RAIRO-Oper. Res. 51 (2017) 101–122
DOI: 10.1051/ro/2016006

RAIRO Operations Research
www.rairo-ro.org

ANALYSIS OF D-POLICY DISCRETE-TIME $GEO/G/1$ QUEUE WITH SECOND J-OPTIONAL SERVICE AND UNRELIABLE SERVER *

Shaojun Lan[1] and Yinghui Tang[1]

Abstract. This paper is concerned with a discrete-time $Geo/G/1$ queueing system with D-policy and J-optional services in which the service station may be subject to failures at random during serving the customers. All the arriving customers require the first essential service, whereas some of them may opt for a second service from the J additional services with some probability. As soon as the system becomes empty, the server will not restart the service until the sum of the service times of the waiting customers in the system reaches or exceeds some given positive integer D. Applying the total probability decomposition law, renewal theory, and probability generating function technique, the queueing indices and reliability measures are investigated simultaneously in our work. Both the probability generating function of the transient queue length distribution and the explicit formulas of the steady-state queue length distribution at time epoch n^+ are derived. Meanwhile, the stochastic decomposition property is presented for the proposed model. Various reliability indices, including the transient and steady-state unavailability, the expected number of breakdowns during $(0^+, n^+]$, and the equilibrium failure frequency, are discussed. Finally, the optimum value of D for minimizing the system cost is numerically discussed under a given cost structure.

Mathematics Subject Classification. 60K25, 68M20, 90B22.

Received April 27, 2015. Accepted January 21, 2016.

1. Introduction

With the fast development of information era, there has been an increasing interest in investigating discrete-time queueing theory during the last few decades. Actually, since the digital computer and communication systems, such as broadband integrated services digital network (BISDN), time division multiple access (TDMA) and asynchronous transfer mode (ATM), operate on a discrete-time basis where the events (arrival of packets and their forward transmissions) can only occur at regularly spaced epochs, discrete-time queues are more suitable than their continuous-time counterparts for characterizing the behaviors of data communication and computer networks, which has become a powerful inducement to the research of discrete-time queueing theory. More comprehensive discussions and applications in the field of discrete-time queue can be found in the monographs by Hunter [8], Bruneel and Kim [3], and Woodward [25].

Keywords. Discrete-time queue, D-policy, unreliable server, second J-optional service, cost optimization.

* *This research is supported by the National Natural Science Foundation of China (No. 7157 1127).*

[1] School of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, Sichuan, P.R. China.
sjlan89@163.com; tangyh@uestc.edu.cn

In recent years, considerable attention has been paid to the study of queueing systems in which the server can provide the second optional service for the customers since the seminal paper was written by Madan [16]. The queueing situations incorporating the second optional service occur in many real service systems. For example, at a barber's shop all arriving customers may require a hair-cut but only some of them may further demand a shave after their hair-cut. In Madan [16], the first regular service time is assumed to be generally distributed, while the second optional service time is governed by an exponential distribution. Later, as a generalization of Madan [16], several researchers have extensively investigated the related topic concerning additional optional service and interested readers may refer to Atencia and Moreno [1], Wang and Zhao [19], Zhang and Zhu [26], and references therein. All the aforementioned papers suppose that the server provides two-phase optional service. However, in the real world we often encounter various situations where the servers may provide multiple optional services for all the units. Taking bank services as an example, a client may require some different types of services like opening an account, deposits or withdrawals, checking account balance, credit transfer, and so forth. The queueing models with multi-optional services not only improve the quality of the service and the satisfaction of customers, but also increase the economic benefits of system. Li and Wang [14] studied an $M/G/1$ retrial queue with feedback in which the server can provide the second multi-optional service for customers and the service station is unreliable. The various steady-state solutions for queueing measures and reliability indices are obtained by supplementary variable method. Ke [12] obtained some stationary results for an $M^{[X]}/G/1$ queueing model with startup time and J additional options for service. Jain and Upadhyaya [9] examined a batch arrival queue with N-policy and Bernoulli vacation where the server may be subject to breakdowns during customer's service and the customers can demand for a second multi-optional service with some probability. Kumar [13] investigated a discrete-time $Geo^X/G/1$ retrial queue wherein all the arriving customer require the first essential service and only a part of them may opt for one of other M-optional services. Some main performance measures were derived and the sensitivity analysis for several system characteristics is conducted by some numerical experiments. Recently, Jain *et al.* [10] provided an extensive analysis for an unreliable queueing system with multi-optional services and multi-optional vacations.

On the other hand, in many real life scenarios such as computer and communication networks, flexible manufacturing system, transportation system and production system, we often meet the case that the service station may fail more or less frequently when rendering service to the customers. The breakdowns of service facilities will result in a temporarily unavailable period of systems and therefore the performances of the systems will be heavily affected. In this context, the research of repairable queueing system is well worth doing from the viewpoint of queueing and reliability theory. In the recent past, remarkable contributions considering the unreliable queueing systems have been made by many authors, such as Atencia and Moreno [2], Tang *et al.* [17], Wang and Zhang [18], Lin and Ke [15], Gao *et al.* [7], Wang *et al.* [22], Gao and Liu [6], Wang [21].

In the present paper, we propose to investigate a discrete-time $Geo/G/1$ queueing model with J additional options for service and unreliable server, where the system operates under the control of D-policy, *i.e.*, the system is turned off once the customers in the system are served exhaustively and is turned on when the sum of the service times of all waiting customers exceeds a predetermined positive integer D. To the best of our knowledge, there is no research work on the proposed model. Our study generalizes the discrete time version of Choudhury [5] and Wang *et al.* [20] by assuming that the server can provide J-optional services and the service station may break down randomly. Employing the total probability decomposition law, renewal theory and probability generating function technique, we first discuss the queueing characteristics including the solutions for the transient queue size distribution and the explicit formulas of the stationary queue length distribution at time epoch n^+. Then, some reliability measures such as the unavailability of service station, the expected number of system failures during time interval $(0^+, n^+]$, the steady-state breakdown frequency of service station are derived. Further, in order to save the operating cost, we develop a long-run expected cost function per unit time to discuss the optimum value of D and some numerical experiments are provided to illustrate the effect of different system parameters and cost elements on the optimum value D^* and the corresponding minimum expected cost per unit time.

The incorporation of J-optional services, unreliable service station and D-policy is not only a generalization in mathematics, but also makes our model closer to real-world situations. A practical justification of the considered model is the following pump manufacturing system based on the work of Ke [12]. A pump manufacturing industry manufactures diverse types of pumps that require shafts of various dimensions. The shafts arrive at the Computer Numerical Control (CNC) copy turning center according to a Bernoulli process. When the total workload of the arriving shafts exceeds a given value D, the CNC machine commences the copy turning process. Some processed shafts may be defective and need to be further processed to meet the required standard. The CNC machine may break down at random and can be repaired. In this situation, the CNC machine, the workload D, the reprocessing of the defective shafts and the failures of CNC machine are corresponding to the server, D-policy, optional service and unreliable server, respectively, in the queueing terminology.

The remainder of this paper is structured as follows. In the following section, the considered mathematical model is formulated and some preliminaries are presented. Section 3 analyzes some queueing characteristics including the transient-state queue length distribution at arbitrary epoch n^+, the steady-state queue size distribution, and the mean of stationary queue length. In Section 4, some reliability indices of the service station are obtained. Moreover, in Section 5, we develop a long-run expected cost function per unit time to study numerically the optimal value for minimizing the system cost. At last, some conclusions are drawn in Section 6.

2. Model formulation and some preliminaries

We deal with a discrete-time $Geo/G/1$ queueing system with D-policy and second J-optional service in which the service station is unreliable. Different from the continuous-time queue, all the queueing activities (arrivals, departures and repairs), in discrete time queue, are nonnegative integer-valued random variables. The time axis is slotted into equal time intervals and the time axis is marked with $0, 1, 2, \ldots, n \ldots$ All the arrivals and departures only happen at boundary epochs of time slots in discrete-time case. In view of this fact, one arrival and one departure may take place simultaneously within a slot. So, it is necessary to stipulate the order of arrivals and departures. Generally speaking, there are two types of discrete-time models, namely, the early arrival system (EAS) and the late arrival system (LAS). And the late arrival system (LAS) can further be subdivided into late arrival system with delayed access (LAS-DA) and late arrival system with immediate access (LAS-IA). More details regarding these concepts can be referred to Hunter [8]. In the present research, we consider the late arrival system with delayed access (LAS-DA), that is, the arrivals and the breakdowns take place within (n^-, n), $n = 0, 1, 2, \ldots$, and the departures and the completion of the repairs occur within (n, n^+), $n = 1, 2, \ldots$ Moreover, we assume that there is no customer arrival within $(0^-, 0)$ and no departure within $(0, 0^+)$. To make it clearer, the various time epochs at which queueing events occur are displayed in Figure 1. The detailed mathematical model is described as follows.

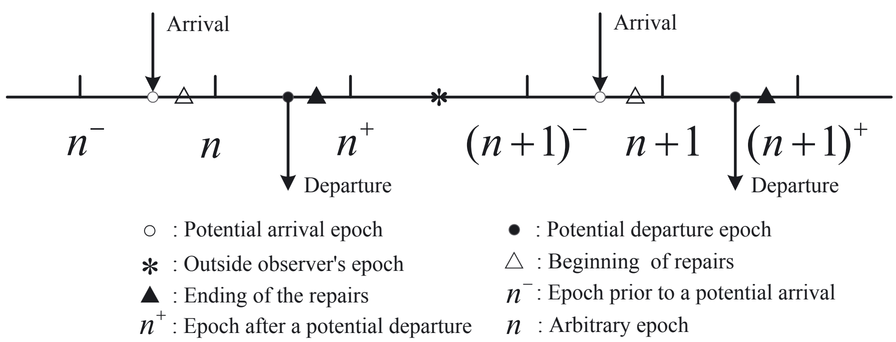

Figure 1. Various time epochs in a late arrival system with delayed access (LAS-DA).

Assumption 1. Arrival process: The customers' inter-arrival times $\{\tau_k, k \geq 1\}$ are independent and identically distributed (i.i.d.) discrete-time random variables with the same geometrical distribution $P\{\tau_k = j\} = \lambda(1-\lambda)^{j-1}$, $j = 1, 2, \ldots$, $0 < \lambda < 1$. That is to say, a customer arrives with probability λ, and no customer arrives with probability $1-\lambda$ in every time interval (n^-, n), $n = 1, 2, \ldots$

Assumption 2. Service process: There is only one service station in the system and the customers are served one by one according to the first-come-first-served (FCFS) discipline. The server renders the first essential service (FES) to all the arriving customers. The service times of FES, denoted by $\{\chi_k^{(0)},\ k \geq 1\}$, are i.i.d. discrete-time random variables and arbitrarily distributed with the probability mass function (p.m.f.) $P\{\chi_k^{(0)} = j\} = g_j^{(0)}$, $j \geq 1$, and probability generating function (p.g.f.) $G_0(z) = \sum_{j=1}^{\infty} g_j^{(0)} z^j$, $|z| < 1$. The expected value of service time is μ_0 $(1 \leq \mu_0 < \infty)$ and the second moment $E[(\chi_k^{(0)})^2]$ is finite. After the FES of a customer is completed, the customer may opt for the ith $(1 \leq i \leq J)$ type of second J-optional service with probability θ_i, in which scenario the ith type service of the customer starts at once, or may leave the system forever with complementary probability θ_0 (*i.e.*, $\theta_0 = 1 - \sum_{i=1}^{J} \theta_i$), in which scenario the next customer (if any) will be served with FES. The service times of the ith type of J additional services, designated by $\{\chi_k^{(i)},\ k \geq 1\}$, are i.i.d. with common distribution $P\{\chi_k^{(i)} = j\} = g_j^{(i)}$, p.g.f. $G_i(z) = \sum_{j=1}^{\infty} g_j^{(i)} z^j$, $|z| < 1$, and finite mean μ_i and the second moment $E[(\chi_k^{(i)})^2]$.

Assumption 3. Control policy: Once all the customers in the queue are processed exhaustively, the server remains in the system. When the total first essential service times of waiting customers is no less than a given positive integer D, the server resumes its service immediately until the system empties again.

Assumption 4. Unreliable service station: When a customer is being served with any type service (FES or second J-optional service), the service station may break down at random. The lifetime of service station, denoted by X, is assumed to be geometrically distributed with parameter α, which means that a breakdown probably occurs with probability α in a slot. The failures only take place in server's busy state. As soon as the service station fails, it is urgently sent to fix at a repair facility and the customer just being served before server failure has to wait until the server is repaired. The repair time, denoted by R, follows a general distribution with p.m.f. $P\{R = j\} = r_j$, $j = 0, 1, \ldots$, p.g.f. $R(z) = \sum_{j=0}^{\infty} r_j z^j$ and finite mean β. It is supposed that after repairing, the server functions as well as before the breakdown, and the remaining service of interrupted customer will go on (*i.e.*, the customer's service time is cumulative). Also, we assume that the service station is new at initial time $n^+ = 0^+$.

Assumption 5. As usual, we assume that various random variables involved in the system are mutually independent of each other. It is also supposed that the server will stay standby and wait for the first arrival if there is no customer at initial time $n^+ = 0^+$. After the first busy period, the system begins to take the D-policy.

For later discussions, we first present some preliminaries as follows.

Definition 2.1. "System idle period" is the time length from the instant when the system becomes empty until the instant when the first customer arrives. Let $\tilde{\tau}_k$ $(k = 1, 2, \ldots)$ be the kth system idle period. Thus, $\tilde{\tau}_k$ are independent mutually and satisfy the same geometric distribution with parameter λ (*i.e.*, $P\{\tilde{\tau}_k = j\} = \lambda(1-\lambda)^{j-1}$, $j \geq 1$) due to the Markov property.

Definition 2.2. "Server idle period" commences when the system is completely empty and finishes when server begins to deal with the waiting customers.

Definition 2.3. "Generalized service time" is the time length between the time when the service of a customer commences and the time when the service of the customer is completed, which may contain some repair times of the service station owing to its failures during the service period of the customer.

Denote by $\tilde{\chi}^{(0)}$ and $\tilde{\chi}^{(i)}$ $(i = 1, 2, \ldots)$ the generalized service time of a customer in FES and the ith type service of second J-optional service, respectively. For $j \geq 1$, $i \geq 1$, let $\tilde{g}_j^{(0)} = P\left\{\tilde{\chi}^{(0)} = j\right\}$ be the distribution function of generalized service time in FES and $\tilde{g}_j^{(i)} = P\left\{\tilde{\chi}^{(i)} = j\right\}$ be the distribution function of generalized service time in the ith type service of second J-optional service, respectively. Thus, from Tang *et al.* [17], we have

$$\begin{aligned}
\tilde{g}_j^{(0)} &= P\left\{\tilde{\chi}^{(0)} = j\right\} \\
&= \sum_{l=1}^{j} P\left\{\chi^{(0)} = l\right\} \sum_{k=0}^{l} P\left\{\sum_{s=1}^{k} R_s = j - l\right\} \binom{l}{k} \alpha^k (1-\alpha)^{l-k}, \qquad (2.1) \\
\tilde{g}_j^{(i)} &= P\left\{\tilde{\chi}^{(i)} = j\right\} \\
&= \sum_{l=1}^{j} P\left\{\chi^{(i)} = l\right\} \sum_{k=0}^{l} P\left\{\sum_{s=1}^{k} R_s = j - l\right\} \binom{l}{k} \alpha^k (1-\alpha)^{l-k}, \ i \geq 1, \qquad (2.2)
\end{aligned}$$

where R_s represents the sth repair time. For $|z| < 1$, the probability generating functions of $\tilde{\chi}^{(0)}$ and $\tilde{\chi}^{(i)}$ are given respectively by

$$\tilde{G}_0(z) = \sum_{j=1}^{\infty} \tilde{g}_j^{(0)} z^j = G_0(\alpha z R(z) + z(1-\alpha)) \tag{2.3}$$

and

$$\tilde{G}_i(z) = \sum_{j=1}^{\infty} \tilde{g}_j^{(i)} z^j = G_i(\alpha z R(z) + z(1-\alpha)). \tag{2.4}$$

The respective expected values of $\tilde{\chi}^{(0)}$ and $\tilde{\chi}^{(i)}$ are

$$E\left[\tilde{\chi}^{(0)}\right] = \frac{\mathrm{d}}{\mathrm{d}z}\tilde{G}_0(z)|_{z=1} = \mu_0(1 + \alpha\beta) \tag{2.5}$$

and

$$E\left[\tilde{\chi}^{(i)}\right] = \frac{\mathrm{d}}{\mathrm{d}z}\tilde{G}_i(z)|_{z=1} = \mu_i(1 + \alpha\beta). \tag{2.6}$$

Definition 2.4. "Total generalized service time" is the time length from the time when the service of a customer begins until the time when the customer is served completely and leaves the system, which consists of the FES, the probable second J-optional service and the potential repair times due to server breakdowns during every service.

Let $\hat{\chi}$ be the total generalized service time of a customer and $\hat{g}_j = P\{\hat{\chi} = j\}$ be the corresponding p.m.f.. According to the model assumptions, we can get

$$\hat{\chi} = \theta_0\tilde{\chi}^{(0)} + \theta_1\left(\tilde{\chi}^{(0)} + \tilde{\chi}^{(1)}\right) + \theta_2\left(\tilde{\chi}^{(0)} + \tilde{\chi}^{(2)}\right) + \ldots + \theta_J\left(\tilde{\chi}^{(0)} + \tilde{\chi}^{(J)}\right).$$

Thus,

$$\hat{g}_j = P\{\hat{\chi} = j\} = \theta_0 P\left\{\tilde{\chi}^{(0)} = j\right\} + \sum_{i=1}^{J} \theta_i P\left\{\tilde{\chi}^{(0)} + \tilde{\chi}^{(i)} = j\right\}, \ j \geq 1. \tag{2.7}$$

From the above equation, the p.g.f. of $\hat{\chi}$ is given by

$$\hat{G}(z) = \sum_{j=1}^{\infty} \hat{g}_j z^j = \theta_0 \tilde{G}_0(z) + \sum_{i=1}^{J} \theta_i \tilde{G}_i(z) \tilde{G}_0(z), \ |z| < 1, \tag{2.8}$$

which implies that the mean of total generalized service time $\hat{\chi}$ is

$$E[\hat{\chi}] = \left(\mu_0 + \sum_{i=1}^{J} \theta_i \mu_i \right) (1 + \alpha\beta). \tag{2.9}$$

Definition 2.5. "Generalized busy period" denotes the time interval between the time when the server begins to serve customers and the moment when the system becomes idle again.

Denote by b the length of the server's generalized busy period initiated with only one customer. Then, similar to the analysis in Bruneel and Kim [3], we have lemma as follows.

Lemma 2.6. *Let $B(z) = \sum_{j=1}^{\infty} P\{b = j\} z^j$ be the p.g.f. of b. For $|z| < 1$, $B(z)$ is the root of the equation $B(z) = \hat{G}\left(\left(\bar{\lambda} + \lambda B(z)\right) z\right)$, and the mean value is given by*

$$E[b] = \begin{cases} \frac{\rho}{\lambda(1-\rho)}, & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases}$$

where $\bar{\lambda} = 1 - \lambda$, $\rho = \lambda \left(\mu_0 + \sum_{i=1}^{J} \theta_i \mu_i \right) (1 + \alpha\beta)$ represents the traffic intensify of the considered model.

Let $b^{\langle i \rangle}$ be the server's generalized busy period initiated with $i\,(i \geq 1)$ customers. Due to the Markov property of the arrival interval, $b^{\langle i \rangle}$ can be expressed as $b^{\langle i \rangle} = b_1 + b_2 + \ldots + b_i$, where $b_1, b_2, \ldots, b_i$ are independent of each other and follow the same distribution as b. Thus, the p.g.f. of $b^{\langle i \rangle}$ is

$$\sum_{j=i}^{\infty} P\left\{ b^{\langle i \rangle} = j \right\} z^j = B^i(z), \ |z| < 1.$$

Let $N(n^+)$ designate the number of customers in the system at time epoch n^+ and $Q_j(n^+) = P\{b > n^+, N(n^+) = j\}$, $j \geq 1$ denote the transient probability of there being j customers at epoch n^+ in generalized busy period b, where the epoch $n^+ = 0^+$ is the beginning of b. According to the meaning of b, we have the boundary condition $Q_1(0^+) = 1$, $Q_j(0^+) = 0$, $j = 2, 3, \ldots$ With the similar discussion in reference Wei *et al.* [23], we have the following lemma.

Lemma 2.7. *Let $q_j^+(z) = \sum_{n=0}^{\infty} Q_j(n^+) z^n$, $|z| < 1$ be the p.g.f. of $Q_j(n^+)$, then the recursive formulas of $q_j^+(z)$ can be expressed as*

$$q_1^+(z) = \frac{B(z)\left[1 - \hat{G}\left(\bar{\lambda} z\right)\right]}{\left(1 - \bar{\lambda} z\right) \hat{G}\left(\bar{\lambda} z\right)},$$

$$q_j^+(z) = \frac{1}{\hat{G}\left(\bar{\lambda} z\right)} \Bigg\{ B(z) \sum_{n=j-1}^{\infty} z^n \sum_{k=n+1}^{\infty} \hat{g}_k \binom{n}{j-1} \lambda^{j-1} \bar{\lambda}^{n-j+1} + \sum_{i=1}^{j-1} \frac{q_{j-i}^+(z)}{B^i(z)}$$
$$\times \left[B(z) - \hat{G}\left(\bar{\lambda} z\right) - \sum_{m=1}^{i} \sum_{k=m}^{\infty} \binom{k}{m} \hat{g}_k z^k [\lambda B(z)]^m \bar{\lambda}^{k-m} \right] \Bigg\}, \ j = 2, 3, \ldots,$$

where $B(z)$ is defined as in Lemma 2.6, $\hat{g}_k$ is given by (2.7), $\hat{G}\left(\bar{\lambda} z\right) = \sum_{j=1}^{\infty} \hat{g}_j \left(\bar{\lambda} z\right)^j$, $\binom{k}{m} = \frac{k!}{m!(k-m)!}$, $k \geq m \geq 1$, $\binom{k}{0} = 1$ and $\binom{k}{m} = 0$ if $k < m$. $\sum_{i=1}^{j} = 0$ if $j < i$.

3. Analysis of system queue length distributions

In this section, by total probability decomposition law and renewal theory, we first derive the p.g.f. of transient queue length distribution at any epoch n^+. Then, based on the transient analysis, the explicit formulas for steady-state queue size are obtained.

3.1. The transient distribution of the queue length at epoch n^+

Let $P_{ij}(n^+) = P\{N(n^+) = j\,|N(0^+) = i\}$ be the conditional probability of there being j customers at epoch n^+ under arbitrary initial state $N(0^+) = i\,(i = 0, 1, \ldots)$. The p.g.f. of $P_{ij}(n^+)$ is given by $p^+_{ij}(z) = \sum_{n=0}^{\infty} P_{ij}(n^+)\, z^n,\ (i, j = 0, 1, 2, \ldots)$.

Theorem 3.1. *For $|z| < 1$, we have*

$$p^+_{00}(z) = \frac{1}{1-\bar{\lambda}z}\left\{1 + \frac{f(z)\,B(z)}{1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)B(z)]^k}\right\}, \tag{3.1}$$

$$p^+_{i0}(z) = \frac{1}{1-\bar{\lambda}z}\cdot\frac{B^i(z)}{1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)B(z)]^k},\ \ i \ge 1, \tag{3.2}$$

where $f(z) = \frac{\lambda z}{1-\bar{\lambda}z}$, $A_k(D) = Y^{(k-1)}(D) - Y^{(k)}(D)$, $Y^{(k)}(D) = P\big\{\sum_{j=1}^{k}\chi_j^{(0)} < D\big\}$, $k = 1, 2, \ldots, D$, $Y^{(0)}(D) = 1$, $Y^{(D)}(D) = 0$.

Proof. Let $S_k = \sum_{i=1}^{k}\chi_i^{(0)}$, $l_k = \sum_{i=1}^{k}\tau_i$, $k = 1, 2, \ldots$, and $S_0 = l_0 = 0$. It is noted that $P_{00}(n^+)$ indicates that there is no customer in the queue at epoch n^+ under initial state $N(0^+) = 0$, *i.e.*, epoch n^+ is located in the system idle period. Based on the previous model assumptions, the beginning and ending epochs of the server's generalized busy period are renewal points. Employing renewal theory and the total probability decomposition law, we obtain

$$\begin{aligned} P_{00}\left(n^+\right) = {} & P\left\{0 \le n^+ < \tilde{\tau}_1\right\} + P\left\{\tilde{\tau}_1 + b_1 \le n^+ < \tilde{\tau}_1 + b_1 + \tilde{\tau}_2\right\} \\ & + \sum_{k=1}^{D} P\left\{\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{k-1} \le n^+, S_{k-1} < D \le S_k, N\left(n^+\right) = 0\right\}, \end{aligned} \tag{3.3}$$

where $\tilde{\tau}_k (k \ge 1)$ represent the length of the kth system idle period, $b_k\,(k \ge 1)$ denote the length of kth server's generalized busy period. The first term of (3.3) means that epoch n^+ is located in the first system idle period, which is equal to

$$\sum_{t=n+1}^{\infty} P\{\tilde{\tau}_1 = t\} = \sum_{t=n+1}^{\infty} \lambda\bar{\lambda}^{t-1} = \bar{\lambda}^n. \tag{3.4}$$

The second term of equation (3.3) is the probability that epoch n^+ is in the second system idle period. Since customers arrive at system according to Bernoulli process, the completion epoch of first generalized busy period b_1 is a renewal point. So we have

$$P\left\{\tilde{\tau}_1 + b_1 \le n^+ < \tilde{\tau}_1 + b_1 + \tilde{\tau}_2\right\} = \sum_{t=2}^{n} P\{\tilde{\tau}_1 + b_1 = t\}\,\bar{\lambda}^{n-t}. \tag{3.5}$$

The representation "$\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{k-1} \le n^+,\ S_{k-1} < D \le S_k$" in the third term of (3.3) means that there are k customers at the starting point of the second server's generalized busy period b_2. As the initiation point of b_2 is a renewal epoch, the third term of equation (3.3) is

$$\sum_{k=1}^{D} A_k(D) \sum_{t=2}^{n} P\left\{\tilde{\tau}_1 + b_1 = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_2 + l_{k-1} = m\right\} P_{k0}\left((n-t-m)^+\right). \tag{3.6}$$

Substituting equations (3.4)–(3.6) into equation (3.3), and Multiplying (3.3) by z^n and summing over n, it finally leads to

$$p_{00}^{+}(z) = \frac{1}{1-z\bar{\lambda}} + \frac{f(z)\,B(z)}{1-z\bar{\lambda}} + f(z)B(z)\sum_{k=1}^{D} A_k(D)\left[f(z)\right]^k p_{k0}^{+}(z). \tag{3.7}$$

For $i \ge 1$, $P_{i0}(n^+)$ denotes there being no customer in the system at instant n^+ under initial condition $N(0^+) = i$. Similar to the analysis of $P_{00}(n^+)$, we can get

$$\begin{aligned} P_{i0}\left(n^+\right) = &\sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\}\bar{\lambda}^{n-t} + \sum_{k=1}^{D} A_k(D) \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \\ &\times \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_1 + l_{k-1} = m\right\} P_{k0}\left((n-t-m)^+\right). \end{aligned} \tag{3.8}$$

Multiplying (3.8) by z^n and summing over n, it leads to

$$p_{i0}^{+}(z) = \frac{B^i(z)}{1-z\bar{\lambda}} + B^i(z)\sum_{k=1}^{D} A_k(D)\left[f(z)\right]^k p_{k0}^{+}(z). \tag{3.9}$$

From (3.7) and (3.9), the relationship between $p_{00}^{+}(z)$ and $p_{i0}^{+}(z)$ is given by

$$p_{i0}^{+}(z) = \frac{B^{i-1}(z)}{f(z)}\left[p_{00}^{+}(z) - \frac{1}{1-z\bar{\lambda}}\right],\ i = 1, 2, \ldots \tag{3.10}$$

Taking (3.10) into (3.7) gives (3.1). Furthermore, we can get (3.2) by substituting (3.1) into (3.10). ☐

Theorem 3.2. *For $|z| < 1$, $i \ge 1$, and $j = 1, 2, \ldots, D-1$, we get*

$$p_{0j}^{+}(z) = f(z)\left\{q_j^{+}(z) + \frac{\theta_j(z)}{1 - \sum\limits_{k=1}^{D} A_k(D)\left[f(z)B(z)\right]^k}\right\}, \tag{3.11}$$

$$p_{ij}^{+}(z) = \sum_{r=1}^{i} B^{r-1}(z)\, q_{j-i+r}^{+}(z) + \frac{B^{i-1}(z)\,\theta_j(z)}{1 - \sum\limits_{k=1}^{D} A_k(D)\left[f(z)B(z)\right]^k}, \tag{3.12}$$

where $q_j^{+}(z)$ is determined by Lemma 2.7, and

$$\begin{aligned} \theta_j(z) = &\frac{B(z)\left[f(z)\right]^j Y^{(j)}(D)}{1-z\bar{\lambda}} + \sum_{k=1}^{j} A_k(D)\left[f(z)\right]^k \sum_{r=1}^{k} B^r(z) q_{j-k+r}^{+}(z) \\ &+ \sum_{k=j+1}^{D} A_k(D)\left[f(z)\right]^k \sum_{r=k-j+1}^{k} B^r(z) q_{j-k+r}^{+}(z). \end{aligned}$$

Proof. For $j = 1, 2, \ldots, D-1$, there are j customers in the system at epoch n^+ if and only if time point n^+ is located in the generalized busy period of server or server idle period with j customers. With the same method used in Theorem 3.1, it gives

$$\begin{aligned}
P_{0j}\left(n^+\right) = &P\left\{\tilde{\tau}_1 \le n^+ < \tilde{\tau}_1 + b_1, N\left(n^+\right) = j\right\} \\
&+ P\left\{\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{j-1} \le n^+ < \tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_j, S_j < D\right\} \\
&+ \sum_{k=1}^{D} P\left\{\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{k-1} \le n^+, S_{k-1} < D \le S_k, N\left(n^+\right) = j\right\} \\
= &\sum_{t=1}^{n} P\left\{\tilde{\tau}_1 = t\right\} Q_j\left((n-t)^+\right) \\
&+ \sum_{t=2}^{n} P\left\{\tilde{\tau}_1 + b_1 = t\right\} \sum_{m=j}^{n-t} P\left\{\tilde{\tau}_2 + l_{j-1} = m\right\} P\left\{\tau_j > (n-t-m)^+\right\} Y^{(j)}(D) \\
&+ \sum_{k=1}^{D} A_k(D) \sum_{t=2}^{n} P\left\{\tilde{\tau}_1 + b_1 = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_2 + l_{k-1} = m\right\} P_{kj}\left((n-t-m)^+\right).
\end{aligned} \tag{3.13}$$

Analogously, for $i \ge 1$, we obtain

$$\begin{aligned}
P_{ij}(n^+) = &\sum_{r=1}^{i} \sum_{m=r-1}^{n} P\left\{b_1 + b_2 + \ldots + b_{r-1} = m\right\} Q_{j-i+r}\left((n-m)^+\right) \\
&+ \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \sum_{m=j}^{n-t} P\left\{\tilde{\tau}_1 + l_{j-1} = m\right\} P\left\{\tau_j > (n-t-m)^+\right\} Y^{(j)}(D) \\
&+ \sum_{k=1}^{D} A_k(D) \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_1 + l_{k-1} = m\right\} P_{kj}\left((n-t-m)^+\right).
\end{aligned} \tag{3.14}$$

Multiplying both equations (3.13) and (3.14) by z^n and adding over n from 0 to 1, it gives

$$\begin{aligned}
p_{0j}^+(z) = &f(z)\, q_j^+(z) + \frac{B(z)\, Y^{(j)}(D)\,[f(z)]^{j+1}}{1 - z\bar{\lambda}} \\
&+ f(z) B(z) \sum_{k=1}^{D} A_k(D)\,[f(z)]^k p_{kj}^+(z).
\end{aligned} \tag{3.15}$$

$$\begin{aligned}
p_{ij}^+(z) = &\sum_{r=1}^{i} B^{r-1}(z)\, q_{j-i+r}^+(z) + \frac{B^i(z)\, Y^{(j)}(D)\,[f(z)]^j}{1 - z\bar{\lambda}} \\
&+ B^i(z) \sum_{k=1}^{D} A_k(D)\,[f(z)]^k p_{kj}^+(z).
\end{aligned} \tag{3.16}$$

Solving (3.15) and (3.16) leads to the expressions of $p_{0j}^+(z)$ and $p_{ij}^+(z)$. $\square$

Theorem 3.3. *For $|z|<1$, $i\geq 1$ and $j=D,D+1,\ldots$, we have*

$$p_{0j}^{+}(z)=f(z)\left\{q_{j}^{+}(z)+\frac{\sigma_{j}(z)}{1-\sum\limits_{k=1}^{D}A_{k}(D)\left[f(z)B(z)\right]^{k}}\right\}, \tag{3.17}$$

$$p_{ij}^{+}(z)=\sum_{r=1}^{i}B^{r-1}(z)\,q_{j-i+r}^{+}(z)+\frac{B^{i-1}(z)\,\sigma_{j}(z)}{1-\sum\limits_{k=1}^{D}A_{k}(D)\left[f(z)B(z)\right]^{k}}, \tag{3.18}$$

where $\sigma_{j}(z)=\sum_{k=1}^{D}A_{k}(D)\left[f(z)\right]^{k}\sum_{r=1}^{r}B^{r}(z)q_{j-k+r}^{+}(z)$.

Proof. For $j=D,D+1,\ldots$, $N(n^{+})=j$ indicates that epoch n^{+} is located in the server's generalized busy period with j customers in the system. Using the same derivation process as in Theorem 3.2, we can complete the proof of Theorem 3.3. □

3.2. The recursive formulas for the steady-state distribution of queue length at epoch n^{+}

On the basis of the transient distribution of the queue length at arbitrary epoch n^{+} derived in Theorems 3.1–3.3, the recursive expressions of steady-state queue length distribution at epoch n^{+} will be investigated in this subsection.

Theorem 3.4. *Let $p_{j}^{+}=\lim\limits_{n\to\infty}P\left\{N\left(n^{+}\right)=j\right\}$, $j=0,1,2,\ldots$ be the probability that there are j customers in the system under steady state. Then, we have*

(a) *If $\rho=\lambda\left(\mu_{0}+\sum_{i=1}^{J}\theta_{i}\mu_{i}\right)(1+\alpha\beta)\geq 1$, then $p_{j}^{+}=0$, $j=0,1,2,\ldots$*

(b) *For $\rho=\lambda\left(\mu_{0}+\sum_{i=1}^{J}\theta_{i}\mu_{i}\right)(1+\alpha\beta)<1$, the equilibrium distribution $\{p_{j}^{+},j\geq 0\}$ at epoch n^{+} exists and forms a probability distribution. The recursive expressions are given by*

$$p_{0}^{+}=\frac{1-\rho}{1+\sum\limits_{k=1}^{D-1}Y^{(k)}(D)}, \tag{3.19}$$

$$p_{j}^{+}=\lambda p_{0}^{+}\left\{\frac{Y^{(j)}(D)}{\lambda}+\sum_{k=1}^{j}A_{k}(D)\sum_{r=1}^{k}q_{j-k+r}^{+}(1)+\sum_{k=j+1}^{D}A_{k}(D)\sum_{r=k-j+1}^{k}q_{j-k+r}^{+}(1)\right\},\quad j=1,2,\ldots,D-1, \tag{3.20}$$

$$p_{j}^{+}=\lambda p_{0}^{+}\sum_{k=1}^{D}A_{k}(D)\sum_{r=1}^{k}q_{j-k+r}^{+}(1),\ j=D,D+1,\ldots, \tag{3.21}$$

where

$$q_{1}^{+}(1)=\frac{1-\hat{G}(\bar{\lambda})}{\lambda\hat{G}(\bar{\lambda})},$$

$$q_{j}^{+}(1)=\frac{1}{\hat{G}(\bar{\lambda})}\left\{\sum_{n=j-1}^{\infty}\sum_{k=n+1}^{\infty}\hat{g}_{k}\binom{n}{j-1}\lambda^{j-1}\bar{\lambda}^{n-j+1}+\sum_{i=1}^{j-1}q_{j-i}^{+}(1)\left[1-\hat{G}(\bar{\lambda})-\sum_{m=1}^{i}\sum_{k=m}^{\infty}\binom{k}{m}\hat{g}_{k}\lambda^{m}\bar{\lambda}^{k-m}\right]\right\},\ j\geq 2.$$

Proof. Before discussing the steady-state queue length distribution, we first present the stability condition of the system under study. To this end, we define ς_m, $m = 1, 2, \ldots$ to be, after $n^+ = 0^+$, the mth time epoch at which the service of a customer is just finished and the system becomes empty. Based on the model assumptions and the memoryless property of the geometric distribution, the time interval $T_m = \varsigma_m - \varsigma_{m-1}(m = 1, 2, \ldots; \varsigma_0 = 0)$ is the length of the mth regeneration cycle. In order to keep the system stable, we should prove the expected length of a regeneration cycle is finite by the theorem on generative process (see Wolff [24]). Thus, our aim is to prove $E[T_m] < \infty$.

Obviously, a regeneration cycle in our system consists of a server's generalized busy period and a server's idle period. Let Q_b be the number of customers in the system at the initiation point of a server's generalized busy period, B be the server's generalized busy period beginning with Q_b customers, I be the server idle period. Since one customer requires at least one time slot for serving in discrete-time queue, there are, based on the D-policy, at most D customers in the system at the beginning of a server's generalized busy period. Thus, we have

$$P\left\{Q_b = r\right\} = P\left\{\sum_{k=1}^{r-1} \chi_k^{(0)} < D \leq \sum_{k=1}^{r} \chi_k^{(0)}\right\},\ r = 1, 2, \ldots, D-1.$$

$$P\left\{Q_b = D\right\} = P\left\{\sum_{k=1}^{D-1} \chi_k^{(0)} < D\right\}.$$

From the above two formulas, the mean of Q_b is given by

$$E\left[Q_b\right] = \sum_{k=1}^{D} kP\left\{Q_b = k\right\} = \sum_{k=0}^{D-1} Y^{(k)}(D).$$

Furthermore, the mean of the server's generalized busy period is

$$E\left[B\right] = E\left[b\right] \cdot E\left[Q_b\right], \tag{3.22}$$

where $E[b]$ is determined by Lemma 2.6. Since the inter-arrival times during server idle period follow independent and identical geometric distribution with mean $1/\lambda$, the expected length of server idle period is

$$E[I] = \frac{E\left[Q_b\right]}{\lambda}\cdot \tag{3.23}$$

From (3.22), (3.23) and Lemma 2.6, $E\left[T_m\right]$ can be expressed as

$$E\left[T_m\right] = E[B] + E[I] = \begin{cases} \frac{\sum\limits_{k=0}^{D-1} Y^{(k)}(D)}{\lambda(1-\rho)} < \infty, & \rho < 1, \\ \infty, & \rho \geq 1. \end{cases} \tag{3.24}$$

Therefore, we must have $\rho < 1$ for the system to be stable.

Now we analyze the steady-state queue length distribution. In discrete-time situation we have $p_j^+ = \lim\limits_{z \to 1^-} (1-z)\, p_{ij}^+(z)$ (see Jury [11]). Applying Lemma 2.6, Theorems 3.1–3.3 and L'Hospital rule, the expressions of Theorem 3.4 can be derived. Further, by manipulating direct calculations, the formula $\sum_{j=0}^{\infty} p_j^+ = 1$ holds, *i.e.*, $\left\{p_j^+,\ j = 0, 1, 2, \ldots\right\}$ is a probability distribution. □

Theorem 3.5. *Let* $\pi^+(z) = \sum_{j=0}^{\infty} p_j^+ z^j,\ |z| < 1$ *be the p.g.f of stationary queue size distribution* $\left\{p_j^+,\ j = 0, 1, 2, \ldots\right\}$ *at epoch* n^+*. Conditioning on* $\rho = \lambda(\mu_0 + \sum_{i=1}^{J} \theta_i \mu_i)(1 + \alpha\beta) < 1$*, we get*

$$\pi^+(z) = \frac{(1-\rho)(1-z)\,\hat{G}\left(z\lambda + \bar{\lambda}\right)}{\hat{G}\left(z\lambda + \bar{\lambda}\right) - z} \cdot \frac{1 + \sum\limits_{k=1}^{D-1} z^k Y^{(k)}(D)}{1 + \sum\limits_{k=1}^{D-1} Y^{(k)}(D)}, \tag{3.25}$$

and the mean of steady-state queue length, denoted by $E\left[L^{+}\right]$*, is presented by*

$$E[L^{+}] = \rho + \frac{\lambda^{2} E\left[\hat{\chi}\left(\hat{\chi}-1\right)\right]}{2\left(1-\rho\right)} + \frac{\sum\limits_{k=1}^{D-1} k Y^{(k)}(D)}{1+\sum\limits_{k=1}^{D-1} Y^{(k)}(D)}, \tag{3.26}$$

where $\hat{\chi}$ *denotes the total generalized service time,* $E\left[\hat{\chi}\left(\hat{\chi}-1\right)\right] = \sum_{j=2}^{\infty} j(j-1) P\left\{\hat{\chi}=j\right\}$.

Proof. Utilizing the expressions of p_j^{+} given in Theorem 3.4 and noticing that

$$\sum_{j=1}^{\infty} q_j^{+}\left(1\right) z^{j} = \frac{z\left[1-\hat{G}\left(z\lambda+\bar{\lambda}\right)\right]}{\lambda\left[\hat{G}\left(z\lambda+\bar{\lambda}\right)-z\right]},$$

the expression (3.25) can be obtained by some algebraic simplifications on $\pi^{+}\left(z\right) = \sum_{j=0}^{\infty} p_j^{+} z^{j}$. Meanwhile, (3.26) can be derived by using $E\left[L^{+}\right] = \frac{\mathrm{d}}{\mathrm{d}z}\left[\pi^{+}\left(z\right)\right]|_{z=1}$. □

Corollary 3.6. *In the discrete-time Geo/G/1 repairable queue with second J-optional service and D-policy, the steady-state queue size* L^{+} *can be decomposed into the sum of two independent random variables:* $L^{+} = L_0^{+} + L_d^{+}$*.* L_0^{+} *is the steady-state queue length of the Geo/G/1 queue with unreliable server and J-additional services and the corresponding p.g.f. is* $((1-\rho)(1-z)\hat{G}(z\lambda+\bar{\lambda}))/(\hat{G}(z\lambda+\bar{\lambda})-z)$*.* L_d^{+} *is the additional queue size caused by D-policy, and the distribution of* L_d^{+} *satisfies*

$$P\left\{L_d^{+}=r\right\} = \frac{Y^{(r)}(D)}{1+\sum\limits_{k=1}^{D-1} Y^{(k)}(D)}, \quad r=0,1,2,\ldots,D-1.$$

Proof. From (3.25), the steady-state queue length consists of two independent parts, and the p.g.f. of L_d^{+} is given by

$$\pi_d^{+}\left(z\right) = \frac{1+\sum\limits_{k=1}^{D-1} z^{k} Y^{(k)}(D)}{1+\sum\limits_{k=1}^{D-1} Y^{(k)}(D)}. \tag{3.27}$$

Hence the p.m.f. of additional queue size L_d^{+} can be derived by $P\left\{L_d^{+}=r\right\} = \frac{1}{r!}\cdot\frac{\mathrm{d}^{r}}{\mathrm{d}z^{r}}\left[\pi_d^{+}\left(z\right)\right]|_{z=0}$. □

Remark 3.7 (Special cases)**.** In this remark, we consider some special cases of our model by taking specific values for the parameters.

(1) When $D=1$, $\theta_0=1$, $\alpha=0$, the considered model reduces to the standard *Geo/G/1* discrete-time queueing system. For $\rho=\lambda\mu_0<1$ and $|z|<1$, we obtain

$$\pi^{+}\left(z\right) = \frac{\left(1-\rho\right)\left(1-z\right) G_0\left(\bar{\lambda}+z\lambda\right)}{G_0\left(\bar{\lambda}+z\lambda\right)-z},$$

which matches with the result in Hunter [8].

(2) When $D = 1$, $J = 1$, our model is equivalent to a discrete-time $Geo/G/1$ queue with breakdowns and second optional service. For $\rho = \lambda(\mu_0 + \theta_1\mu_1)(1 + \alpha\beta) < 1$ and $|z| < 1$, we have

$$\pi^+(z) = \frac{(1-\rho)(1-z)\,G\left(\bar{\lambda} + z\lambda\right)}{G\left(\bar{\lambda} + z\lambda\right) - z},$$

where $G\left(\bar{\lambda} + z\lambda\right) = (1-\theta_1)\,\tilde{G}_0\left(\bar{\lambda} + z\lambda\right) + \theta_1\tilde{G}_1\left(\bar{\lambda} + z\lambda\right)\tilde{G}_0\left(\bar{\lambda} + z\lambda\right)$, $\tilde{G}_0(z)$ and $\tilde{G}_1(z)$ are determined by (2.3) and (2.4), respectively.

(3) When $\theta_0 = 1$, the model investigated in this paper becomes a discrete-time $Geo/G/1$ queue with D-policy and unreliable server. For $\rho = \lambda\mu_0(1 + \alpha\beta) < 1$ and $|z| < 1$, we have

$$\pi^+(z) = \frac{(1-\rho)(1-z)\,\tilde{G}_0\left(\bar{\lambda} + z\lambda\right)}{\tilde{G}_0\left(\bar{\lambda} + z\lambda\right) - z} \cdot \frac{1 + \sum\limits_{k=1}^{D-1} z^k Y^{(k)}(D)}{1 + \sum\limits_{k=1}^{D-1} Y^{(k)}(D)},$$

where $\tilde{G}_0(z)$ is given by (2.3), $Y^{(k)}(D) = P\left\{\sum_{j=1}^{k}\chi_j^{(0)} < D\right\}$.

4. Reliability indices of the service station

In this section, some reliability indices, including the transient and the steady-state unavailability of service station, the expected number of service breakdowns during time interval $(0^+, n^+]$ and the equilibrium failure frequency of the service station, are investigated. In the interest of presenting the stochastic decomposition property of these reliability measures, we first introduce a lemma.

Lemma 4.1. *Let $H_i(n^+) = P\{$the time n^+ is in the server's generalized busy period$\,|N(0^+) = i\}$, $i \geq 0$, and the p.g.f. of $H_i(n^+)$ is $h_i(z) = \sum_{n=0}^{\infty} H_i(n^+)\,z^n$, $|z| < 1$. Then, we have*

$$h_0(z) = \frac{f(z)}{1-z}\left\{1 - \frac{B(z)\left[1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)]^k\right]}{1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)B(z)]^k}\right\}, \tag{4.1}$$

$$h_i(z) = \frac{1}{1-z}\left\{1 - \frac{B^i(z)\left[1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)]^k\right]}{1 - \sum\limits_{k=1}^{D} A_k(D)\,[f(z)B(z)]^k}\right\},\ i \geq 1. \tag{4.2}$$

And for all $i = 0, 1, 2, \ldots$,

$$H = \lim_{n\to\infty} H_i\left(n^+\right) = \begin{cases} \lambda\left(\mu_0 + \sum\limits_{i=1}^{J}\theta_i\mu_i\right)(1+\alpha\beta), & \rho < 1, \\ 1, & \rho \geq 1. \end{cases} \tag{4.3}$$

Proof. Let $S_k = \sum_{i=1}^{k} \chi_i^{(0)}$, $l_k = \sum_{i=1}^{k} \tau_i$, $k = 1, 2, \ldots$, and $S_0 = l_0 = 0$. By a similar probabilistic argument to the proof of Theorem 3.1, it follows that

$$\begin{aligned}
H_0\left(n^+\right) &= P\{\text{the time } n^+ \text{ is in the server's generalized busy period} \,\big|\, N\left(0^+\right) = 0\} \\
&= P\left\{\tilde{\tau}_1 \le n^+ < \tilde{\tau}_1 + b_1\right\} + \sum_{k=1}^{D} P\{\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{k-1} \le n^+, S_{k-1} < D \\
&\quad \le S_k, \text{the time } n^+ \text{ is in the server's generalized busy period}\} \\
&= \sum_{t=1}^{n} \lambda \bar{\lambda}^{t-1} \left[1 - P\left\{b_1 \le (n-t)^+\right\}\right] \\
&\quad + \sum_{k=1}^{D} A_k(D) \sum_{t=2}^{n} P\left\{\tilde{\tau}_1 + b_1 = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_2 + l_{k-1} = m\right\} \\
&\quad \times H_k\left((n-t-m)^+\right).
\end{aligned} \tag{4.4}$$

Using the analysis method of $H_0\left(n^+\right)$, for $i \ge 1$, we get

$$\begin{aligned}
H_i\left(n^+\right) &= P\{\text{the time } n^+ \text{ is in the server's generalized busy period} \,\big|\, N\left(0^+\right) = i\} \\
&= 1 - \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \\
&\quad + \sum_{k=1}^{D} A_k(D) \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_2 + l_{k-1} = m\right\} \\
&\quad \times H_k\left((n-t-m)^+\right).
\end{aligned} \tag{4.5}$$

Multiplying both equations (4.4) and (4.5) by z^n and adding over n from 0 to 1, it yields

$$h_0(z) = \frac{f(z)\left[1 - B(z)\right]}{1 - z} + f(z) B(z) \sum_{k=1}^{D} A_k(D) \left[f(z)\right]^k h_k(z). \tag{4.6}$$

$$h_i(z) = \frac{1 - B^i(z)}{1 - z} + B^i(z) \sum_{k=1}^{D} A_k(D) \left[f(z)\right]^k h_k(z). \tag{4.7}$$

Solving (4.6) and (4.7), the desired expressions $h_0(z)$ and $h_i(z)$ can be obtained. Finally, by Final Value Theorem (Jury [11]), it follows that $\lim_{n\to\infty} H_i\left(n^+\right) = \lim_{z\to 1^-} (1-z) \cdot h_i(z)$. By Lemma 2.6 and L'Hospital rule, we can achieve the equilibrium result given by (4.3). □

4.1. The unavailability of service station

Before discussion, we consider a classical discrete-time repairable system with a single unit. The lifetime of the unit, denoted by X, obeys a geometric distribution with parameter α. The system breaks down if and only if failures happen to the unit. As soon as the unit is subject to unpredictable breakdowns, it is immediately sent to mend. The repair time, denoted by R, has an arbitrary distribution with p.m.f. $P\{R = j\} = r_j$, $j = 0, 1, \ldots$, p.g.f. $R(z) = \sum_{j=0}^{\infty} r_j z^j$ and finite expected value β. When the unit is repaired, it renews and starts to operate immediately. Also, suppose that the unit is new at initial point $n^+ = 0^+$, and X and R are independent of each other. For $n \ge 0$, $|z| < 1$, let

$$\Psi\left(n^+\right) = P\left\{\text{the unit is under repair at time } n^+\right\}, \ \phi(z) = \sum_{n=0}^{\infty} z^n \Psi\left(n^+\right).$$

$M(n^+) = E\{\text{the number of unit failures during } (0^+, n^+]\}$, $m(z) = \sum_{n=0}^{\infty} z^n M(n^+)$.

Similar to the counterpart of continuous-time case in Cao and Cheng [4], we have the following lemma.

Lemma 4.2. *If $|z| < 1$, then we have*

$$\phi(z) = \frac{z\alpha[1 - R(z)]}{(1 - z)[1 - z + z\alpha(1 - R(z))]},$$
$$m(z) = \frac{z\alpha}{(1 - z)[1 - z + z\alpha(1 - R(z))]},$$

and the steady-state unavailability and breakdown frequency are respectively given by

$$\lim_{n\to\infty} \Psi(n^+) = \lim_{z\to 1^-} (1 - z)\cdot\phi(z) = \frac{\alpha\beta}{1 + \alpha\beta},$$
$$\lim_{n\to\infty} \frac{M(n^+)}{n} = \lim_{z\to 1^-} (1 - z)^2 m(z) = \frac{\alpha}{1 + \alpha\beta}.$$

We now set out to analyze the transient unavailability at time epoch n^+ and the steady-state unavailability $(n \to \infty)$. The unavailability of service station at time epoch n^+ is the probability that the service station is under repair at time point n^+. Let $\Psi_i(n^+) = P\{$ the unit is under repair at time $n^+ \,|N(0^+) = i\}$, $i \geq 0$, and the p.g.f. of $\Psi_i(n^+)$ is $\phi_i(z) = \sum_{n=0}^{\infty} z^n \Psi_i(n^+)$, $|z| < 1$.

Theorem 4.3. *For $|z| < 1$, $i = 0, 1, \ldots$, it follows that*

$$\phi_i(z) = \phi(z)[(1 - z)h_i(z)], \tag{4.8}$$

and the steady-state unavailability, denoted by Ψ, is

$$\Psi = \lim_{n\to\infty} \Psi_i(n^+) = \lim_{z\to 1^-} (1 - z)\cdot\phi_i(z) = \begin{cases} \lambda\left(\mu_0 + \sum\limits_{i=1}^{J} \theta_i\mu_i\right)\alpha\beta, & \rho < 1, \\ \frac{\alpha\beta}{1+\alpha\beta}, & \rho \geq 1, \end{cases} \tag{4.9}$$

where $\rho = \lambda\left(\mu_0 + \sum_{i=1}^{J} \theta_i\mu_i\right)(1 + \alpha\beta)$, $h_i(z)$ and $\phi(z)$ are determined by Lemmas 4.1 and 4.2, respectively.

Proof.

(1) Note that from the model assumptions, the failure of service station takes place only in server's generalized busy period. That is to say, the service station breaks down at time n^+ if and only if the time n^+ is located in some generalized busy period of server and the service station is under repair at time n^+. By the same analysis method as in the proof of Theorem 3.1, $\Psi_0(n^+)$ is expressed as

$$\begin{aligned}
\Psi_0(n^+) &= P\{\text{the unit is under repair at time } n^+ \,|N(0^+) = 0\} \\
&= P\{\tilde{\tau}_1 \leq n^+ < \tilde{\tau}_1 + b_1 \text{, the unit is under repair at time } n^+\} \\
&\quad + \sum_{k=1}^{D} P\{\tilde{\tau}_1 + b_1 + \tilde{\tau}_2 + l_{k-1} \leq n^+, S_{k-1} < D \leq S_k \text{, the unit is under repair at time } n^+\} \\
&= \sum_{t=1}^{n} \lambda\bar{\lambda}^{t-1}\Gamma_1\left((n - t)^+\right) \\
&\quad + \sum_{k=1}^{D} A_k(D) \sum_{t=2}^{n} P\{\tilde{\tau}_1 + b_1 = t\} \sum_{m=k}^{n-t} P\{\tilde{\tau}_2 + l_{k-1} = m\} \\
&\quad \times \Psi_k\left((n - t - m)^+\right),
\end{aligned} \tag{4.10}$$

where $\mathit{\Gamma}_i(n^+) = P\left\{0 \le n^+ < b^{\langle i\rangle}, \text{the unit is under repair at time } n^+\right\}$, $i \ge 1$. Similarly, for $i \ge 1$, $\Psi_i(n^+)$ is given by

$$\Psi_i\left(n^+\right) = \mathit{\Gamma}_i\left(n^+\right) + \sum_{k=1}^{D} A_k(D) \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \sum_{m=k}^{n-t} P\left\{\tilde{\tau}_2 + l_{k-1} = m\right\} \times \Psi_k\left(\left(n-t-m\right)^+\right). \tag{4.11}$$

(2) Since the service station behaves in renewal process alternating with failed state and normal state in server's generalized busy period and the lifetime X is governed by geometric distribution, we conduct the total probability decomposition of $\Psi(n^+)$ with $b^{\langle i\rangle}$ and get that

$$\begin{aligned}\Psi\left(n^+\right) &= P\{\text{the unit is under repair at time } n^+, b^{\langle i\rangle} > n^+ \ge 0\} \\ &\quad + P\{\text{the unit is under repair at time } n^+, b^{\langle i\rangle} \le n^+\} \\ &= \mathit{\Gamma}_i\left(n^+\right) + \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \Psi\left(\left(n-t\right)^+\right),\ i \ge 1,\end{aligned}$$

which leads to

$$\sum_{n=0}^{\infty} z^n \mathit{\Gamma}_i\left(n^+\right) = \phi(z)\left[1 - B^i(z)\right],\ |z| < 1. \tag{4.12}$$

Multiplying both (4.10) and (4.11) by z^n and adding over n from 0 to 1, and utilizing (4.12), it leads to

$$\phi_0(z) = f(z)\,\phi(z)\left[1 - B(z)\right] + f(z)B(z)\sum_{k=1}^{D} A_k(D)\left[f(z)\right]^k \phi_k(z), \tag{4.13}$$

$$\phi_i(z) = \phi(z)\left[1 - B^i(z)\right] + B^i(z)\sum_{k=1}^{D} A_k(D)\left[f(z)\right]^k \phi_k(z). \tag{4.14}$$

From the above two equations, (4.1) and (4.2), (4.8) can be obtained. Further, the stationary unavailability Ψ in (4.9) can be gained by

$$\begin{aligned}\lim_{n\to\infty} \Psi_i\left(n^+\right) = \lim_{z\to 1^-}(1-z)\,\phi_i(z) &= \lim_{z\to 1^-}(1-z)\,\phi(z) \cdot \lim_{z\to 1^-}(1-z)\,h_i(z) \\ &= \lim_{n\to\infty} \Psi\left(n^+\right) \cdot \lim_{n\to\infty} H_i\left(n^+\right),\end{aligned}$$

where $\lim_{n\to\infty} H_i(n^+)$ and $\lim_{n\to\infty} \Psi(n^+)$ are presented by Lemmas 4.1 and 4.2, respectively. □

4.2. The expected number of service station failures during $\left(0^+, n^+\right]$

In this subsection, we will study the expected number of service station breakdowns in time interval $(0^+, n^+]$. Denote by $M_i(n^+) = E\{$ the number of service station failures during $(0^+, n^+]\,|N(0^+) = i\}$ the mean of failure number of service station during $(0^+, n^+]$ under initial condition $N(0^+) = i$, $i = 0, 1, \ldots$ The p.g.f. of $M_i(n^+)$ is $m_i(z) = \sum_{n=0}^{\infty} z^n M_i(n^+)$, $|z| < 1$.

Theorem 4.4. *For $|z| < 1$, $i = 0, 1, \ldots$, we have*

$$m_i(z) = m(z) \cdot \left[(1-z)\,h_i(z)\right], \tag{4.15}$$

and the steady-state failure frequency of service station, denoted by M, is given by

$$M = \lim_{n\to\infty} \frac{M_i(n^+)}{n} = \lim_{n\to\infty} (1-z)^2 m_i(z) = \begin{cases} \lambda\left(\mu_0 + \sum\limits_{i=1}^{J} \theta_i\mu_i\right)\alpha, & \rho < 1, \\ \frac{\alpha}{1+\alpha\beta}, & \rho \geq 1, \end{cases} \tag{4.16}$$

where $\rho = \lambda\left(\mu_0 + \sum_{i=1}^{J} \theta_i\mu_i\right)(1+\alpha\beta)$, $h_i(z)$ and $m(z)$ are given by Lemmas 4.1 and 4.2, respectively.

Proof.

(1) Similar to the analysis of (4.10) and (4.11), we have the following two equations.

$$\begin{aligned} M_0(n^+) = {} & \sum_{t=1}^{n} \lambda\bar{\lambda}^{t-1}\left[W_1\left((n-t)^+\right) + T_1\left((n-t)^+\right)\right] \\ & + \sum_{k=1}^{D} A_k(D) \sum_{t=2}^{n} P\{\tilde{\tau}_1 + b_1 = t\} \sum_{m=k}^{n-t} P\{\tilde{\tau}_2 + l_{k-1} = m\} \\ & \times M_k\left((n-t-m)^+\right). \end{aligned} \tag{4.17}$$

$$\begin{aligned} M_i(n^+) = {} & W_i(n^+) + T_i(n^+) \\ & + \sum_{k=1}^{D} A_k(D) \sum_{t=i}^{n} P\left\{b^{\langle i\rangle} = t\right\} \sum_{m=k}^{n-t} P\{\tilde{\tau}_2 + l_{k-1} = m\} \\ & \times M_k\left((n-t-m)^+\right), \ i \geq 1, \end{aligned} \tag{4.18}$$

where $T_i(n^+) = \mathrm{E}\{$the number of service station failures during $\left(0^+, b^{\langle i\rangle}\right]$, $b^{\langle i\rangle} \leq n^+\}$, $W_i(n^+) = \mathrm{E}\{$the number of service station failures during $(0^+, n^+]$, $b^{\langle i\rangle} > n^+\}$, $i = 1, 2, \ldots$.

(2) Let

$$t_i(z) = \sum_{n=0}^{\infty} T_i(n^+) z^n, \ w_i(z) = \sum_{n=0}^{\infty} W_i(n^+) z^n.$$

By performing a similar discussion in (4.12), it gives

$$t_i(z) + w_i(z) = m(z)\left[1 - B^i(z)\right], \ i = 1, 2, \ldots, \tag{4.19}$$

where $m(z)$ is determined by Lemma 4.2. Multiplying (4.17) and (4.18) by z^n and summing over n, respectively, and applying (4.19), which leads to (4.15). Meanwhile, the occurrence rate of service station failures, in steady state, can be obtained by

$$\begin{aligned} \lim_{n\to\infty} \frac{M_i(n^+)}{n} = \lim_{z\to 1^-} (1-z)^2 m_i(z) &= \lim_{z\to 1^-} (1-z)^2 m(z) \cdot \lim_{z\to 1^-} (1-z) h_i(z) \\ &= \lim_{n\to\infty} \frac{M(n^+)}{n} \cdot \lim_{n\to\infty} H_i(n^+), \end{aligned}$$

where $\lim\limits_{n\to\infty} H_i(n^+)$ and $\lim\limits_{n\to\infty} \frac{M(n^+)}{n}$ are given by Lemmas 4.1 and 4.2, respectively. □

Remark 4.5. From (4.8) and (4.15), one can see that the transient behaviors of reliability indices of service station meet the stochastic decomposition property. Also, (4.3), (4.9) and (4.16) reveal that the steady-state results are independent of the initial state $N(0^+) = i$, $i \geq 0$ and D-policy.

5. Optimization for system cost

In this section, for the sake of demonstrating the applicability of the results obtained in the previous discussion, an optimization analysis is conducted from the economic point of view. We first develop an expected operating cost function per unit time for the repairable queueing model discussed in this paper, in which D is decision variable. Then, we give some numerical calculations to find the optimum D, say D^*, to minimize the long-run expected cost per unit time. Let us define the cost structure as follows.

C_h $\equiv$ holding cost per unit time for each customer present in the system (this cost originates from the customer's sojourn time that consists of the waiting time and the total generalized service time).

$C_s \equiv$ fixed cost per unit time for per busy cycle (this cost is due to the switch-over between busy period and idle period).

C_r $\equiv$ repair cost per unit time of the broken service station (this cost originates from the unpredictable breakdowns occurring in server's generalized busy period).

$C_d \equiv$ fixed cost incurred due to every failure of the service station (this cost may be from the depreciation when the breakdown occurs).

Utilizing the definitions of each cost element listed above and the corresponding system performance measures obtained previously, the total expected cost function per unit time is given by

$$\begin{aligned} C\left(D\right) &= C_h E\left[L^+\right] + C_s \frac{1}{E\left[T_m\right]} + C_r \Psi + C_d M \\ &= C_h \left\{ \rho + \frac{\lambda^2 E\left[\hat{\chi}\left(\hat{\chi}-1\right)\right]}{2\left(1-\rho\right)} + \frac{\sum\limits_{k=1}^{D-1} k Y^{(k)}(D)}{1+\sum\limits_{k=1}^{D-1} Y^{(k)}(D)} \right\} + C_s \frac{\lambda\left(1-\rho\right)}{\sum\limits_{k=0}^{D-1} Y^{(k)}(D)} \\ &\quad + C_r \lambda \left(\mu_0 + \sum_{i=1}^{J} \theta_i \mu_i \right) \alpha\beta + C_d \lambda \left(\mu_0 + \sum_{i=1}^{J} \theta_i \mu_i \right) \alpha, \;\; \rho < 1. \end{aligned} \tag{5.1}$$

From (5.1) we can see that the cost function is extremely complex and non-linear, which poses a hard task to achieve the analytic results for the optimum value of D. Therefore, we will search the optimum value D^* for $C\left(D\right)$ through numerical examples.

Example 5.1. We consider a practical problem concerning the aforementioned pump manufacturing system. Assume that the shafts of various dimensions arrive at CNC copy turning center according to a Bernoulli process with parameter λ. From an economic point of view, it is desirable that the CNC machine begins to deal with the shafts whenever the total service times of the waiting shafts reach a fixed value D. All the arriving shafts require the first essential service (FES) that is a random variable $\chi^{(0)}$ with a mean μ_0. After the first essential service, some shafts are of excellent quality, whereas some processed shafts may be defective. The imperfect shafts may be one of the summarized J defective types and need to be reprocessed (re-served) to meet the required specification. The defective shafts belonging to the ith $(1 \leq i \leq J)$ type will demand a second service $\chi^{(i)}$ with a mean μ_i. Furthermore, the lifetime of the CNC machine has a random length X following a geometric distribution with parameter α, which indicates that the CNC machine may break down with probability α in a slot. Once the CNC machine is subject to failure, it emergently is sent to repair. The repair time R is a random variable.

The production system can be modeled by the queueing system investigated in this paper. For convenience of computations, the corresponding parameters are given as follows.

(1) The arrival rate of shafts is $\lambda = 0.15$.
(2) The FES $\chi^{(0)}$ is geometrically distributed with parameter $\sigma_0 = 0.6$.

TABLE 1. The average cost per unit time for different values of D.

D	$C(D)$	D	$C(D)$	D	$C(D)$	D	$C(D)$	D	$C(D)$
1	34.2770	9	32.2706	17	34.2383	25	36.4741	33	38.7877
2	32.6070	10	32.4815	18	34.5102	26	36.7605	34	39.0796
3	31.9998	11	32.7075	19	34.7848	27	37.0479	35	39.3721
4	31.7765	12	32.9451	20	35.0620	28	37.3361	36	39.6649
5	31.7355	13	33.1918	21	35.3412	29	37.6251	37	39.9581
6	31.7956	14	33.4456	22	35.6222	30	37.9148	38	40.2517
7	31.9174	15	33.7053	23	35.9048	31	38.2052	39	40.5456
8	32.0799	16	33.9698	24	36.1889	32	38.4962	40	40.8398

(3) It is assumed that there are two summarized defective types (*i.e.*, $J = 2$), and the flawed shafts belonging to the ith type receive a second service $\chi^{(i)}$ following geometric distribution with parameter $\sigma_i (i = 1, 2)$. We take $\sigma_1 = 0.75$, $\sigma_2 = 0.8$.
(4) The repair time R of the broken service station obeys geometric distribution with parameter $\eta = 0.67$.
(5) Other variables are selected as $\alpha = 0.25$, $\theta_0 = 0.65$, $\theta_1 = 0.2$, $\theta_2 = 0.15$, $C_h = 1$, $C_s = 60$, $C_r = 100$, $C_d = 210$.

Substituting these parameters into (5.1) and developing MATLAB program, the numerical results for different value of D are reported in Table 1. With the information of Table 1, one can observe that the minimum long-run expected operating cost per unit time is $C(D) = 31.7355$ at $D = 5$, which indicates that the system designers should design the workload threshold D as 5.

Example 5.2. In this example, the sensitivity analyses of $C(D)$ and D with respect to different system parameters and cost elements will be carried out, which can provide significant insight for decision makers to make the system profitable. Suppose that there are FES and two additionally optional services for customers, and the corresponding distributions are governed by geometric distributions with parameters σ_0, σ_1, and σ_2, respectively. The repair time of the broken service facility has geometric distribution with parameter η. Other parameters and notations are identical to those given in previous sections. The following five cases are considered in this example.

Case 1. $\sigma_0 = 0.9$, $\sigma_1 = 0.85$, $\sigma_2 = 0.88$, $\alpha = 0.1$, $\eta = 0.8$, $\theta_0 = 0.86$, $\theta_1 = 0.13$, $\theta_2 = 0.01$, $C_h = 10$, $C_s = 300$, $C_r = 550$ and $C_d = 660$ for different values of λ (see Tab. 2).
Case 2. $\lambda = 0.25$, $\alpha = 0.1$, $\eta = 0.8$, $\theta_0 = 0.86$, $\theta_1 = 0.13$, $\theta_2 = 0.01$, $C_h = 10$, $C_s = 300$, $C_r = 550$ and $C_d = 660$ for different values of σ_0, σ_1 and σ_2 (see Tab. 3).
Case 3. $\lambda = 0.25$, $\sigma_0 = 0.9$, $\sigma_1 = 0.85$, $\sigma_2 = 0.88$, $\theta_0 = 0.86$, $\theta_1 = 0.13$, $\theta_2 = 0.01$, $C_h = 10$, $C_s = 300$, $C_r = 550$ and $C_d = 660$ for different values of α and η (see Table 4).
Case 4. $\lambda = 0.25$, $\sigma_0 = 0.9$, $\sigma_1 = 0.85$, $\sigma_2 = 0.88$, $\alpha = 0.1$, $\eta = 0.8$, $\theta_0 = 0.86$, $\theta_1 = 0.13$, $\theta_2 = 0.01$, $C_s = 300$ and $C_d = 660$ for different values of C_h and C_r (see Tab. 5).
Case 5. $\lambda = 0.25$, $\sigma_0 = 0.9$, $\sigma_1 = 0.85$, $\sigma_2 = 0.88$, $\alpha = 0.1$, $\eta = 0.8$, $\theta_0 = 0.86$, $\theta_1 = 0.13$, $\theta_2 = 0.01$, $C_h = 10$ and $C_r = 550$ for different values of C_s and C_d (see Tab. 6).

Tables 2–4 display the effect of various system parameters on the optimum threshold value D^* and its corresponding minimum average cost $C(D^*)$. From Table 2, it is observed that both $C(D^*)$ and D^* increase as arrival rate λ increases. This is due to the fact that with the growth of λ, the number of customers in the system becomes larger and therefore the total holding cost increases. One can see from Table 3 that $C(D^*)$ decreases with an increasing σ_0, σ_1 or σ_2 but D* increases as σ_0 or σ_1 increases. Actually, the increment of service rate will lead to the decrease of average system size, which in turn reduces the holding cost of customers. Table 4 reveals that there is an increase in $C(D^*)$ but a decrease in D^* as the failure rate α increases. The reverse trend is shown with the increase of repair rate η. Such situations match with many real life congestion scenarios.

TABLE 2. The optimal threshold value D^* and its minimum average cost $C(D^*)$ for different values of λ.

λ	ρ	D^*	$C(D^*)$
0.10	0.1435	3	34.7392
0.15	0.2152	3	47.5912
0.20	0.2870	3	59.7395
0.25	0.3587	4	71.0741
0.30	0.4305	4	81.6169
0.35	0.5022	4	91.7493
0.40	0.5739	4	101.5580
0.45	0.6457	4	111.1997

TABLE 3. The optimal threshold value D^* and its minimum average cost $C(D^*)$ for different values of σ_0, σ_1, and σ_2.

σ_0	$\sigma_1 = 0.75$, $\sigma_2 = 0.8$			σ_1	$\sigma_0 = 0.79$, $\sigma_2 = 0.8$			σ_2	$\sigma_0 = 0.79$, $\sigma_1 = 0.75$		
	ρ	D^*	$C(D^*)$		ρ	D^*	$C(D^*)$		ρ	D^*	$C(D^*)$
0.33	0.9054	2	183.6507	0.26	0.5002	3	89.8115	0.2	0.4188	4	78.3768
0.39	0.7734	3	130.3424	0.32	0.4738	4	85.8590	0.3	0.4141	4	77.6939
0.45	0.6773	4	112.5890	0.38	0.4557	4	83.2393	0.4	0.4118	4	77.3816
0.51	0.6037	4	101.7851	0.44	0.4426	4	81.4181	0.5	0.4104	4	77.2034
0.57	0.5457	4	94.0253	0.50	0.4327	4	80.0779	0.6	0.4095	4	77.0883
0.63	0.4987	4	88.0206	0.56	0.4248	4	79.0501	0.7	0.4088	4	77.0080
0.69	0.4599	4	83.2066	0.62	0.4185	4	78.2367	0.8	0.4083	4	76.9488
0.75	0.4273	4	79.2320	0.68	0.4133	4	77.5768	0.9	0.4079	4	76.9033

TABLE 4. The optimal threshold value D^* and its minimum average cost $C(D^*)$ for different values of α and η.

α	$\eta = 0.8$			η	$\alpha = 0.1$		
	ρ	D^*	$C(D^*)$		ρ	D^*	$C(D^*)$
0.15	0.3786	4	92.5206	0.27	0.4369	3	114.9561
0.20	0.3986	3	113.8532	0.34	0.4126	3	101.1224
0.25	0.4185	3	135.1991	0.41	0.3966	3	92.1368
0.30	0.4384	3	156.5619	0.48	0.3853	3	85.8268
0.35	0.4584	3	177.9435	0.55	0.3768	4	81.1369
0.40	0.4783	3	199.3461	0.62	0.3703	4	77.4898
0.45	0.4982	3	220.7721	0.69	0.3651	4	74.5921
0.50	0.5181	3	242.2245	0.76	0.3608	4	72.2342

In fact, the higher the failure rate is, the more repair cost the system managers have to be charged. And under a relatively higher repair rate, the waiting customers have a greater chance to be served, which can reduce the system queue length and the system cost.

Tables 5 and 6 exhibit the influence of the cost elements C_h, C_r, C_s and C_d on $C(D^*)$ and D^*. We can see from Tables 5 and 6 that $C(D^*)$ increases with an increase in C_h, C_r, C_s, or C_d while D^* decreases as C_h increases, which is quite obvious. In addition, it is noted that D^* is insensitive to C_r and C_d. This is because the failure and repair of the system are not related to the threshold D.

From the numerical results and the corresponding analysis, we overall conclude the following observations which may provide some insight for system designers and decision makers so as to help them model real time system.

TABLE 5. The optimal threshold value D^* and its minimum average cost $C(D^*)$ for different values of C_h and C_r.

C_h	$C_r = 500$			C_r	$C_h = 10$		
	ρ	D^*	$C(D^*)$		ρ	D^*	$C(D^*)$
5	0.3587	5	61.0794	100	0.3587	4	53.1386
10	0.3587	4	69.0813	300	0.3587	4	61.1099
15	0.3587	3	74.7446	500	0.3587	4	69.0813
20	0.3587	3	80.2761	700	0.3587	4	77.0526
25	0.3587	2	84.5318	900	0.3587	4	85.0240
30	0.3587	2	88.1808	1100	0.3587	4	92.9953
35	0.3587	2	91.8298	1300	0.3587	4	100.9667
40	0.3587	2	95.4788	1500	0.3587	4	108.9380

TABLE 6. The optimal threshold value D^* and its minimum average cost $C(D^*)$ for different values of C_s and C_d.

C_s	$C_d = 600$			C_d	$C_s = 300$		
	ρ	D^*	$C(D^*)$		ρ	D^*	$C(D^*)$
200	0.3587	3	63.4873	450	0.3587	4	62.3854
240	0.3587	3	65.7776	500	0.3587	4	63.9796
280	0.3587	3	68.0679	550	0.3587	4	65.5739
320	0.3587	4	69.9479	600	0.3587	4	67.1682
360	0.3587	4	71.6811	650	0.3587	4	68.7624
400	0.3587	4	73.4143	700	0.3587	4	70.3567
440	0.3587	4	75.1475	750	0.3587	4	71.9510
480	0.3587	4	76.8807	800	0.3587	4	73.5452

- The system designers may introduce an admission control policy to regulate the arrival rate. Thus, the average system size can be controlled to economize the system cost.
- By increasing the service rate, repair rate, or reducing the failure rate, the system designers can improve the quality of service at reasonable cost.
- It is also beneficial for system managers to lower properly the values of cost elements C_h, C_r, C_s and C_d.

6. Conclusions

This paper explored a discrete-time $Geo/G/1$ queue with J-optional services, D-policy, and unreliable service station. Employing the total probability decomposition law, renewal theory and probability generating function technique, the analytical expressions for transient queue length distribution and steady-state queue length distribution were derived. Especially, the recursive formulas of the steady-state queue length distribution (given by Thm. 3.4) are computationally tractable to handle the congestion problems in real-life scenarios. Furthermore, some main reliability indices were discussed, which may be useful for system designers and engineering managers to improve the reliability and availability of systems. Finally, a cost model is established to investigate cost optimization problem for the system and some numerical examples are conducted to show the effect of different system parameters and cost elements on the optimum value of D and its corresponding minimum expected cost per unit time. The analysis of this paper provides a potentially practical application in telecommunication systems, queueing networks, flexible manufacturing systems, inventory problems and so forth. For future research, using the analytical technique in our discussion, one can extend the present work to more complex queueing models by taking the concepts of bulk arrival and server vacations.

Acknowledgements. The authors would like to sincerely thank the anonymous reviewers and the editor for their valuable comments and constructive suggestions which are very helpful in improving the presentation of this paper.

References

[1] I. Atencia and P. Moreno, $Geo/G/1$ retrial queue with 2nd optional service. *Int. J. Oper. Res.* **1** (4) (2006) 340−362.

[2] I. Atencia and P. Moreno, A discrete-time $Geo/G/1$ retrial queue with server breakdowns. *Asia-Pac. J. Oper. Res.* **23** (2006) 247−271.

[3] H. Bruneel and B.G. Kim, *Discrete-Time Models for Communication Systems including ATM*. Kluwer Academic Publishers, Boston (1993).

[4] J.H. Cao and K. Cheng, *Introduction to Reliability Mathematics*. Higher Education Press, Beijing (2006).

[5] G. Choudhury, An $M/G/1$ queueing system with two phase service under D-Policy. *Int. J. Inf. Manag. Sci.* **16** (2005) 1−17.

[6] S. Gao and Z. Liu, A repairable $Geo^X/G/1$ retrial queue with Bernoulli feedback and impatient customers. *Acta Math. Appl. Sin., English Series* **30** (2014) 205−222.

[7] S. Gao, Z. Liu and H. Dong, A repairable discrete-time retrial queue with recurrent customers, Bernoulli feedback and general retrial times. *Oper. Res.* **12** (2012) 367−383.

[8] J.J. Hunter, *Mathematical Techniques of Applied Probability*. Vol. 2: *Discrete Time Models: Techniques and Applications*. Academic Press, New York (1983).

[9] M. Jain and S. Upadhyaya, Optimal repairable $M^x/G/1$ queue with multi-optional services and Bernoulli vacation. *Int. J. Oper. Res.* **7** (2010) 109−132.

[10] M. Jain, G.C. Sharma and R. Sharma, Unreliable server $M/G/1$ queue with multi-optional services and multi-optional vacations. *Int. J. Math. Oper. Res.* **5** (2013) 145−169.

[11] E.I. Jury, *Theory and Application of the Z-Transform Method*. John Wiley & Sons, New York (1964).

[12] J.C. Ke, An $M^{[x]}/G/1$ system with startup server and J additional options for service. *Appl. Math. Model.* **32** (2008) 443−458.

[13] M.S. Kumar, A discrete-time $Geo^{[X]}/G/1$ retrial queue with general retrial time and M-additional options for service. *RAIRO: OR* **45** (2011) 131−152.

[14] J. Li and J. Wang, An $M/G/1$ retrial queue with second multi-optional service, feedback and unreliable server. *Appl. Math.-A J. Chin. Univ.* **21** (2006) 252−262.

[15] C.H. Lin and J.C. Ke, On the discrete-time system with server breakdowns: Computational algorithm and optimization algorithm. *Appl. Math. Comput.* **218** (2011) 3624−3634.

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

[17] Y. Tang, X. Yun and S. Huang, Discrete-time $Geo^X/G/1$ queue with unreliable server and multiple adaptive delayed vacations. *J. Comput. Appl. Math.* **220** (2008) 439−455.

[18] J. Wang and P. Zhang, A discrete-time retrial queue with negative customers and unreliable server. *Comput. Ind. Eng.* **56** (2009) 1216−1222.

[19] J. Wang and Q. Zhao, A discrete-time $Geo/G/1$ retrial queue with starting failures and second optional service. *Comput. Math. Appl.* **53** (2007) 115-127.

[20] K.H. Wang, C.C. Kuo and J.C. Ke, Optimal control of the D-policy $M/G/1$ queueing system with server breakdowns. *Amer. J. Appl. Sci.* **5** (2008) 565−573.

[21] T.Y. Wang, An unreliable $Geo/G/1$ queue with startup and closedown times under randomized finite vacations. *Appl. Math. Model.* **39** (2015) 1383−1399.

[22] T.Y. Wang, J.C. Ke and F.M. Chang, Analysis of a discrete-time queue with server subject to vacations and breakdowns. *J. Ind. Prod. Eng.* **30** (2013) 54−66.

[23] Y. Wei, M. Yu, Y. Tang and J.Gu, Queue size distribution and capacity optimum design for N-policy $Geo^{\lambda_1,\lambda_2,\lambda_3}/G/1$ queue with setup time and variable input rate. *Math. Comput. Model.* **57** (2013) 1559−1571.

[24] R.W. Wolff, *Stochastic Modeling and the Theory of Queues*. Prentice Hall. Inc., New Jersey (1989).

[25] M.E. Woodward, *Communication and Computer Networks: Modelling with Discrete-Time Queues*. IEEE Computer Society Press, Los Alamitos (1994).

[26] F. Zhang and Z. Zhu, A discrete-time unreliable $Geo/G/1$ retrial queue with balking customers, second optional service, and general retrial times. *Math. Probl. Eng.* **2013** (2013) 1−12.