



# Optimal Control Policy of M/G/1 Queueing System with Delayed Randomized Multiple Vacations Under the Modified $\text{Min}(N, D)$ -Policy Control

Le Luo<sup>1,2</sup> · Ying-Hui Tang<sup>1</sup> · Miao-Miao Yu<sup>1</sup> · Wen-Qing Wu<sup>3</sup>

Received: 29 June 2021 / Revised: 27 February 2022 / Accepted: 28 February 2022 /

Published online: 18 April 2022

© Operations Research Society of China, Periodicals Agency of Shanghai University, Science Press, and Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Based on the number of customers and the server's workload, this paper proposes a modified  $\text{Min}(N, D)$ -policy and discusses an M/G/1 queueing model with delayed randomized multiple vacations under such a policy. Applying the well-known stochastic decomposition property of the steady-state queue size, the probability generating function of the steady-state queue length distribution is obtained. Moreover, the explicit expressions of the expected queue length and the additional queue length distribution are derived by some algebraic manipulations. Finally, employing the renewal reward theorem, the explicit expression of the long-run expected cost per unit time is given. Furthermore, we analyze the optimal policy for economizing the expected cost and compare the optimal  $\text{Min}(N, D)$ -policy with the optimal  $N$ -policy and the optimal  $D$ -policy by using numerical examples.

**Keywords** M/G/1 queue · Modified  $\text{Min}(N, D)$ -policy · Randomized multiple vacations · Queue length generating function · Optimal joint control policy

This work was supported by the National Natural Science Foundation of China (No. 71571127) and the National Natural Science Youth Foundation of China (No. 72001181).

✉ Ying-Hui Tang  
tangyh@sicnu.edu.cn

Le Luo  
18990722176@163.com

Miao-Miao Yu  
mmyu75@163.com

Wen-Qing Wu  
wwqing0704@163.com

<sup>1</sup> School of Mathematical Sciences, Sichuan Normal University, Chengdu 610068, Sichuan, China

<sup>2</sup> Nanchong Vocational and Technical College, Nanchong 637000, Jiangsu, China

<sup>3</sup> Visual Computing and Virtual Reality Key Laboratory of Sichuan Province, Sichuan Normal University, Chengdu 610068, Sichuan, China

## 1 Introduction

It is known that different control policies are useful to control the queue size and effectively reduce the customers' waiting time. If the number of arrivals during the server vacation is too large, system congestion may occur, which will degrade the quality of service and the customers' satisfaction. In addition, a queueing system without a control policy will generate a large amount of switching costs if the system frequently switches its state for a long time. Therefore, in order to effectively control the queue size and overcome the cost problem caused by the frequent switching, the investigation concerning the queueing system with a control policy is worth doing. The earliest works on the control policy of queueing systems were the  $N$ -policy proposed by Yadin and Naor [1], the  $T$ -policy presented by Heyman [2], and the  $D$ -policy developed by Balachandran [3, 4]. In some  $M/G/1$  queues, the  $N$ -policy and  $D$ -policy operate according to exhaustive service rule, i.e., the server is turned off when the system becomes empty. The  $N$ -policy means that the server resumes its service when there are  $N$  customers present in the system. In contrast, under the conventional  $D$ -policy, the server starts to serve only when the sum of the service times of all waiting customers firstly exceeds a predetermined threshold  $D$  ( $D \geq 0$ ). We know that the stochastic decomposition property proposed in reference [5] cannot be applied to the classical  $D$ -policy queue. Thus, the research on  $D$ -policy queue has always been a very challenging task in queueing theory. Authors like Dshalalow [6], Artalejo [7, 8] and Lee et al. [9, 10] did some pioneering works in studying the queues with the conventional  $D$ -policy. On the other hand, the absence of the server or non-availability of the service in the system can be termed as vacation. Queues with vacations have been extensively investigated by many researchers. Doshi [11] provided an exhaustive survey of such work through 1985. Here, we refer the readers to the monograph by Tian [12] for more details. In the study of vacation models, most works were concentrated on the study of models with multiple vacations and single vacation. However, the randomized multiple vacation policy (also called multiple adaptive vacation policy, first proposed by Tian [13]) is more general than multiple and single vacation policies.

From the engineering point of view, employing a control involves installing some instruments, including precision counting and scanning gauges. When the manufacturing environment changes and the system owner wants to switch to another control policy, it is impossible, in many cases, to discard the existing hardware system. Therefore, with the fast development of the information era, there has been an increasing interest in investigating some queueing systems with the joint control policy. For example, Lee and Seo [14] studied about the  $M/G/1$  queue with the dyadic  $\text{Min}(N, D)$ -policy in which the idle server resumes its service if either  $N$  customers accumulate in the system or the sum of the service times of the waiting customers exceeds  $D$ , whichever occurs first. Lee et al. [15] extended it to the  $\text{MAP}/G/1$  queueing system. Until now, an enormous number of works on the joint control policy and vacation queues have been published. Some more extensive and extended research works on

continuous-time and discrete-time queues with the threshold policies and vacation mechanism can refer to the references [16–31].

On the one hand, the meaning of the conventional  $D$ -policy above is “the server resumes its service when the cumulative service times of the waiting customers firstly reach or exceed  $D$ ”. However, in many practical applications, it is very difficult for the server to run  $D$ -policy designed according to the cumulative service times of the waiting customers in the system, because before each customer’s service is completed, the service time of the customer is randomly unknown. On the other hand, there are many queueing systems with vacation policies, the server cannot leave for vacation immediately after the service is completed and the system becomes empty, he/she has to go through a period of delayed time to prepare for the vacation. Sometimes this delayed period is very necessary, such as bank staff need to sort out accounts before going off work or leaving. In addition, the length of the server’s vacation time depends not only on the amount of auxiliary works which the server is engaged in during the vacation, but also on the constraints of the customers’ arrival process. Consequently, the number of consecutive vacations is often uncertain. Multiple adaptive vacation schedule is a more flexible vacation policy and is more general than most classical vacation policies in the sense that the well known multiple vacation policy and single vacation policy become two special cases of this policy. Under this framework, we can investigate many different vacation policies between these two extreme cases for the purpose of better allocation of the server’s time to perform primary jobs (serving queue) and to do secondary jobs (vacations). Thus, inspired by the facts mentioned above, we will propose a new continuous-time M/G/1 queueing model with delayed randomized multiple vacations under the modified  $\text{Min}(N, D)$ -policy, in which the idle server resumes its service if either  $N$  customers accumulate in the system or the accumulative workload (not the cumulative service times) of the waiting customers in the system exceeds  $D$ , whichever occurs first. In our model, the states of the server can be in the working state, the idle state (on duty) and the vacation state (off duty). Utilizing the stochastic decomposition theorem of the steady-state queue size in [5, 12], we present the stochastic decomposition structure of steady-state queue size. Moreover, the explicit expressions of the expected queue length and the additional queue-length distribution are derived. By numerical examples, we investigate the long-run expected cost per unit time of the system and discuss the optimal  $\text{Min}(N, D)$ -policy, the optimal  $N$ -policy and the optimal  $D$ -policy, respectively. The comparison of the minimum costs and the optimal threshold values for three different policies is presented. When the number of consecutive vacations is a fixed positive integer  $J$ , the optimal three-dimensional control policy  $(N^*, D^*, J^*)$  is also given.

The remainder of this paper is organized as follows. In Sect. 2, the details assumptions of the model are given. In Sect. 3, we firstly introduce several definitions and lemmas, and then obtain the probability generating function of the steady-state queue-length distribution by applying the well-known stochastic decomposition property of the steady-state queue size [5, 12]. Moreover, the explicit expressions of the expected queue size and the additional queue-length distribution are derived by some algebraic manipulations. Section 4 is devoted to obtain a long-run expected cost function to discuss a cost optimization problem, and numerically find the optimal control policy for economizing the system cost. At last, the conclusions are presented in Sect. 5.

## 2 Model Formulation

Considering a continuous-time M/G/1 queueing model with delayed randomized multiple vacations under the modified  $\text{Min}(N, D)$ -policy, the detailed assumptions of the model are given as follows:

- (1) The inter-arrival time  $\tau_n (n \geq 1)$  is exponentially distributed with cumulative distribution function  $F(t) = 1 - e^{-\lambda t}$ ,  $t \geq 0, \lambda > 0$ .
- (2) The workload of the server for the  $n$ th customer, denoted by  $W_n (n \geq 1)$ , refers to the quantity of events included in the completed service items required by the customer and follows an arbitrary distribution  $W(x)$ ,  $x \geq 0$ .
- (3) The service time required to complete the workload  $W_n$ , that is, the service time of the  $n$ th customer, denoted by  $\chi_n (n \geq 1)$ , is arbitrarily distributed with  $G(t)$  which is supposed to have finite mean  $1/\mu$ .
- (4) The queueing system operates under the modified  $\text{Min}(N, D)$ -policy based on delayed randomized multiple vacations, that is, once the system becomes empty, the server goes through a delayed period  $Y$  with arbitrary distribution  $Y(t)$  before taking vacation. During the delayed period  $Y$ , if there are customers arriving in the system, the server starts its service at once until the system becomes empty and the server restarts a delayed period  $Y$ . Otherwise, the server immediately takes a vacation with a random length  $V$  which obeys arbitrary distribution  $V(t)$ . Limited by the amount of current auxiliary work, the server is required to take a random maximum number of vacations (denoted by  $H$ ) after the system becomes empty, and  $H$  obeys a generally discrete distribution with

$$P\{H = l\} = h_l, l = 1, 2, 3, \dots$$

and probability generating function (p.g.f)  $H(z) = \sum_{l=1}^{\infty} h_l z^l$ ,  $|z| < 1$ . In addition, for a certain positive integer  $k (1 \leq k \leq H)$ , due to the vacation interruption mechanism, if the number of customers arriving in the system reaches  $N (N \geq 1)$  or the total workload of the server for all the waiting customers is not less than a given threshold  $D (D \geq 0)$ , whichever occurs first, the server immediately interrupts the vacation and returns to system to serve. During the  $k$ th vacation, if there are less than  $N$  customers in the system as well as the total workload of the server for all the waiting customers is less than  $D$ , the server remains vacation state until the  $k$ th vacation ends and begins to serve at once. If no customer arrives in the system during the  $k$ th vacation, the server will prepare for  $(k + 1)$ th vacation. This pattern repeats until the  $H$ th vacation expires. If the total  $H$  vacations have been finished, the server will go back to system to stay idle and wait for the next arrival. Briefly speaking, there are four cases of starting a new server busy period after the system becomes empty:

- (a) Starting a new server busy period during a delayed period (arrival occurs during a delayed period);
- (b) Starting a new server busy period during a vacation (arrival occurs during vacation period and the queue size reaches  $N$  or the total workload of the server for all the waiting customers is not less than a given threshold  $D$ );

- (c) Starting a new server busy period at the completion instant of a vacation (arrival occurs during vacation period);
  - (d) Starting a new server busy period during idle period (no arrival occurs during the given  $H$  vacations).
- (5) The inter-arrival time  $\tau$ , the workload  $W$  of the server for each customer, the service time  $\chi$ , the server's continuous vacation number  $H$ , the delayed period  $Y$  and the server vacation time  $V$  are all independent of each other.

**Remarks** (1) The workload of the server for each customer refers to the quantity of events included in the completed service items required by the customer. The unit of measurement for the workload may be a counting unit, a weight unit, etc. For example, if a customer needs a certain factory to process 100 products, the workload of this factory for this customer is 100 products. If a truck arrives at a warehouse to load 3 000 kg(3 tons) of materials, the workload of the server for the truck is 3 000 kg(3 tons).

- (2) In practical applications, the modified  $D$ -policy proposed in this paper is very easy for the server to run the  $D$ -policy designed by the cumulative workload of the server for the waiting customers in the system, because before starting the service, the server can easily know the workload of the server for the waiting customers in the system according to the customers' service items.
- (3) The queueing model studied in this paper generalizes the queueing models that have been studied in references [13, 17–19, 27]. For example, when  $P\{Y = \infty\} = 1$ , the queueing model considered in this paper is equivalent to the classic queueing model studied in [32]. When  $P\{Y = 0\} = 1$  and  $N = 1$ , it is equivalent to the queueing model considered in [12, 13]. When  $P\{Y = 0\} = 1$  and  $D \rightarrow \infty$ , the considered queueing model in this paper can be reduced to the queueing model investigated in [27].
- (4) For later use, the following notations are adopted throughout this paper:  $g^*(s) = \int_0^\infty e^{-st} G(t) dt$  denotes the Laplace transform of the corresponding distribution  $G(t)$ ,  $g(s) = \int_0^\infty e^{-st} dG(t)$  denotes the Laplace–Stieltjes transform of the corresponding distribution  $G(t)$ ,  $G^{(k)}(t)$  denotes the  $k$ -fold convolution of  $G(t)$ , i.e.,  $G^{(k)}(t) = \int_0^t G^{(k-1)}(t-x) dG(x)$ ,  $k \geq 1$  and  $G^{(0)}(t) = 1$ ,  $\rho = \lambda/\mu$  denotes the traffic intensity of the system,  $\Re(s)$  denotes the real part of the complex variable  $s$ .

### 3 The Stochastic Decomposition of the Stationary Queue Size

In this section, applying the well-known stochastic decomposition property of the steady-state queue size [5, 12], the probability generating function of the steady-state queue-length distribution is obtained. Moreover, the explicit expressions of the expected queue size and the additional queue-length distribution are derived by some algebraic manipulations. For later discussions, some definitions and lemmas will be presented as follows.

**Definition 1** (*Server busy period*) It is the time interval when the server starts to serve customers until the system becomes empty again. Thus, the server busy period here is equivalent to the system busy period of the standard M/G/1 queueing system in [32].

Denote by  $b$ , the length of the server busy period that starts from only one customer, and let  $B(t) = P\{b \leq t\}, t \geq 0, b(s) = \int_0^\infty e^{-st} dB(t)$ , we have the following lemma.

**Lemma 1** (see [32]) For  $\Re(s) > 0, b(s)$  is the unique solution of the equation  $z = g(s + \lambda - \lambda z)$  in  $|z| < 1$ , and  $B(t)$  can be expressed as:

$$B(t) = \sum_{k=1}^\infty \int_0^t \frac{(\lambda x)^{k-1}}{k!} e^{-\lambda x} dG^{(k)}(x), \quad t \geq 0,$$

and

$$\lim_{t \rightarrow \infty} B(t) = \lim_{s \rightarrow 0^+} b(s) = \begin{cases} 1, & \rho \leq 1 \\ \omega, & \rho > 1 \end{cases}, \quad E[b] = \begin{cases} \frac{\rho}{\lambda(1-\rho)}, & \rho < 1, \\ \infty, & \rho \geq 1, \end{cases}$$

where  $\omega(0 < \omega < 1)$  is the root of the equation  $z = g(\lambda - \lambda z)$  in  $(0,1)$ .

Let  $b^{<i>}$  be the length of the server busy period evoked by  $i(i \geq 1)$  customers. Because the arrival process is a Poisson process, the probability distribution function of  $b^{<i>}$  is given by

$$P\{b^{<i>} \leq t\} = B^{(i)}(t), \quad t \geq 0, i \geq 1.$$

**Definition 2** (*Server non-busy period*) It is the time period from the moment that the system just becomes empty until the moment when the server returns to the system after vacation and starts to serve customers.

**Definition 3** (*System idle period*) It is a period of time during which the system is continuously idle (no customers). Obviously, the system idle period is the remaining time of an arrival interval. Let  $\hat{\tau}_j$  represent the length of the  $j$ th system idle period, then  $\{\hat{\tau}_j, j \geq 1\}$  are independent identically distributed random variables each with distribution  $F(t) = 1 - e^{-\lambda t}, t \geq 0$ .

**Definition 4** (*System busy period*) It is the time interval that starts at the instant at which the first customer arriving at the idle system and ends at the instant when the system becomes empty again.

**Definition 5** (*Busy cycle*) It is the time interval that from the moment that the server non-busy period just begins until the moment that the next server busy period ends.

**Theorem 1** Let  $P(z)$  be the p.g.f. of steady-state queue-length distribution for the queueing system studied in this paper, then for  $\rho < 1$  and  $|z| < 1$ , we have the stochastic decomposition of the steady-state queue-length as follows:

$$P(z) = \frac{(1 - \rho)(1 - z)g(\lambda(1 - z))}{g(\lambda(1 - z)) - z}$$

$$\frac{1 - v(\lambda) + y(\lambda)[1 - H(v(\lambda))] \sum_{k=1}^{N-1} z^k \int_0^\infty W^{(k)}(D)F^{(k+1)}(t)dV(t)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}, \tag{1}$$

and the average steady-state queue size, denoted by  $E[L]$ , is presented by

$$E[L] = \rho + \frac{\lambda^2 E[\chi^2]}{2(1 - \rho)} + \frac{y(\lambda)[1 - H(v(\lambda))] \sum_{k=1}^{N-1} k \int_0^\infty W^{(k)}(D)F^{(k+1)}(t)dV(t)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}, \tag{2}$$

where  $\Delta_N = \sum_{m=1}^N W^{(m-1)}(D) \int_0^\infty F^{(m)}(t)dV(t)$ ,  $E[\chi^2] = \int_0^\infty t^2 dG(t)$ .

**Proof** See Appendix A.

**Theorem 2** For  $\rho < 1$ , the probability distribution of the additional queue size  $L_1$  is presented by

$$P\{L_1 = 0\} = \frac{1 - v(\lambda)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}, \tag{3}$$

$$P\{L_1 = j\} = \frac{y(\lambda)[1 - H(v(\lambda))] \int_0^\infty W^{(j)}(D)F^{(j+1)}(t)dV(t)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}, \tag{4}$$

$j = 1, 2, \dots, N - 1.$

**Proof** Let  $I(z) = 1 - v(\lambda) + y(\lambda)[1 - H(v(\lambda))] \sum_{k=1}^{N-1} z^k \int_0^\infty W^{(k)}(D)F^{(k+1)}(t)dV(t)$ , then the p.g.f. of the additional queue size distribution can be rewritten as:

$$P_{L_1}(z) = \frac{I(z)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}.$$

Through a simple calculation, we can obtain the following results:

$$I^{(0)}(0) = 1 - v(\lambda),$$

$$I^{(j)}(0) = (j!)y(\lambda)[1 - H(v(\lambda))] \int_0^\infty W^{(j)}(D)F^{(j+1)}(t)dV(t),$$

$j = 1, 2, \dots, N - 1.$

Thus, employing  $P\{L_1 = j\} = \frac{1}{j!} \cdot \frac{d^j}{dz^j} [P_{L_1}(z)]|_{z=0}$ , we can get Eqs. (3)–(4).

### 4 The Optimal Control Policy Under the Cost Model

In practice, the operating cost of system is closely related to the system benefit. Therefore, from the perspective of economic profit, taking optimal control of operating

cost into account is very needed. In this section, in order to discuss the cost optimization problem, the cost structure model is first established as follows:

- (1)  $r \equiv$  fixed setup cost for per busy cycle (this cost is due to the server being turned on each time);
- (2)  $h \equiv$  fixed holding cost per unit time for each customer present in the system (this cost originates from the customer’s sojourn time that consists of waiting time and service time).

Let  $F(N, D)$  be the long-run expected cost per unit time of the system. Applying the renewal of reward process theorem for a busy cycle that is defined as the Definition 5 above, it leads to

$$F(N, D) = h \cdot E[L] + \frac{r}{E[U]}, \tag{5}$$

where  $E[L]$  denotes the average steady-state queue size, and  $U$  denotes a busy cycle that consists of a server non-busy period (see Definition 2 above) and a server busy period (see Definition 1 above), denoted by  $I$  and  $B$ , respectively.

Next, we will find the expression of the expected busy cycle  $E[U]$  in the objective function. Clearly, based on Lemma 1, we obtain the average length of the server busy period as follows:

$$E[B] = E[b] \cdot E[L_2] = \frac{\rho}{\lambda(1 - \rho)} \cdot E[L_2],$$

where  $E[L_2]$  is given by Eq. (A7) in appendix.

We note that the number of customers in the system at the beginning of a server busy period is equal to the number of customers who arrived during the server non-busy period, and the arrival process is a Poisson process with rate  $\lambda$ , then the expected length of the server non-busy period is given by

$$E[I] = \frac{E[L_2]}{\lambda}.$$

Therefore, we can obtain the expected length of the busy cycle as follows:

$$E[U] = E[I] + E[B] = \frac{E[L_2]}{\lambda(1 - \rho)}. \tag{6}$$

Substituting Eqs. (2), (A7) and (6) into Eq. (5), we have

$$F(N, D) = h \left\{ \rho + \frac{\lambda^2 E[X^2]}{2(1 - \rho)} + \frac{y(\lambda)[1 - H(v(\lambda))] \sum_{m=1}^{N-1} m W^{(m)}(D) \int_0^\infty F^{(m+1)}(t) dV(t)}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N} \right\}$$



$$+ \frac{r\lambda(1 - \rho)[1 - v(\lambda)]}{[1 - v(\lambda)][1 - y(\lambda) + y(\lambda)H(v(\lambda))] + y(\lambda)[1 - H(v(\lambda))]\Delta_N}, \tag{7}$$

where  $\Delta_N$  is determined in Theorem 1.

It is observed from Eq. (7) that the cost function  $F(N, D)$  is extremely complex and nonlinear with respect to the decision variables  $N$  and  $D$ , which poses a hard task to find the analytic results for the optimum values  $N^*$  and  $D^*$  directly. From Eq. (5), we see that the total cost consists of two parts, one part is the cost incurred by customers waiting in the system, which is determined by  $hE[L]$ . The other part is the setup cost per unit time, denoted by  $\frac{r}{E[U]}$ . The dyadic  $\min(N, D)$  policy reduces the setup cost caused by frequently on/off a queueing system, but it also increases the cost of the customer’s waiting time. Therefore, the total cost function is a convex function of two variables  $N$  and  $D$ , and then, we can choose optimal values  $(N^*, D^*)$  for the optimization problem. In order to solve the optimization problem in Eq. (7), numerical calculation examples are provided to demonstrate how to search the optimal values  $N^*$  and  $D^*$ .

**Numerical Example 1** To demonstrate the model’s application, we consider a practical situation related to a production system. In order to effectively use the remaining productivity of the production system, the manager decided to accept some outside production tasks to increase the revenue of the system. Therefore, the manager designs a strategy: When the system does not has orders to be produced which belong to the normal production task, the factory begins to prepare to undertake the accepted outside task. But it will take a random period  $Y$  to do this preparation. That is, the factory goes through a delayed period  $Y$  before undertaking the accepted outside task. During the delayed period  $Y$ , if there are normal orders that need to be processed arrive in the system, the factory starts its production at once until the system becomes empty and the factory restarts a delayed period  $Y$ . Otherwise, the factory immediately undertakes the production of accepted outside tasks for a random length  $V$ . In order to control the queue size of normal orders during the random length  $V$  and reduce the cost that the system frequently switches its state for a long time, the system operates under the above modified  $\text{Min}(N, D)$ -policy based on delayed randomized multiple vacations. In this case, the specific meaning of each cost above is stated in more detail as follows.

$r \equiv$  fixed setup cost for each production normal order during per busy cycle;

$h \equiv$  fixed holding cost per unit time for each normal order present in the system (this cost originates from the order’s sojourn time that consists of waiting time and production time).

Suppose that the inter-arrival time  $\tau$  of normal orders, the production time  $\chi$  of each normal order, the delayed period  $Y$ , and the time  $V$  for the factory to continuously undertake the accepted outside task follow exponential distribution with parameters  $\lambda, \mu, \beta$  and  $\theta$ , respectively. Besides, we also assume that the number  $H$  of consecutive times that the factory undertakes the production of accepted outside tasks each with a random length  $V$  is geometrically distributed with parameter  $\alpha(0 < \alpha < 1)$ , and the production workload  $W$  of the factory for each normal order follows the distribution  $W(x) = 1 - e^{-\epsilon x}$ . Under these specific distributions and the above cost model, the

expression of the cost function  $F(N, D)$  above can be rewritten as:

$$\begin{aligned}
 F(N, D) = & h \left\{ \frac{\lambda}{\mu - \lambda} \right. \\
 & + \frac{\lambda\beta \sum_{m=1}^{N-1} m \left(\frac{\lambda}{\lambda+\theta}\right)^m \cdot \left[1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!}\right]}{(\lambda + \beta)(\lambda + \alpha\theta) + \lambda\beta \sum_{m=1}^{N-1} \left(\frac{\lambda}{\lambda+\theta}\right)^m \cdot \left[1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!}\right]} \left. \right\} \\
 & + \frac{r\lambda \left(1 - \frac{\lambda}{\mu}\right) \cdot (\lambda + \beta)(\lambda + \alpha\theta)}{(\lambda + \beta)(\lambda + \alpha\theta) + \lambda\beta \sum_{m=1}^{N-1} \left(\frac{\lambda}{\lambda+\theta}\right)^m \cdot \left[1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!}\right]}. \tag{8}
 \end{aligned}$$

Setting  $\lambda = 1, \mu = 2, \varepsilon = 4, \theta = 0.1, \beta = 0.5, \alpha = 0.2, r = 100, h = 2$ , and putting them into Eq. (8), we have

$$\begin{aligned}
 F(N, D) = & 2 \left\{ 1 + \frac{5 \sum_{m=1}^{N-1} m \left(\frac{10}{11}\right)^m \cdot \left[1 - e^{-4D} \sum_{i=0}^{m-1} \frac{(4D)^i}{i!}\right]}{\frac{153}{10} + 5 \sum_{m=1}^{N-1} \left(\frac{10}{11}\right)^m \cdot \left[1 - e^{-4D} \sum_{i=0}^{m-1} \frac{(4D)^i}{i!}\right]} \right\} \\
 & + \frac{765}{\frac{153}{10} + 5 \sum_{m=1}^{N-1} \left(\frac{10}{11}\right)^m \cdot \left[1 - e^{-4D} \sum_{i=0}^{m-1} \frac{(4D)^i}{i!}\right]}. \tag{9}
 \end{aligned}$$

Since the decision variable  $N$  is a discrete variable, the decision variable  $D$  is a continuous variable, for a fixed  $N(N = 1, 2, 3, \dots)$ , we can find the corresponding optimum solution  $D_N^*$  by  $\frac{\partial F(N, D)}{\partial D} = 0$ . All the calculations have been done on the MATLAB software package. The numerical results are displayed in Table 1 and Fig. 1.

It can be seen from Table 1 and Fig. 1 that the optimal control threshold value  $(N^*, D^*) = (12, 7.45)$  and the corresponding minimum cost  $F(N^*, D^*) = 24.8970$ .

**Table 1** The optimal threshold values and the minimum costs for different values of  $N$

$N$	$D_N^*$	$F(N, D_N^*)$	$\frac{\partial F(N, D)}{\partial D} \Big _{D=D_N^*}$
1	–	52.000 0	–
2	2.86	41.006 0	$-3.65 \times 10^{-4}$
6	4.78	27.443 1	$-1.05 \times 10^{-4}$
11	6.34	24.931 2	$9.81 \times 10^{-5}$
12	7.45	<b>24.897 0</b>	$1.72 \times 10^{-5}$
13	6.51	24.932 6	$6.41 \times 10^{-4}$
18	2.14	25.240 9	$6.98 \times 10^{-4}$
23	2.42	25.422 5	$-6.80 \times 10^{-4}$

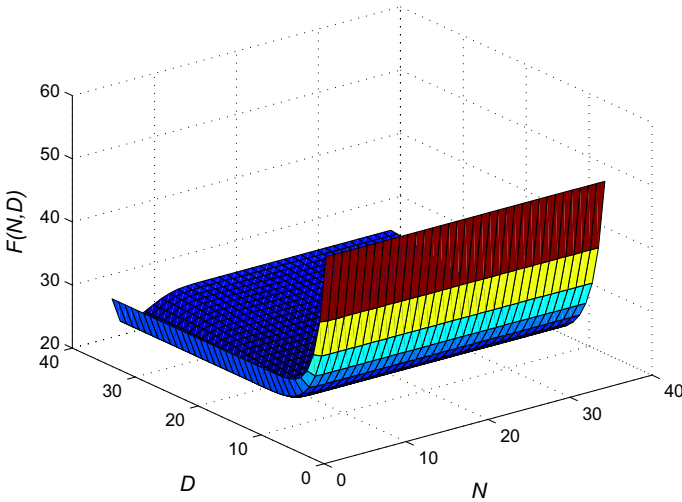


Fig. 1 The changes of  $F(N, D)$  with different values of  $N$  and  $D$

Table 2 shows the optimal values and the minimum costs for three policies under the cost function 8 when  $\mu = 2, \varepsilon = 4, \theta = 0.1, \beta = 0.5, \alpha = 0.2, r = 100, h = 2$ , with different inter-arrival rate.

**Numerical Example 2** Based on Numerical Example 1 mentioned above, if the number  $H$  of consecutive times that the factory undertakes the production of accepted outside tasks is a fixed positive integer  $J$ , that is,  $P\{H = J\} = 1$ . Then, the expression of the cost function  $F(N, D)$  can be simplified as:

$$\begin{aligned}
 F(N, D, J) = h & \left\{ \frac{\lambda}{\mu - \lambda} \right. \\
 & + \frac{\beta \left[ 1 - \left( \frac{\theta}{\lambda + \theta} \right)^J \right] \sum_{m=1}^{N-1} m \left( \frac{\lambda}{\lambda + \theta} \right)^m \cdot \left[ 1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!} \right]}{\lambda + \beta + \beta \left[ 1 - \left( \frac{\theta}{\lambda + \theta} \right)^J \right] \sum_{m=1}^{N-1} \left( \frac{\lambda}{\lambda + \theta} \right)^m \cdot \left[ 1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!} \right]} \Bigg\} \\
 & + \frac{r\lambda \left( 1 - \frac{\lambda}{\mu} \right) \cdot (\lambda + \beta)}{\lambda + \beta + \beta \left[ 1 - \left( \frac{\theta}{\lambda + \theta} \right)^J \right] \sum_{m=1}^{N-1} \left( \frac{\lambda}{\lambda + \theta} \right)^m \cdot \left[ 1 - e^{-\varepsilon D} \sum_{i=0}^{m-1} \frac{(\varepsilon D)^i}{i!} \right]} \cdot \quad (10)
 \end{aligned}$$

Table 2 Comparison of the minimum costs and the optimal threshold values for three different policies

$\rho$	N-policy	D-policy	Min(N, D)-policy
0.5	$N^* = 12$ $F(N^*, \infty) = 24.8970$	$D^* = 8.52$ $F(\infty, D^*) = 30.5009$	$N^* = 12, D^* = 7.45$ $F(N^*, D^*) = 24.8970$
0.7	$N^* = 11$ $F(N^*, \infty) = 25.9700$	$D^* = 3.51$ $F(\infty, D^*) = 36.6837$	$N^* = 11, D^* = 6.12$ $F(N^*, D^*) = 25.9700$

Setting  $\lambda = 1.8, \mu = 2, \varepsilon = 1, \theta = 0.1, \beta = 4, r = 100, h = 4$ , and putting them into Eq. (10), we have

$$F(N, D, J) = 4 \left\{ 9 + \frac{20 \left[ 1 - \left( \frac{1}{19} \right)^J \right] \sum_{m=1}^{N-1} m \left( \frac{18}{19} \right)^m \left( 1 - e^{-D} \sum_{i=0}^{m-1} \frac{D^i}{i!} \right)}{29 + 20 \left[ 1 - \left( \frac{1}{19} \right)^J \right] \sum_{m=1}^{N-1} \left( \frac{18}{19} \right)^m \left( 1 - e^{-D} \sum_{i=0}^{m-1} \frac{D^i}{i!} \right)} \right\} + \frac{522}{29 + 20 \left[ 1 - \left( \frac{1}{19} \right)^J \right] \sum_{m=1}^{N-1} \left( \frac{18}{19} \right)^m \left( 1 - e^{-D} \sum_{i=0}^{m-1} \frac{D^i}{i!} \right)}. \tag{11}$$

Since  $N$  and  $J$  are discrete variables, the variable  $D$  can be consecutively taken, we first take the fixed  $N$  to find minimum cost  $F(N, D_N^*, J_N^*)$  and the corresponding values of variables  $D$  and  $J$ . Then, by comparing the minimum values of  $F(N, D_N^*, J_N^*)$  at different values of  $N$ , the three-dimensional optimal control policy  $(N^*, D^*, J^*)$  and the corresponding minimum cost  $F(N^*, D^*, J^*)$  are determined.

Obviously, Table 3 and Fig. 2 show that the cost  $F(N, D, J)$  keep decreasing until a minimum and then increase as  $N$  grows. We observe that the minimum operating cost is 47.250 7 under  $N^* = 3, D^* = 12.27, J^* = 4$ .

### 5 Conclusions

In this paper, we proposed the modified  $\text{Min}(N, D)$ -policy based on the customers' number and server's workload, and studied an  $M/G/1$  queueing model with delayed randomized multiple vacations under the control of the modified  $\text{Min}(N, D)$ -policy. The well-known stochastic decomposition property of the steady-state queue size was used to obtain the probability generating function of the steady-state queue-length distribution. Also, the explicit expressions of the expected queue size and the additional queue-length distribution were derived by directly algebraic manipulations. To demonstrate the model's application, we further discussed the operating cost of system. Several numerical examples were provided to determine the optimal control policy for

**Table 3** The optimal threshold values and the minimum costs for different values of  $N$

$N$	$(D_N^*, J_N^*)$	$F(N, D_N^*, J_N^*)$	$\frac{\partial F(N, D, J_N^*)}{\partial D} \Big _{D=D_N^*}$
1	---	54.000 0	---
2	(13.18, 4)	48.467 6	$-6.32 \times 10^{-6}$
3	(12.27, 4)	<b>47.250 7</b>	$-6.07 \times 10^{-5}$
4	(10.42, 4)	47.404 4	$6.52 \times 10^{-4}$
5	(2.21, 4)	47.932 9	$4.92 \times 10^{-4}$
6	(2.04, 4)	48.222 8	$-8.73 \times 10^{-4}$
7	(1.93, 4)	48.285 1	$-3.67 \times 10^{-4}$

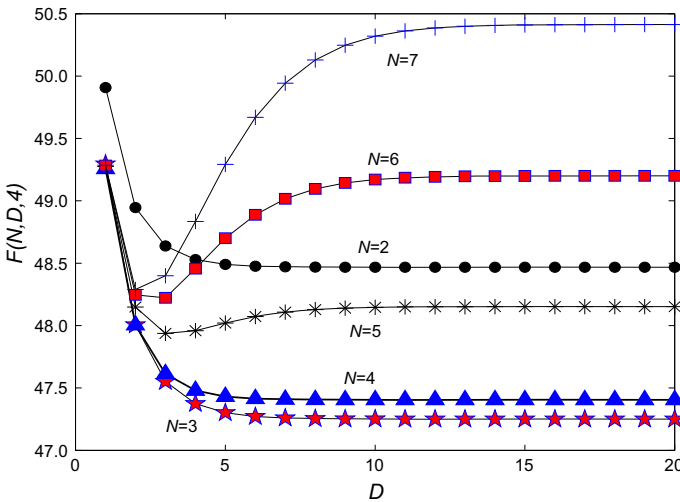


Fig. 2 The changes of  $F(N, D, 4)$  for different values of  $N$  and  $D$

economizing the system cost. Thus, the analysis of this paper will provide a potentially practical application for system managers and decision makers in related application areas such as flexible manufacturing systems.

### Appendix A. Proof of Theorem 1

**Proof** Let  $S_0 = I_0 = A_0 = 0, I_n = \sum_{i=1}^n \tau_i, S_n = \sum_{i=1}^n V_i, A_n = \sum_{i=1}^n W_i, n \geq 1$ . According to the assumption (4) of the model above, there are four cases of starting a new server busy period after the system becomes empty:

- (a) Starting a new server busy period during a delayed period (arrival occurs during a delayed period);
- (b) Starting a new server busy period during a vacation (arrival occurs during vacation period and the queue size reaches  $N$  or the total workload of the server for all the waiting customers is not less than a given threshold  $D$ );
- (c) Starting a new server busy period at the completion instant of a vacation (arrival occurs during vacation period);
- (d) Starting a new server busy period during idle period (no arrival occurs during the given  $H$  vacations).

Due to the above definition of the workload of the server for each customer, we know that the cumulative workload of the server for the waiting customers is completely determined by the quantity of events included in these customer’s service items, not by the cumulative service times of the customers. Thus, the beginning instant of a server busy period is the time point of regeneration of the system. Applying the mutual independence of the inter-arrival time  $\tau$ , the workload  $W$ , the service time  $\chi$ , the delayed period  $Y$ , the maximum vacation number  $H$  and the server vacation time  $V$ , we have that the lengths of sub-busy periods are identical and independently distributed

in this system, and furthermore, the decomposition property of the steady-state queue-length in Refs. [5, 12] holds. Therefore, the steady-state queue size discussed in this paper can be decomposed into the sum of two independent parts: one is the stationary queue size of the classical M/G/1 queueing system, and the other is that the additional queue size  $L_1$  caused by the delayed randomized multiple vacations and the modified  $\text{Min}(N, D)$ -policy. So that, the p.g.f. of the steady-state queue-length distribution, denoted by  $P(z)$ , is equal to the product of the p.g.f. of the steady-state queue-length distribution of the classical M/G/1 queueing system and the p.g.f. of the additional queue-length distribution ([12], Theorem 2.3.2), that is,

$$P(z) = P_{M/G/1}(z) \cdot P_{L_1}(z). \tag{A1}$$

In [32], the p.g.f. of the steady-state queue-length distribution of the classical M/G/1 queueing system is given by

$$P_{M/G/1}(z) = \frac{(1 - \rho)(1 - z)g(\lambda(1 - z))}{g(\lambda(1 - z)) - z}. \tag{A2}$$

In [12], the p.g.f. of the additional queue size distribution is presented by

$$P_{L_1}(z) = \frac{1 - L_2(z)}{(1 - z)E[L_2]}, \tag{A3}$$

where  $L_2$  is the random variable denoting the number of customers present in the system at the beginning instant of a server busy period,  $E[L_2]$  is the expected value of  $L_2$ , and  $L_2(z)$  is the p.g.f. of  $L_2$ .

Based on the delayed randomized multiple vacations and the modified  $\text{Min}(N, D)$ -policy above, the probability distribution of the random variable  $L_2$  can be derived by the law of total probability decomposition.

(1) The event  $\{L_2 = 1\}$  can be divided into the following four cases:

- After the system becomes empty, the first arrival occurs during the delayed period;
- After the system becomes empty, the first arrival occurs during the  $k$ th vacation  $V_k (k = 1, 2, \dots)$  and the workload  $W$  of the server for this customer is less than  $D$ ;
- After the system becomes empty, the first arrival occurs during the  $k$ th vacation  $V_k (k = 1, 2, \dots)$  and the workload  $W$  of the server for this customer is not less than  $D$ ;
- After the system becomes empty, the first arrival occurs after the end of  $H$  vacations (no arrival occurs during the  $H$  vacations).

Therefore,

$$P\{L_2 = 1\} = P\{0 \leq \hat{\tau} < Y\} + \sum_{l=1}^{\infty} h_l \sum_{k=1}^l P\{Y + S_{k-1} \leq \hat{\tau} < Y + S_k, Y + S_{k-1}$$

$$\begin{aligned}
 &\leq \widehat{\tau} \leq Y + S_k < \widehat{\tau} + l_1, W < D\} \\
 &+ \sum_{l=1}^{\infty} h_l \sum_{k=1}^l P\{Y + S_{k-1} \leq \widehat{\tau} < Y + S_k, W \geq D\} \\
 &+ P\{\widehat{\tau} > Y + S_H\} \\
 &= 1 - y(\lambda) \\
 &+ W(D) \sum_{l=1}^{\infty} h_l \sum_{k=1}^l \int_0^{\infty} \int_0^{\infty} \int_t^{t+x} p\{y \leq t + x < y + l_1\} \\
 &\quad dF(y)dV(x)dY(t) * V^{(k-1)}(t) \\
 &+ [1 - W(D)] \sum_{l=1}^{\infty} h_l \sum_{k=1}^l \int_0^{\infty} \int_0^{\infty} \int_t^{t+x} p\{t \leq y < t + x\} \\
 &\quad dF(y)dV(x)dY(t) * V^{(k-1)}(t) \\
 &+ \sum_{l=1}^{\infty} h_l \int_0^{\infty} [1 - F(t)]dY(t) * V^{(l)}(t) \\
 &= 1 - W(D)y(\lambda)[1 - H(v(\lambda))] \cdot \frac{1 - v(\lambda) - \int_0^{\infty} \lambda x e^{-\lambda x} dV(x)}{1 - v(\lambda)}, \quad (A4)
 \end{aligned}$$

where  $\widehat{\tau}$  represents the remaining inter-arrival time of  $\tau$  and  $W$  denotes the workload  $W$  of the server for this customer.

(2) For  $m(2 \leq m \leq N - 1)$ , the event  $\{L_2 = m\}$  can be divided into the following two cases:

- After the system becomes empty, there are  $m$  customers arrive during the  $k$ th vacation  $V_k (k = 1, 2, \dots)$  and the total workload of the server for  $m$  customers is less than  $D$ ;
- After the system becomes empty, there are  $m$  customers arrive during the  $k$ th vacation  $V_k (k = 1, 2, \dots)$  and the total workload of the server for  $m$  customers is not less than  $D$  for the first time.

Thus, the probability that the event  $\{L_2 = m\}$  occurs is given by

$$\begin{aligned}
 P\{L_2 = m\} &= \sum_{l=1}^{\infty} h_l \sum_{k=1}^l P\{Y + S_{k-1} \leq \widehat{\tau} < Y + S_k, Y + S_{k-1} \\
 &\quad < \widehat{\tau} + l_{m-1} \leq Y + S_k < \widehat{\tau} + l_m, A_m < D\} \\
 &+ \sum_{l=1}^{\infty} h_l \sum_{k=1}^l P\{Y + S_{k-1} \leq \widehat{\tau}, \widehat{\tau} + l_{m-1} < Y + S_k, \\
 &\quad A_{m-1} < D \leq A_m\} \\
 &= \frac{W^{(m)}(D)y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \int_0^{\infty} \frac{(\lambda x)^m}{m!} e^{-\lambda x} dV(x)
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{[W^{(m-1)}(D) - W^{(m)}(D)]y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \\
 & \cdot \sum_{n=m}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x). \tag{A5}
 \end{aligned}$$

(3) The event  $\{L_2 = N\}$  means that after the system becomes empty, there are  $N$  customers arrive during the  $k$ th vacation  $V_k (k = 1, 2, \dots)$  and the total workload of the server for the first  $N - 1$  customers is less than  $D$ . So that

$$\begin{aligned}
 P\{L_2 = N\} &= \sum_{l=1}^{\infty} h_l \sum_{k=1}^l P\{Y + S_{k-1} \leq \hat{\tau}, \hat{\tau} + l_{N-1} < Y + S_k, \\
 A_{N-1} < D\} &= \frac{W^{(N-1)}(D)y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \sum_{n=N}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x). \tag{A6}
 \end{aligned}$$

Therefore, the average number of customers at the beginning of the server busy period is given by

$$\begin{aligned}
 E[L_2] &= \sum_{m=1}^N m \cdot P\{L_2 = m\} \\
 &= 1 - y(\lambda) + y(\lambda)H(v(\lambda)) \\
 &+ \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ \sum_{m=1}^{N-1} m W^{(m)}(D) \int_0^{\infty} \frac{(\lambda x)^m}{m!} e^{-\lambda x} dV(x) \right. \\
 &+ \sum_{m=1}^{N-1} m [W^{(m-1)}(D) - W^{(m)}(D)] \sum_{n=m}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \\
 &+ \left. N W^{(N-1)}(D) \sum_{n=N}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right\} \\
 &= 1 - y(\lambda) + y(\lambda)H(v(\lambda)) \\
 &+ \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ - \sum_{m=1}^{N-1} m W^{(m)}(D) \sum_{n=m+1}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right. \\
 &+ \left. \sum_{m=0}^{N-1} (m+1) W^{(m)}(D) \sum_{n=m+1}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right\} \\
 &= 1 - y(\lambda) + y(\lambda)H(v(\lambda)) \\
 &+ \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ \sum_{m=1}^{N-1} W^{(m)}(D) \sum_{n=m+1}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right. \\
 &+ \left. \sum_{n=1}^{\infty} \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right\}
 \end{aligned}$$



$$\begin{aligned}
 &= 1 - y(\lambda) + y(\lambda)H(v(\lambda)) \\
 &+ \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ \sum_{m=0}^{N-1} W^{(m)}(D) \int_0^\infty \left[ 1 - \sum_{n=0}^m \frac{(\lambda x)^n}{n!} e^{-\lambda x} \right] dV(x) \right\} \\
 &= 1 - y(\lambda) + y(\lambda)H(v(\lambda)) + \frac{y(\lambda)[1 - H(v(\lambda))]\Delta_N}{1 - v(\lambda)}. \tag{A7}
 \end{aligned}$$

The p.g.f.  $L_2(z)$  is given by

$$\begin{aligned}
 L_2(z) &= \sum_{k=1}^N z^k \cdot P\{L_2 = k\} \\
 &= z[1 - y(\lambda) + y(\lambda)H(v(\lambda))] \\
 &+ \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ \sum_{k=1}^N z^k W^{(k-1)}(D) \sum_{n=k}^\infty \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right. \\
 &+ \sum_{k=1}^{N-1} z^k W^{(k)}(D) \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} dV(x) \\
 &\left. - \sum_{k=1}^{N-1} z^k W^{(k)}(D) \sum_{n=k}^\infty \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \right\} \\
 &= z[1 - y(\lambda) + y(\lambda)H(v(\lambda))] + \frac{y(\lambda)[1 - H(v(\lambda))]}{1 - v(\lambda)} \left\{ z[1 - v(\lambda)] \right. \\
 &+ (z - 1) \sum_{k=1}^{N-1} z^k W^{(k)}(D) \sum_{n=k+1}^\infty \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} dV(x) \left. \right\} \\
 &= z + \frac{(z - 1)y(\lambda)[1 - H(v(\lambda))] \sum_{k=1}^{N-1} z^k \int_0^\infty W^{(k)}(D) F^{(k+1)}(t) dV(t)}{1 - v(\lambda)} \tag{A8}
 \end{aligned}$$

Substituting Eqs. (A2), (A3), (A7) and (A8) in Eq. (A1), it gives Eq. (1). Then, Eq. (2) can be derived by using  $E[L] = \frac{d}{dz}[P(z)]|_{z=1}$ .

**References**

- [1] Yadin, M., Naor, P.: Queueing system with a removable service station. *Oper. Res. Q.* **14**(4), 393–405 (1963)
- [2] Heyman, D.P.: *T*-policy for the M/G/1 queue. *Manage. Sci.* **23**(7), 775–778 (1977)
- [3] Balachandran, K.R.: Control policies for a single server system. *Manage. Sci.* **19**(4), 1013–1018 (1973)
- [4] Balachandran, K.R., Tijms, H.C.: On the *D*-policy for M/G/1 queue. *Manage. Sci.* **21**, 1073–1076 (1975)
- [5] Fuhrman, S.W., Cooper, R.B.: Stochastic decomposition in the M/G/1 queue with generalized vacations. *Oper. Res.* **33**, 1117–1129 (1985)

- [6] Dshalalow, J.H.: Queueing processes in bulk systems under  $D$ -policy. *J. Appl. Probab.* **35**, 976–989 (1998)
- [7] Artalejo, J.R.: On the M/G/1 queue with  $D$ -policy. *Appl. Math. Model.* **25**, 1055–1069 (2001)
- [8] Artalejo, J.R.: A note on the optimality of the  $N$ - and  $D$ -policies for the M/G/1 queue. *Oper. Res. Lett.* **30**, 375–376 (2002)
- [9] Lee, H.W., Song, K.S.: Queue length analysis of MAP/G/1 queue under  $D$ -policy. *Stoch. Model.* **20**(3), 363–380 (2004)
- [10] Lee, H.W., Baek, J.W., Jeon, J.: Analysis of  $M^X/G/1$  queue under  $D$ -policy. *Stoch. Anal. Appl.* **23**, 785–808 (2005)
- [11] Doshi, B.: Queueing systems with vacations-A survey. *Que. Syst.* **1**, 29–66 (1986)
- [12] Tian, N.S.: *Stochastic Service System with Vacations*. Peking University Press, Beijing (2001)
- [13] Tian, N.S.: Queue M/G/1 with adaptive multistage vacations. *Math. Applicata.* **5**(4), 12–18 (1992)
- [14] Lee, H.W., Seo, W.J.: The performance of the M/G/1 queue under the dyadic  $\text{Min}(N, D)$ -policy and its cost optimization. *Perform. Eval.* **65**, 742–758 (2008)
- [15] Lee, H.W., Seo, W.J., Lee, S.W., Jeon, J.: Analysis of the MAP/G/1 queue under the  $\text{Min}(N, D)$ -policy. *Stoch. Model.* **26**, 98–123 (2010)
- [16] Yu, M.M., Tang, Y.H., Fu, Y.H., Pan, L.M.: GI/Geom/1/MWV queue with changeover time and searching for the optimum service rate in working vacation period. *J. Comput. Appl. Math.* **235**(8), 2170–2184 (2011)
- [17] Jing, C.X., Cui, Y., Tian, N.S.: Analysis of the M/G/1 queueing system with  $\text{Min}(N, V)$ -policy. *Oper. Res. Manag. Sci.* **15**(3), 53–58 (2006)
- [18] Tang, Y.H., Wu, W.Q., Liu, Y.P., Liu, X.Y.: The queue length distribution of M/G/1 queueing system with  $\text{Min}(N, V)$ -policy based on multiple server vacations. *Syst. Eng.-Theory Practice.* **34**(6), 1533–1546 (2014)
- [19] Tang, Y.H., Wu, W.Q., Liu, Y.P.: Analysis of M/G/1 queueing system with  $\text{Min}(N, V)$ -policy based on single server vacation. *Acta Math. Appl. Sin.* **37**(6), 976–996 (2014)
- [20] Luo, C.Y., Tang, Y.H., Yu, K.Z., Ding, C.: Optimal  $(r, N)$ -policy for discrete-time Geo/G/1 queue with different input rate and setup time. *Appl. Stoch. Models Business Indus.* **31**(4), 405–423 (2015)
- [21] Wei, Y.Y., Tang, Y.H., Yu, M.M.: Queue length distribution and optimum policy for M/G/1 queueing system under  $\text{Min}(N, D)$ -policy. *J. Syst. Sci. Math. Sci.* **35**(6), 729–744 (2015)
- [22] Lan, S.J., Tang, Y.H.: Geo/G/1 discrete-time queue with multiple server vacations and  $\text{Min}(N, V)$ -policy. *Syst. Eng.-Theory Practice.* **35**(3), 799–810 (2015)
- [23] Lan, S.J., Tang, Y.H., Yu, M.M.: System capacity optimization design and optimal threshold  $N^*$  for a Geo/G/1 discrete-time queue with single server vacation and under the control of  $\text{Min}(N, V)$ -policy. *J. Indus. Manag. Optim.* **12**(4), 1435–1464 (2016)
- [24] Lan, S.J., Tang, Y.H.: Reliability analysis of discrete-time  $\text{Geo}^{\lambda_1, \lambda_2}/G/1$  repairable queue with Bernoulli feedback and  $\text{Min}(N, D)$ -policy. *J. Syst. Sci. Math. Sci.* **36**(11), 2070–2086 (2016)
- [25] Lan, S.J., Tang, Y.H.: The structure of departure process and optimal control strategy  $N^*$  for Geo/G/1 discrete-time queue with multiple server vacations and  $\text{Min}(N, V)$ -policy. *J. Syst. Sci. Complex.* **30**(6), 1382–1402 (2017)
- [26] Cai, X.L., Tang, Y.H.: M/G/1 Repairable queueing system with warm standby failure and  $\text{Min}(N, V)$ -policy based on single vacation. *Acta Math. Appl. Sin.* **40**(5), 702–726 (2017)
- [27] Jiang, S.L., Tang, Y.H.: M/G/1 queueing system with multiple adaptive vacations and  $\text{Min}(N, V)$ -policy. *J. Syst. Sci. Math. Sci.* **37**(8), 1866–1884 (2017)
- [28] Wang, M., Tang, Y.H.: Optimal control policy of M/G/1 queueing system based on  $\text{Min}(N, D, V)$ -policy and single server vacation. *J. Syst. Sci. Math. Sci.* **38**(9), 1067–1084 (2018)
- [29] Luo, L., Tang, Y.H.: M/G/1 queueing system with  $\text{Min}(N, D, V)$ -policy control. *Oper. Res. Trans.* **23**(2), 1–16 (2019)
- [30] Wei, Y.Y., Tang, Y.H., Yu, M.M.: Recursive solution of queue length distribution for Geo/G/1 queue with delayed  $\text{Min}(N, D)$ -policy. *J. Syst. Sci. Inform.* **8**(4), 367–386 (2020)
- [31] Li, D., Tang, Y.: H: M/G/1 queueing system based on  $\text{Min}(N, D, V)$ -policy and single server vacation without interruption. *J. Syst. Sci. Math. Sci.* **41**(2), 533–556 (2021)
- [32] Tang, Y.H., Tang, X.W.: *Queueing Theory-Foundations and Analysis Techniques*. Science Press, Beijing (2006)