)$

number of customers. It is also worthy to note that, if p is fixed, the bigger N is, the larger $E[L_s]$ becomes. It is because, the bigger N is, the server needs to serve more customers so that it can accumulate N negative feedbacks.

In Figure 4, we assume that $N = 4$ and explore the impact of the parameters μ and p on $E[B]$. As expected, Figure 4 reveals that $E[B]$ decreases with the increase of p and μ, respectively.

In Figure 5, we respectively assume $\mu = 3, \lambda = 1$ and $\mu = 3, \lambda = 5$, and pay attention to the curves of the expected duration of a busy period $E[T]$ with the change of p from 0.1 to 0.9 for different values of ξ. Without loss of generality, we assume that $N = 1$. Then, from Figure 5(a), we see that $E[T]$ increases with the increase of p as $\lambda < \mu$, from Figure 5(b), we observe that $E[T]$ is a decreasing function of p as $\lambda > \mu$. We also see that, if p is fixed, the smaller ξ is, the larger $E[T]$ becomes. Actually, the phenomena depicted in Figure 5 can be interpreted by equation

$$E[T] = \pi_{0,-1}E[T_{1,-1}] + \sum_{k=0}^{N-1} \pi_{0,k}e[T_{1,k}].$$

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

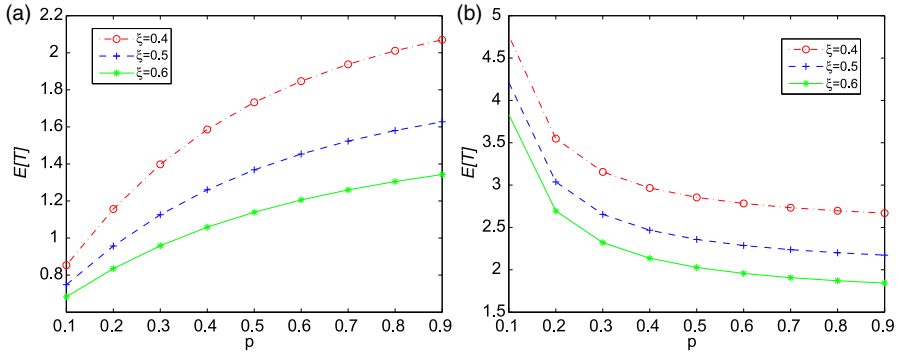


FIGURE **5.** $E[T]$ versus p for different ξ (a) $E[T]$ versus p ($\mu = 3, \lambda = 1$), (b) $E[T]$ versus p ($\mu = 3, \lambda = 5$)

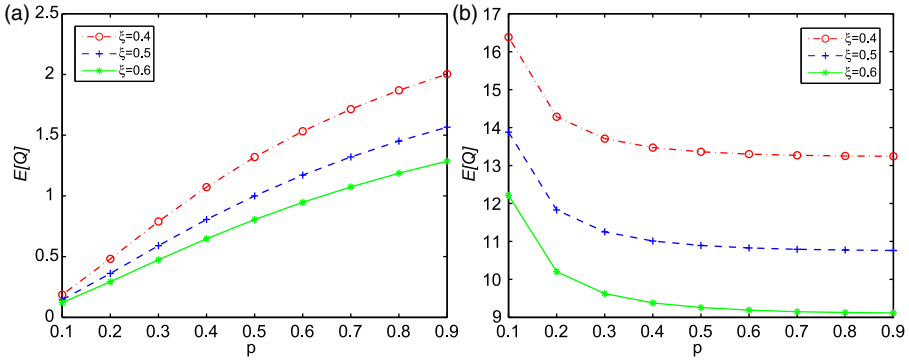


FIGURE **6.** $E[Q]$ versus p for different ξ (a) $E[Q]$ versus p ($\mu = 3, \lambda = 1$), (b) $E[Q]$ versus p ($\mu = 3, \lambda = 5$)

Intuitively, the mean values of $E[T_{1,-1}]$ and $E[T_{1,k}]$ are decreasing functions of p and $\pi_{0,-1}$ is an increasing function of p. However, if $\lambda < \mu$, as p increases, the value of $\pi_{0,-1}E[T_{1,-1}]$ may increase, which may play a dominant role in increasing the value of $E[T]$. If $\lambda > \mu$, as p increases, the value of $\sum_{k=0}^{N-1} \pi_{0,k}E[T_{1,k}]$ decreases, which plays a dominant role in decreasing $E[T]$.

In Figure 6, we assume $\mu = 3, \lambda = 1$ and $\mu = 3, \lambda = 5$, and plot the trend of the change for the expected number of lost customers in a busy period $E[Q]$ as p increases from 0.1 to 0.9 with different values of ξ. From Figures 6(a) and 6(b), we can easily see that $E[Q]$ increases with the increase of p as $\lambda < \mu$, and $E[Q]$ decreases with the increase of p as $\lambda > \mu$. It is obvious that, in Figure 6, if p is fixed, the smaller ξ is, the larger $E[Q]$ becomes. It is reasonable that, the smaller ξ is, the times that the server stays in maintenance period become longer, which leads to a larger value of $E[Q]$.

8. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

This paper is devoted to the study of a clearing queueing system with the server being subjected to maintenance after N negative feedbacks. Different from the queueing models subject to disasters, the model in consideration undergoes a maintenance procedure and all present customers are forced to abandon the system as soon as the server receives exactly

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

N negative feedbacks. By using the matrix analytic method, we first derived the stationary queue length distribution. Then, based on the results, we derived some performance measures. Furthermore, we gave detailed analysis on the sojourn time of an arbitrary customer, the busy period of the system and total number of lost customers in a busy period. Finally, some numerical examples are conducted to find the optimal value N and check the impacts of system parameters on some key performance measures, we expect that the results can be applied to more practical queueing systems.

Apart from the results obtained by the present model, many interesting and important topics on this queueing model have not yet been fully explored. It would be interesting to consider further extensions of the present queueing model. One direction is to give detailed analysis of the various generalizations of the present queueing model. For example, we could consider the case that N is a random variable instead of a constant, that is, we could investigate the queueing model with several distributions (e.g., geometric distribution, a finite mixture of geometric distributions, and general distribution) for N. We can also consider the queueing model with generally distributed service and maintenance times (i.e., the M/G/1 type queueing system). Actually, by using the same method, it may be difficult to investigate these generalized models, therefore, we expect to find another method to analyze these models in future research.

Another interesting direction for future research is to take another maintenance policy into consideration on the basis of the present queueing model, see for example, planned maintenance policy and preventive maintenance policy, and further explore the impact of varying maintenance cost on the long-run rate of cost. Since we only focus our analysis on a single production line of the production system, in fact, we can also take another production line into consideration and study this type model (i.e., we consider a multi-server multi-queue queueing system). Because of the dependency among the queues, it will be difficult to analyze the model and derive the explicit results by same method. (For the queueing system with multiple queues, we think that we can study the tail asymptotics property of the steady-state distributions, and then develop some approximation methods to calculate the steady-state probabilities.) The detailed analysis of this case requires a separate discussion and is left for future research.

Acknowledgement

The authors would like to thank the editor and the two referees for their helpful suggestions and comments to improve the quality of this paper.

References

1. Artalejo, J.R. & Gómez-Corral, A. (1998). Analysis of a stochastic clearing system with repeated attempts. *Communications in Statistics: Stochastic Models* 14: 623–645.
2. Boudali, O. & Economou, A. (2012). Optimal and equilibrium balking strategies in the single server Markovian queue with catastrophes. *European Journal of Operational Research* 218: 708–715.
3. Boudali, O. & Economou, A. (2013). The effect of catastrophes on the strategic customer behavior in queueing systems. *Naval Research Logistics* 60: 571–587.
4. Chakravarthy, S.R. (2009). A disaster queue with Markovian arrivals and impatient customers. *Applied Mathematics and Computation* 214: 48–59.
5. Dimou, S. & Economou, A. (2013). The single server queue with catastrophes and geometric reneging. *Methodology and Computing in Applied Probability* 15(3): 595–621.
6. Dimou, S., Economou, A., & Fakinos, D. (2011). The single server vacation queueing model with geometric abandonments. *Journal of Statistical Planning and Inference* 141: 2863–2877.
7. Economou, A. & Manou, A. (2013). Equilibrium balking strategies for a clearing system in alternating environment. *Annals of Operations Research* 208: 489–514.

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press

8. Giorno, V., Nobile, A.G., & Spina, S. (2014). On some time non-homogeneous queueing systems with catastrophes. *Applied Mathematics and Computation* 245: 220–234.
9. Jiang, T. & Liu, L. (2015). The GI/M/1 queue in a multi-phase service environment with disasters and working breakdowns. *International Journal of Computer Mathematics* 94(4): 707–726.
10. Jiang, T. & Liu, L. (2016). Analysis of a GI/M/1 queue in a multi-phase service environment with disasters. *RAIRO-Operations Research* 51: 79–100.
11. Jiang, T., Liu, L. & Li, J. (2015). Analysis of the M/G/1 queue in multi-phase random environment with disasters. *Journal of Mathematical Analysis and Applications* 430: 857–873.
12. Kapodistria, S., Phung-Duc, T., & Resing, J. (2016). Linear birth/immigration-death process with binomial catastrophes. *Probability in the Engineering and Informational Sciences* 30(1): 79–111.
13. Kim, B.K. & Lee, D.H. (2014). The M/G/1 queue with disasters and working breakdowns. *Applied Mathematical Modelling* 38: 1788–1798.
14. Mytalas, G.C. & Zazanis, M.A. (2015). An M^X/G/1 queueing system with disasters and repairs under a multiple adapted vacation policy. *Naval Research Logistics* 62: 171–189.
15. Neuts, M.F. (1981). *Matrix geometric solutions in stochastic models-an algorithmic approach*. Baltimore and London: The Johns Hopkins University Press.
16. Paz, N. & Yechiali, U. (2014). An M/M/1 queue in random environment with disasters. *Asia-Pacifc Journal of Operational Research* 31(3), Article 1450016, 12 pages.
17. Phung-Duc, T. (2017). Exact solutions for M/M/c/setup queues. *Telecommunication Systems* 64(2): 309–324.
18. Stidham, S. (1974). Stochastic clearing systems. *Stochastic Processes and their Applications* 2: 85–113.
19. Sudhesh, R. (2010). Transient analysis of a queue with system disasters and customer impatience. *Queueing Systems* 66: 95–105.
20. Sudhesh, R., Sebasthi Priya, R., & Lenin, R.B. (2016). Analysis of N-policy queues with disastrous breakdown. *TOP* 24: 612–634.
21. Van Houdt, B. & van Leeuwaarden, J.S.H. (2011). Triangular M/G/1-type and tree-like QBD Markov chains. *INFORMS Journal on Computing* 23(1): 165–171.
22. Yang, W.S., Kim, J.D., & Chae, K.C. (2002). Analysis of M/G/1 stochastic clearing systems. *Stochastic Analysis and Applications* 20: 1083–1100.
23. Yechiali, U. (2007). Queues with system disasters and impatient customers when system is down. *Queueing Systems* 56: 195–202.
24. Zhou, W., Zheng, Z. & Xie, W. (2017). A control-chart-based queueing approach for service facility maintenance with energy-delay tradeoff. *European Journal of Operational Research* 261: 613–625.

https://doi.org/10.1017/S026996481800013X Published online by Cambridge University Press