Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Applied Mathematical Modelling 30 (2006) 49–66
APPLIED MATHEMATICAL MODELLING
www.elsevier.com/locate/apm

On M/G/1 system under *NT* policies with breakdowns, startup and closedown

Jau-Chuan Ke *

Department of Statistics, National Taichung Institute of Technology, No. 129, Sec. 3, Sanmin Rd. Taichung 404, Taiwan, ROC

Received 1 March 2004; received in revised form 1 February 2005; accepted 4 March 2005
Available online 23 May 2005

Abstract

This paper studies the vacation policies of an M/G/1 queueing system with server breakdowns, startup and closedown times, in which the length of the vacation period is controlled either by the number of arrivals during the vacation period, or by a timer. After all the customers are served in the queue exhaustively, the server is shutdown (deactivates) by a closedown time. At the end of the shutdown time, the server immediately takes a vacation and operates two different policies: (i) The server reactivates as soon as the number of arrivals in the queue reaches to a predetermined threshold N or the waiting time of the leading customer reaches T units; and (ii) The server reactivates as soon as the number of arrivals in the queue reaches to a predetermined threshold N or T time units have elapsed since the end of the closedown time. If the timer expires or the number of arrivals exceeds the threshold N, then the server reactivates and requires a startup time before providing the service until the system is empty. If some customers arrive during this closedown time, the service is immediately started without leaving for a vacation and without a startup time. We analyze the system characteristics for each scheme.

Keywords: Delay cycle; Markov process; Regenerative; Vacation period; Waiting time distribution

* Tel.: +886 4 22196332; fax: +886 4 22196331.
E-mail address: jauchuan@ntit.edu.tw

0307-904X/$ - see front matter
doi:10.1016/j.apm.2005.03.022

1. Introduction

1.1. Literature review

Queueing systems with vacation period effectively arise in the stochastic modelling of many computers and communication systems, manufacturing/production, and inventory systems. For a comprehensive survey see Doshi [1]. In general, in order to control the length of the vacation period we either use a queue length N (known as N policy by Yadin and Naor [2]) or a timer T (known as T policy by Heyman [3]). Of particular interest is a tighter control over the length of the vacation period, which can be achieved by combining N policy and T policy. Considerable efforts have been devoted to study these types of the controllable queueing models, such as Teghem [4], Takagi [5], and many others. From those vast and rich literature, we review some well-known research works according to listed below five kinds of classification for the studies on these control policies.

1.1.1. Queueing systems under N policy

Yadin and Naor [2] first introduced the concept of an N policy (without startup) which turns the server on whenever N ($N \geqslant 1$) or more customers are present, turns the server off only when none is present. For a reliable server, the N policy M/G/1 queueing system was first studied by Heyman [6] and was developed by several researchers such as Teghem [4], Tijms [7], Gakis et al. [8], Artalejo [9], Wang and Ke [10] and others. Analytic steady-state solutions of the N policy $M/H_k/1$ queueing system were first obtained by Wang and Yen [11]. For an unreliable server, Wang [12], Wang [13], and Wang et al. [14] derived analytic steady-state solutions of the N policy M/M/1, the N policy $M/E_k/1$, and the N policy $M/H_2/1$ queueing systems, respectively. Later, Wang and Ke [15] investigated three control policies in an M/G/1 queueing system and proved that in three control policies, the probability that the server is busy in the steady-state is equal to the traffic intensity. Recently, Wang et al. [16] extended Wang and Yen's system [11] to unreliable server case.

1.1.2. Queueing systems under T policy

The server returns to provide service immediately as long as there is at least one customer in the system, but when there are no customers in the system, it becomes unavailable for a fixed length of time T (a vacation). After a vacation period of length T, the server returns to the system. It begins to serve if there is at least one customer present in the waiting line; otherwise, the server waits another period of length T and so on until at least one customer is present. This kind of control policy is called the T policy and the T policy for the M/G/1 queueing system with a reliable server was investigated by Heyman [3], Levy and Yechiali [17] and Gakis et al. [8]. Recently, Tadj [18] investigated the T policy for the M/G/1 quorum queueing system by using embedded Markov chain. He obtained the probability generating function of the number of customers in the system, the expected length of the idle period, busy period and busy cycle, and the determination of the optimum value T.

1.1.3. Queueing systems under N policy with general vacations

For some controllable queueing systems with general vacations, it is usually assumed that the server becoming available, or unavailable, completely depends on the number of customers in the system. Every time when the system is empty, the server goes on a vacation. The instance at which the server comes back from a vacation and finds at least N (predetermined threshold) customers in the system it begins serving immediately and exhaustively. This type of control policy is also called N policy queueing systems with vacations. Kella [19] and Lee and Srinivasan [20] first provided detail discussions concerning N policy M/G/1 and $M^{[x]}$/G/1 queueing systems with reliable server and vacations, respectively. Later, Lee et al. [21,22] analyzed in detail the batch arrival M/G/1 queueing system under N policy with a single vacation and repeated vacations. Their results significantly confirmed the stochastic decomposition property given by Fuhrmann and Cooper [23]. Recently, Ke [24] investigated Lee and Srinivasan's system [20] by considering vacations and startup of an unreliable server.

1.1.4. Queueing systems with N policy and general startup

The server startup corresponds to the preparatory work of the server before starting the service. In some actual situations, the server often requires a startup time before starting his each service period. Concerning queueing systems combining N policy with a startup time, the N policy M/G/1 queueing system with startup time was first studied by Minh [25] and was investigated by several researchers such as Medhi and Templeton [26], Takagi [5,27], Lee and Park [28], and so on. Recently, Hur and Paik [29] examined the operation characteristics of M/G/1 queueing system under N policy with server startup and explained how the system's optimal policy and cost behave for various arrival rates. The concept of closedown-time was first introduced by Takagi [5]. Niu and Takahashi [30] studied the performance analysis of the switched virtual connection (SVC) by a closedown time which is corresponding to an inactive timer during which the SVC resource is reserved to anticipate more customers (packets) from the same IP flow. The above authors mentioned focus on reliable servers for the server startup or closedown.

1.1.5. Combined N policy and T policy

Gakis et al. [8] first introduced the concept of a Min(N, T) policy which terminates the server's vacation if either N ($N \geqslant 1$) customers have appeared in the system or T time units have elapsed since the end of a busy period or the end of the previous T time units and at least one customer in the system waits for service, which occurs first. The distributions and the first two moments of the busy and idle periods in the Min(N, T) policy M/G/1 queueing systems with a reliable server operating under six dyadic policies were developed by Gakis et al. [8]. They have also shown that in all policy cases the steady-state probability that the server is busy is equal to the traffic intensity. Doganata [31] first considered the NT policy M/G/1 queueing system with a reliable server and derived the expected values of the performance measures. Alfa and Frigui [32] studied the NT policy MAP/PH/1 queueing system with a reliable server and no startup. They have proved that the vacation period distribution is of phase type. Alfa and Li [33] and Li and Alfa [34] separately studied the NT policy for both M/G/1 and M/M/S queue in a manufacturing system with cost structure. Recently, Hur et al. [35] optimized the operating cost of an M/G/1 queueing system using Min(N, T) policy. They derived steady-state system size distribution, and established a cost function to reveal the characteristics of the cost function and found the optimal operating policy. In this paper our NT policy differs from the NT policy of authors mentioned above. We will consider the NT policy for an M/G/1 queueing system, in which the unreliable server is shutdown by a closedown time when the system is empty, and needs to perform a startup time before starting his each service (completion) period.

Existing research works, including those mentioned above, have never investigated cases involving server breakdowns, startup and closedown. Queuing models with server breakdown, startup and closedown accommodate the real-world situations more closely. There are many applications in which the system is operated only intermittently. The system will be shutdown by a closedown time when no customers are present. Therefore, it would be practical to consider the vacation policies for queuing models, in which an unreliable server is characterized by startup and closedown.

1.2. Model description and assumptions

In this paper, we consider two different kinds of NT-policies for the M/G/1 queueing system where termination of the vacation period is controlled by two threshold parameters N and T. The detailed description of the models are given as follows:

Assumptions of the Model 1. The server terminates his vacation as soon as the number of arrivals reaches N or the waiting time of the leading customer reaches T units since the end of the closedown time.

1. Customers arrive according to a Poisson with rate λ. The service time provided by a single server is an independent and identically distributed random variable (S) with a general distribution function $S(t)$. If at any time any customer arrives, he goes to the service facility for service. Arriving customers forms a single waiting line based on the order of his arrivals. The server can serve only one customer at a time.
2. The server is subject to breakdowns at any time with Poisson breakdown rate α when it is working. Whenever the server fails, it is immediately repaired at a repair facility, where the repair time is an independent and identically distributed random variable (R) with a general distribution function $R(t)$.
3. In case the server breaks down when serving customers, he is sent for repair and the customer who has just being served should wait for the server back to complete his remaining service. Immediately after the server is fixed, he starts to serve customers until the system is empty, and the service time is cumulative. A customer who arrives and finds the server busy or broken down must wait in the queue until a server is available. Although no service occurs during the repair period of a broken server, customers continue to arrive according to a Poisson process.
4. Whenever the system becomes empty, the server shuts down (deactivates) by a closedown time. When the number of arrivals in the queue reaches to a predetermined threshold N or the waiting time of the leading customer reaches T units since the end of the closedown time, the server immediately reactivates and is temporarily unavailable for the waiting customers. He needs a startup time with random length U before starting service. As soon as the server finishes startup, he starts serving the waiting customers until the system becomes empty.
5. If a customer arrives during a closedown time, the service is immediately started without satisfying the conditions of NT vacation and without a startup time.

Assumptions of the Model 2. The server terminates his vacation as soon as the number of arrivals accumulated to N or the time units after the end of the closedown time reaches predetermined T units.

The first three and the last assumptions are the same as those in *Model 1*. However the fourth assumption now is that the system only reactivates if the number of arrivals in the queue waiting for service reaches to predetermined threshold N, or the time units after the end of the closedown time reaches T units. The server still requires to perform a startup before his each completion period is started.

As a practical application fitting our general model is the following *produce to order* system for a product, which based on the work of Zhang et al. [36]. Assume that customer orders for this product arrive according to a random process. It is desirable that the production begins whenever the number of orders reaches a critical value N (the minimum set-up lot size satisfied benefit). In other words, if the number of orders is less than N the production waits until the Nth order arrives. To maintain good business or have some cause it does not make the orders delay too long if there is not up to N available. The management policy is to set up the facility (startup) and begin production when there are N orders in the queue or the first order has been waiting for T units of time. Moreover, the production may be interrupted when facility encounters unpredicted breakdowns. When production interruptions occur (breakdowns), it is emergently recovered with a random time. The facility is shut down by a closedown time whenever the production ends and no orders arrive. The closedown time may be referred to machine maintenance and others.

The objectives of this paper are as follows: First, we construct some propositions for the M/G/1 queue with an unreliable server and general vacations, which will be used frequently in subsequent sections. Second, we derive the system size distribution, the waiting time distribution in the queue and other important system characteristics for two different kinds of NT policy models. Third, we show that the results generalize those of the N policy and the T policy M/G/1 queueing system with a reliable server. Finally, some special cases are also presented in this paper.

2. The M/G/1 queue with an unreliable server and generalized vacations

Before going into the study of this system, let us recall some results in the ordinary M/G/1 queueing system with an unreliable server. Let H be a random variable representing the completion time of a customer (the nth arrival), it is time interval from the server begins to serve the customer (the nth arrival) to the ends of his service, which include the repair time of server because of probable breakdowns in the customer's service time (the service time of the nth arrival). The useful results by Gaver [37] and Tang [38] are as follows:

$$E[H] = E[S](1 + \alpha E[R]), \tag{1}$$

$$E[H^2] = (1 + \alpha E[R])^2 E[S^2] + \alpha E[S]E[R^2], \tag{2}$$

$$\rho_H = \lambda E[H] = \rho(1 + \alpha E[R]), \tag{3}$$

where $\rho = \lambda E[S]$. Note that ρ_H is traffic intensity and it should be assumed to be less than unity.

Now, we consider an M/G/1 queueing system in which the server may meet unpredictable breakdowns while working and may perform generalized vacations (vacation, or startup, or idle) when system is empty.

First, let us define several items used for the generalized vacation model. We call a time interval when the server is either unavailable (for various reasons such as vacation, or startup or idle) an *idle period*. A time interval when the server is working continuously is called a *busy period*. A *vacation period* may contain a number of vacations, just as a *busy period* may contain a number of service times. During the *busy period*, the server may break down and start his repair immediately. This is called *breakdown period*. After the server is repaired, it returns and provides service until there are no customers in the system. The *completion period* is from the end of the *idle period* to no customers in the system, which occurs before the system becomes empty and can be represented as the sum of all *busy periods* and *breakdown periods*. The time interval consisting of an *idle period* and a following *completion period* is called a *busy cycle*.

Further, let a_k be the probability that k customers arrive during a completion time, we have

$$a_k = \int_0^\infty \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} \mathrm{d}H(t), \quad k = 0, 1, 2, \ldots, \tag{4}$$

where $H(t)$ is d.f. of H.

2.1. Preliminary formulas for system size distribution

We first define f as the number of customers that arrive during each vacation, and φ as the number of customers that arrive during a vacation period. The probability distributions and their probability generating functions for f and φ are as follows:

$$f_k = \Pr[f = k], \quad k = 0, 1, 2, \ldots, \tag{5a}$$

$$\varphi_k = \Pr[\varphi = k], \quad k = 1, 2, 3, \ldots, \tag{5b}$$

$$F(z) = \sum_{k=0}^{\infty} f_k z^k, \tag{6a}$$

and

$$\varphi(z) = \sum_{k=1}^{\infty} \varphi_k z^k. \tag{6b}$$

A vacation period begins only when there are no customers in the exhaustive service system. The number φ of customers that arrive during a generalized-vacation period (idle period) may depend on the arrival process during that generalized-vacation period.

We denote by L_n the number of customers left in the system immediately after the nth departing customer. A sequence of random variables $\{L_n; n = 1, 2, 3, \ldots\}$ constitutes a Markov chain. Let us define the steady-state distribution for $\{L_n; n = 1, 2, 3, \ldots\}$ as

$$\pi_k = \lim_{n\to\infty} \Pr[L_n = k], \quad k = 0, 1, 2, \ldots$$

As in the analysis to the ordinary M/G/1 system, the state-transition probability $p_{ji} = \Pr[L_{n+1} = i|L_n = j]$ is given by

$$p_{ji} = \begin{cases} \sum_{k=1}^{i+1} \varphi_k a_{i-k+1} & j = 0, \quad i \geqslant 0, \\ a_{i-j+1} & i \geqslant j-1, \\ 0 & j \geqslant 1, \quad 0 \leqslant i < j-1, \end{cases}$$

which satisfy the following equations:

$$\pi_k = \sum_{j=0}^{\infty} \pi_j p_{jk} \quad \text{and} \quad \sum_{k=0}^{\infty} \pi_k = 1$$

from which we get

$$\pi_k = \pi_0 \sum_{j=1}^{k+1} \varphi_j a_{k-j+1} + \sum_{j=1}^{k+1} \pi_j a_{k-j+1}. \tag{7}$$

Multiplying (7) by z^k and summing over $k = 0, 1, 2, \ldots$, we get the p.g.f. of $\{\pi_k\}$ as

$$\begin{aligned} \Pi(z) &= \pi_0 \sum_{j=1}^{\infty} \varphi_j z^{j-1} \sum_{k=j-1}^{\infty} a_{k-j+1} z^{k-j+1} + \sum_{j=1}^{\infty} \pi_j z^{j-1} \sum_{k=j-1}^{\infty} a_{k-j+1} z^{k-j+1} \\ &= \frac{\pi_0 \varphi(z) H^*(\lambda - \lambda z)}{z} + \frac{(\Pi(z) - \pi_0) H^*(\lambda - \lambda z)}{z}, \end{aligned} \tag{8}$$

where $H^*(\cdot)$ is LST of H.

Solving this equation for $\Pi(z)$, we get

$$\Pi(z) = \frac{\pi_0 [1 - \varphi(z)] H^*(\lambda - \lambda z)}{H^*(\lambda - \lambda z) - z}. \tag{9}$$

We determine π_0 by normalization condition, i.e., $\Pi(1) = 1$, to be

$$\pi_0 = \frac{1 - \rho_H}{E[\varphi]}. \tag{10}$$

Substituting (10) into (9), it finally yields

$$\Pi(z) = \frac{(1 - \rho_H)[1 - \varphi(z)] H^*(\lambda - \lambda z)}{E[\varphi][H^*(\lambda - \lambda z) - z]} \tag{11}$$

from which we have the expected number of customers in the system

$$E[L] = \frac{E[\varphi(\varphi - 1)]}{2E[\varphi]} + \frac{\lambda^2 E[H^2]}{2(1 - \rho_H)} + \rho_H. \tag{12}$$

3. Analysis of the *NT* policy for Model 1

The end of the idle period for *Model 1* can be described by the following two situations:

Case 1: No customers arrive when the server is shutting down.

In this case the server operates the *NT* policy after shutdown. We may imagine an M/G/1 system with vacations that end with $f(>0)$ customers and with a startup time U for the first customer. In this case, *Model 1* is as an extension of the M/G/1 queueing system given in Section 2. For this system, we can use $F(z)U^*(\lambda-\lambda z)$ to denote p.g.f. of the number of customers found in the system at the end of the idle period, where $U^*(\theta)$ is LST of U.

Case 2: Some customers arrive when the server is shutting down.

In this case, we use z to denote p.g.f. of the number of customers found in the system at the end of the idle period.

From the well-known stochastic decomposition results by Fuhrmann and Cooper [23] and using the above inferences listed, we obtain the p.g.f. of the number of customers found in the system at the beginning of the busy period

$$\varphi(z)=D^*(\lambda)F(z)U^*(\lambda-\lambda z)+(1-D^*(\lambda))z. \tag{13}$$

3.1. System size distribution and expected number of customers in the system

Using (11)–(13), the distribution for the number of customers in the system at an arbitrary time is given by

$$\Pi(z)=\frac{(1-\rho_H)[1-D^*(\lambda)F(z)U^*(\lambda-\lambda z)-(1-D^*(\lambda))z]H^*(\lambda-\lambda z)}{[D^*(\lambda)(E[F]+\lambda E[U])+1-D^*(\lambda)][H^*(\lambda-\lambda z)-z]}, \tag{14}$$

$$E[L]=\frac{D^*(\lambda)(E[F(F-1)]+2\lambda E[F]E[U]+\lambda^2E[U^2])}{2[D^*(\lambda)(E[F]+\lambda E[U])+1-D^*(\lambda)]}+\frac{\lambda^2E[H^2]}{2(1-\rho_H)}+\rho_H. \tag{15}$$

In Eqs. (14) and (15), F represents the distribution of the number of arrivals during the vacation period and the probability generating function for F is $F(z)$, which are generalized from those of Section 2, where

$$f_k=\Pr[k\text{ arrivals during the idle period}]=\frac{(\lambda T)^k\mathrm{e}^{-\lambda T}}{k!},\quad k=0,1,\ldots,N-2$$

and

$$f_{N-1}=1-\sum_{k=0}^{N-2}f_k.$$

In this model,

$$F(z)=\sum_{k=0}^{N-1}f_kz^k.$$

It is to be noted that when $T \to \infty$, we have $E[F] = N$ and $E[F(F-1)] = N(N-1)$. Likewise, when $N \to \infty$, we have $E[F] = \lambda T$ and $E[F(F-1)] = \lambda^2 T^2$.

3.2. Waiting time distribution and expected waiting time in the queue

First let $H_{\mathrm{o}}^*(\theta)$ denote the LST of the completion period in the ordinary M/G/1 queueing system with an unreliable server. It is useful results of Tang [38] that the LST of the completion period started with one customer in the M/G/1 queueing system with an unreliable server can be expressed by

$$H_{\mathrm{o}}^*(\theta) = H^*[\theta + \lambda - \lambda H_{\mathrm{o}}^*(\theta)], \tag{16}$$

where $H^*(\cdot)$ is LST of completion time H.

Differentiating (16) with respect to θ, we derive the expected length of the completion period initiated with one customer in the system as

$$E[H_{\mathrm{o}}] = \frac{E[H]}{1-\rho_H}, \tag{17}$$

where $E[H]$ and ρ_H are given by (1) and (3).

The LST $W^*(\theta)$ of the waiting time distribution in the queue can be found as follows. There is a way to derive $W^*(\theta)$ using regenerative arguments. Note that the points in time when each completion period ends are the system's regeneration points. At these points, the number of customers in the system is zero. The time interval between two such successive regeneration points is called a *generalized vacation cycle*, whose length is denoted by Θ_{c}. The LST $\Theta_{\mathrm{c}}^*(\theta)$ and the expectation for Θ_{c} are found by the delay cycle arguments (see [5, Section 1.2]) as

$$\Theta_{\mathrm{c}}^*(\theta) = D^*(\lambda)F(1-(\theta+\lambda-\lambda H_{\mathrm{o}}^*(\theta))/\lambda)U^*(\theta+\lambda-\lambda H_{\mathrm{o}}^*(\theta)) + (1-D^*(\lambda))I_{\mathrm{o}}^*(\theta)H_{\mathrm{o}}^*(\theta), \tag{18}$$

and

$$\begin{aligned} E[\Theta_{\mathrm{c}}^*] &= \frac{D^*(\lambda)(E[F]+\lambda E[U])}{\lambda(1-\rho_H)} + (1-D^*(\lambda))\left(\frac{1}{\lambda}+\frac{E[H]}{1-\rho_H}\right) \\ &= \frac{1}{\lambda(1-\rho_H)}[D^*(\lambda)(E[F]+\lambda E[U]) + 1 - D^*(\lambda)], \end{aligned} \tag{19}$$

where $I_{\mathrm{o}}(\theta) = \lambda/(\lambda+\theta)$ is LST of the idle period for the ordinary M/G/1 queueing system.

To find $W^*(\theta)$, we consider three types of customers according to their arrival points. For these customers that arrive during a delay cycle that is initiated with a generalized-vacation whose LST $F(1-\theta/\lambda)U^*(\theta)$, it follows from the arguments by Takagi [5, Section 1.2] that we have

$$W^*(\theta|\text{delay cycle}) = \frac{(1-\rho_H)[1-F(1-\theta/\lambda)U^*(\theta)]}{(E[F]/\lambda+E[U])[\theta-\lambda+\lambda H^*(\theta)]}. \tag{20}$$

Customers that arrive during the closedown time have zero waiting time. Customers that during the completion period have the LST of the waiting time

$$W^*(\theta|\text{busy}) = \frac{(1-\rho_H)[1-H^*(\theta)]}{E[H][\theta-\lambda+\lambda H^*(\theta)]}. \tag{21}$$

Thus, considering all cases we obtain

$$W^*(\theta) = \frac{1}{E[\Theta_c]}\left[D^*(\lambda)\left(\frac{E[F]+\lambda E[U]}{\lambda(1-\rho_H)}\right)\times W^*(\theta|\text{delay cycle}) + (1-D^*(\lambda))\left(\frac{1}{\lambda}+\frac{E[H]}{1-\rho_H}\times W^*(\theta|\text{busy})\right)\right], \tag{22}$$

which can be simplified as

$$W^*(\theta) = \frac{(1-\rho_H)\{D^*(\lambda)[\lambda-\lambda F(1-\theta/\lambda)U^*(\theta)]+\theta(1-D^*(\theta))\}}{[D^*(\lambda)(E[F]+\lambda E[U])+1-D^*(\lambda)][\theta-\lambda+\lambda H^*(\theta)]}. \tag{23}$$

Differentiating (23) with respect to θ, we finally get the expected waiting time in the queue

$$E[W] = \frac{D^*(\lambda)(E[F(F-1)]+2\lambda E[F]E[U]+\lambda^2E[U^2])}{2\lambda[D^*(\lambda)(E[F]+\lambda E[U])+1-D^*(\lambda)]}+\frac{\lambda E[H^2]}{2(1-\rho_H)}, \tag{24}$$

which is identical to Little's result $\left(E[W]=\frac{E[L]}{\lambda}-E[H]\right)$ obtained through (15).

Note that the second terms of (24) represent the expected waiting time in the queue for the ordinary M/G/1 queueing system. Hence, the first term can be called 'extra expected waiting time in the generalized-vacation period (idle period)'.

3.3. Other system characteristics

In this section we derive the expected length of the completion period, the idle period and the busy cycle.

3.3.1. Expected length of the completion period

Let $H^*_{NT}(\theta)$ be the LST of the completion period for the NT policy M/G/1 queueing system where an unreliable server is characterized by startup and closedown times. This system the server is shutdown (deactivates) by a shutdown time at end of each completion period. When the number of customers in the queue reaches N or the waiting time of the leading customer reaches T units since the end of the closedown time, the server immediately reactivates and the first customer in each completion period needs to wait for a server startup time U before receiving the service. If some customers present when the server is shutting down, he immediately starts serving the waiting customers without waiting the conditions of NT policy and without a startup time. It follows from a property of the Poisson arrival process and the assumption of exhaustive service that those points in time at which each completion period ends are the regeneration points of the system. If there are k customers in the system at the beginning of a completion period, the subsequent completion period will consist of k independent completion times, which each is initiated with a single customer's completion time. Thus, we have

$$H^*_{NT}(\theta) = \sum_{k=1}^{\infty}\varphi_k[H^*_o(\theta)]^k = \varphi[H^*_o(\theta)]. \tag{25}$$

Substituting (13) into (25), it finally yields

$$H^*_{NT}(\theta) = D^*(\lambda)F(H^*_{\mathrm{o}}(\theta))U^*(\lambda - \lambda H^*_{\mathrm{o}}(\theta)) + (1 - D^*(\lambda))H^*_{\mathrm{o}}(\theta), \tag{26}$$

where $H^*_{\mathrm{o}}(\theta)$ is given in (16).

Differentiating (26) with respect to θ and using (17), we obtain the expected length of the completion period for the NT policy M/G/1 queueing system given by

$$E[H_{NT}] = -\frac{\mathrm{d}H^*_{NT}(\theta)}{\mathrm{d}\theta}\bigg|_{\theta=0} = [D^*(\lambda)(E[F] + \lambda E[U]) + 1 - D^*(\lambda)]\frac{E[H]}{1-\rho_H}. \tag{27}$$

3.3.2. Expected length of the busy period and the breakdown period

The expected length of the busy period and the expected length of the breakdown period are denoted by $E[B_{NT}]$ and $E[D_{NT}]$, respectively. Recall that the completion period is the sum of the busy period and the breakdown period which implies $E[H_{NT}] = E[B_{NT}] + E[D_{NT}]$. Hence from (1) and (27) we have

$$E[B_{NT}] = \frac{\{D^*(\lambda)(E[F] + \lambda E[U]) + 1 - D^*(\lambda)\}E[S]}{1-\rho_H} \tag{28}$$

and

$$E[D_{NT}] = \frac{\alpha\{D^*(\lambda)(E[F] + \lambda E[U]) + 1 - D^*(\lambda)\}E[S]E[R]}{1-\rho_H}. \tag{29}$$

3.3.3. Expected length of the idle period

Let I_{NT} be the idle period for the NT policy M/G/1 queueing system in which an unreliable server is characterized by startup and closedown times. The server is shutdown by a closedown time as soon as the system is empty. When closedown time is over, the server operates NT vacation policy and requires a startup time before providing service. If some customers arrives at the system when the server is shutting down, the service is immediately started without taking NT policy and without a startup time. The idle period for this system is composed of the following two cases:

Case 1: *No customers arrive when the server is shutting down*

Let $I_{\text{idle-startup}}$ be the length of the idle period minus the startup period since server shutdown. It can be shown from Kleinrock [39, Chapter 5] that

$$I^*_{\text{idle-startup}}(\lambda - z\lambda) = F(z) = \sum_{k=0}^{N-1} f_k z^k, \tag{30}$$

where $I^*_{\text{idle-startup}}(\theta)$ is LST of $I_{\text{idle-startup}}$.

The expectation of $I^*_{\text{idle-startup}}$ is found from (30) as

$$E[I^*_{\text{idle-startup}}] = -\frac{\mathrm{d}I^*_{\text{idle-startup}}(\theta)}{\mathrm{d}\theta}\bigg|_{\theta=0} = \frac{\mathrm{d}F(z)}{\lambda \mathrm{d}z}\bigg|_{z=1} = \frac{\sum_{k=0}^{N-1} k f_k}{\lambda} = \frac{E[F]}{\lambda}. \tag{31}$$

For this case, let $I_1^*(\theta)$ be the LST of the idle period which is the sum of $I_{\text{idle-startup}}$ and U. It is clear that $I_1^*(\theta)$ is convolution of $I^*_{\text{idle-startup}}(\theta)$ and $U^*(\theta)$. Thus we have

$$I_1^*(\theta) = I^*_{\text{idle-startup}}(\theta) \otimes U^*(\theta) = I^*_{\text{idle-startup}}(\theta)U^*(\theta), \tag{32}$$

which leads to

$$E[I_{NT}] = \frac{E[F]}{\lambda} + E[U]. \tag{33}$$

Case 2: *Some customers arrive when the server is shutting down*

If some customers present when the server is shutting down, he immediately starts providing service for the waiting customers without a startup time. Thus we have the LST of the idle period for this case

$$I_2^*(\theta) = \frac{\lambda}{\theta + \lambda}, \tag{34}$$

which yields

$$E[I_2] = \frac{1}{\lambda}. \tag{35}$$

From the above two cases listed, we have the expected length of the idle period for the NT policy M/G/1 queueing system (*Model 1*).

$$E[I_{NT}] = D^*(\lambda)\left(\frac{E[F]}{\lambda} + E[U]\right) + \left(\frac{1 - D^*(\lambda)}{\lambda}\right). \tag{36}$$

3.4. Expected length of the busy cycle

The busy cycle for the NT policy M/G/1 queueing system, denoted by Ω_{NT}, is the length of time from the beginning of the last idle period to the beginning of the next idle period. From (27) and (36), we have

$$E[\Omega_{NT}] = \frac{D^*(\lambda)\{E[F] + \lambda E[U]\} + 1 - D^*(\lambda)}{\lambda(1 - \rho_H)}. \tag{37}$$

4. Analysis of the *NT* policy for Model 2

In this section, we consider the NT policy for a system consisting of the classical N-policy and T-policy. In this system, the only difference from the system discussed in Section 3 is that the system is reactivated as soon as there are N customers accumulated in the queue or T time units have elapsed since the end of the closedown time. It is to be noted in this case that the next closedown time, startup period and completion period are zero if the time units after the end of the completion period reaches the predetermined time T and there is no customer arrival. This system is as another extension of the M/G/1 queue given in Section 2. From similar arguments as for *Model 1*. For this system, use the p.g.f.

$$\varphi(z) = D^*(\lambda)[F(z) - f_0 + f_0 z]U^*(\lambda - \lambda z) + (1 - D^*(\lambda))z \tag{38}$$

for the number of customers found in the system at the end of the idle period.

4.1. System size distribution and expected number of customers in the system

From (11), (12) and (38), we have the distribution for the number of customers in the system at an arbitrary time

$$\Pi(z) = \frac{(1-\rho_H)\{1 - D^*(\lambda)[F(z) - f_0 + f_0 z]U^*(\lambda - \lambda z) - (1 - D^*(\lambda))z\}H^*(\lambda - \lambda z)}{[D^*(\lambda)(E[F] + f_0 + \lambda E[U]) + 1 - D^*(\lambda)][H^*(\lambda - \lambda z) - z]}, \tag{39}$$

$$E[L] = \frac{D^*(\lambda)\{E[F(F-1)] + 2\lambda(E[F] + f_0)E[U] + \lambda^2 E[U^2]\}}{2[D^*(\lambda)(E[F] + f_0 + \lambda E[U]) + 1 - D^*(\lambda)]} + \frac{\lambda^2 E[H^2]}{2(1-\rho_H)} + \rho_H, \tag{40}$$

where

$$f_k = \Pr[k \text{ arrivals during the idle period}] = \frac{(\lambda T)^k e^{-\lambda T}}{k!}, \quad k = 0, 1, 2, \ldots, N-1$$

and

$$f_N = 1 - \sum_{k=0}^{N-1} f_k.$$

Furthermore, the p.g.f. for F is as follows

$$F(z) = \sum_{k=0}^{N} f_k z^k.$$

4.2. Waiting time distribution and expected waiting time in the queue

Let Θ_c and $\Theta_c^*(\theta)$ be defined as those in Section 3.2. By following arguments similar to *Model 1*, we have

$$\begin{aligned}\Theta_c^*(\theta) = {} & D^*(\lambda)[F(1 - (\theta + \lambda - \lambda H_o^*(\theta))/\lambda) - f_0(\theta + \lambda - \lambda H_o^*(\theta))/\lambda]U^*(\theta + \lambda - \lambda H_o^*(\theta)) \\ & + (1 - D^*(\lambda))I_o^*(\theta)H_o^*(\theta)\end{aligned} \tag{41}$$

and

$$E[\Theta_c^*] = \frac{1}{\lambda(1-\rho_H)}[D^*(\lambda)(E[F] + f_0 + \lambda E[U]) + 1 - D^*(\lambda)]. \tag{42}$$

Similar to the analysis of *Model 1*, customers that arrive during a delay cycle have the LST of the waiting time

$$W^*(\theta|\text{delay cycle}) = \frac{(1-\rho_H)\{1-[F(1-\theta/\lambda)-f_0\theta/\lambda]U^*(\theta)\}}{\{(E[F]+f_0)/\lambda+E[U]\}[\theta-\lambda+\lambda H^*(\theta)]}. \tag{43}$$

Using the expression for the LST of the waiting time depending on the interval in which a customer arrives, we obtain

$$W^*(\theta) = \frac{1}{E[\Theta_c]}\left[D^*(\lambda)\left(\frac{E[F]+f_0+\lambda E[U]}{\lambda(1-\rho_H)}\right)\times W^*(\theta|\text{delay cycle}) + (1-D^*(\lambda))\left(\frac{1}{\lambda}+\frac{E[H]}{1-\rho_H}\times W^*(\theta|\text{busy})\right)\right] \tag{44}$$

from which we get

$$W^*(\theta) = \frac{(1-\rho_H)\{D^*(\lambda)[\lambda-\lambda(F(1-\theta/\lambda)-f_0\theta/\lambda)U^*(\theta)]+\theta(1-D^*(\theta))\}}{[D^*(\lambda)(E[F]+f_0+\lambda E[U])+1-D^*(\lambda)][\theta-\lambda+\lambda H^*(\theta)]}. \tag{45}$$

From (45), we obtain the expected waiting time in the queue

$$E[W] = \frac{D^*(\lambda)\{E[F(F-1)]+2\lambda(E[F]+f_0)E[U]+\lambda^2E[U^2]\}}{2\lambda[D^*(\lambda)(E[F]+f_0+\lambda E[U])+1-D^*(\lambda)]} + \frac{\lambda E[H^2]}{2(1-\rho_H)}, \tag{46}$$

which is identical to Little's result $\left(E[W]=\frac{E[L]}{\lambda}-E[H]\right)$ obtained through (40).

4.3. Other system characteristics

4.3.1. Expected length of the completion period

Let $H^*_{NT}(\theta)$ denote the LST of the completion period for the M/G/1 queueing system, where an unreliable server is characterized by startup and closedown times. At the end of each completion period, the server deactivates through a closedown time. As soon as the number of arrivals in the queue reaches to a predetermined threshold N or T time units have elapsed since the end of the closedown time, the server immediately reactivates but is temporarily unavailable to the waiting customers. He needs a startup time before providing service until the system becomes empty. If some customers arrive at the system when the server is shutting down, he immediately provides his service without waiting the conditions of NT vacation and without a startup time. As analyzed in *Model 1*, we get

$$H^*_{NT}(\theta) = D^*(\lambda)[F(H^*_o(\theta))-f_0+f_0H^*_o(\theta)]U^*(\lambda-\lambda H^*_o(\theta))+(1-D^*(\lambda))H^*_o(\theta), \tag{47}$$

Differentiating (47) with respect to θ and using (17), we obtain the expected length of the completion period for the NT policy M/G/1 queueing system (*Model 2*) given by

$$E[H_{NT}] = \frac{\{D^*(\lambda)(E[F]+f_0+\lambda E[U])+1-D^*(\lambda)\}E[H]}{1-\rho_H}. \tag{48}$$

From (48), we have the expected length of the busy period and the expected length of the breakdown period, respectively, as follows:

$$E[B_{NT}] = \frac{\{D^*(\lambda)(E[F]+f_0+\lambda E[U])+1-D^*(\lambda)\}E[S]}{1-\rho_H} \tag{49}$$

and

$$E[D_{NT}] = \frac{\alpha\{D^*(\lambda)(E[F] + f_0 + \lambda E[U]) + 1 - D^*(\lambda)\}E[S]E[R]}{1 - \rho_H}. \tag{50}$$

4.3.2. Expected length of the idle period

Now let us find the LST $I^*_{NT}(\theta)$ of the idle period (I_{NT}) for the NT policy M/G/1 queueing system, in which the unreliable server applies the same operations described in the first paragraph of this section. It can be shown from Takagi [5, Section 2.2] that

$$I^*_{NT}(\theta) = D^*(\lambda)[F(1 - \theta/\lambda) - f_0 + f_0 I^*_{\mathrm{o}}(\theta)]U^*(\theta) + (1 - D^*(\lambda))I^*_{\mathrm{o}}(\theta) \tag{51}$$

from which we have the expected length of the idle period

$$E[I_{NT}] = D^*(\lambda)\left(\frac{E[F] + f_0}{\lambda} + E[U]\right) + \frac{1 - D^*(\lambda)}{\lambda}. \tag{52}$$

4.3.3. Expected length of the busy cycle

The busy cycle for the NT policy M/G/1 queueing system is denoted by Ω_{NT}. From (48) and (52), we have

$$E[\Omega_{NT}] = \frac{D^*(\lambda)\{E[F] + f_0 + \lambda E[U]\} + 1 - D^*(\lambda)}{\lambda(1 - \rho_H)}. \tag{53}$$

5. Special cases

In this section, we present some existing results in the literature which are special cases of our system.

Case 1. As $D^*(\lambda) = 0$ and $\alpha = 0$, *Model 1* and *Model 2* can be reduced to the ordinary M/G/1 queueing system with a reliable server. In this case, the results coincides with those of Kleinrock's system [39].

Case 2. Suppose that we have $\Pr[U = 0] = 1$ and $\alpha = 0$; then if we put $T = \infty$ and $D^*(\lambda) = 1$, *Model 1* and *Model 2* can be reduced to the N policy M/G/1 queueing system with a reliable server. In this case, the results coincides with those of Heyman's system [6].

Case 3. Suppose that we let $\Pr[U = 0] = 1$ and $\alpha = 0$; then if we put $N = \infty$ and $D^*(\lambda) = 1$, *Model 1* can be reduced to the ordinary M/G/1 queueing system with a reliable server and multiple vacations of fixed length T. In this case, our model can describe the T policy M/G/1 queueing system, Eq. (24) for $E[W]$ and (37) for $E[\Omega_{NT}]$ can be simplified to the following expressions:

$$E[W] = \frac{T}{2} + \frac{\lambda E[S^2]}{2(1 - \rho)}$$

and

$$E[\Omega_{NT}] = \frac{T}{1 - \rho},$$

which are in accordance with those of Heyman's system [3] or Takagi's system [5, Section 2.2] with multiple vacations of fixed length T.

Case 4. Suppose that we let $\Pr[U=0]=1$ and $\alpha=0$; then if we put $N=\infty$ and $D^*(\lambda)=1$, *Model 2* can be reduced to the ordinary M/G/1 queueing system with a reliable server and a single vacation of fixed length T. In this case, Eq. (46) for $E[W]$ and (53) for $E[\Omega_{NT}]$ can be simplified to the following expressions:

$$E[W]=\frac{\lambda T^2}{2(f_0+\lambda T)}+\frac{\lambda E[S^2]}{2(1-\rho)}$$

and

$$E[\Omega_{NT}]=\frac{f_0+\lambda T}{\lambda(1-\rho)},$$

which agree with those of Takagi's system [5, Section 2.2] with a single vacation of fixed length T.

Case 5. Suppose that we let $\alpha=0$; then if we put $T=\infty$ and $D^*(\lambda)=1$, *Model 1* can be reduced to the N policy M/G/1 queueing system with a reliable server and a startup. In this case, Eq. (24) for $E[W]$ can be simplified to the following expression:

$$E[W]=\frac{\lambda E[S^2]}{2(1-\rho)}+\frac{N(N-1)+2N\lambda E[U]+\lambda^2E[U^2]}{2\lambda(N+\lambda E[U])},$$

which is the same as $-\left.\frac{dW^*(\theta)}{d\theta}\right|_{\theta=0}$ for Eq. (5.1) in Takagi [40].

Case 6. Suppose that we let $\alpha=0$; then if we put $T=\infty$, *Model 1* can be reduced to the N policy M/G/1 queueing system with a reliable server, startup and closedown times. In this case, Eq. (24) for $E[W]$ can be simplified to the following expression:

$$E[W]=\frac{\lambda E[S^2]}{2(1-\rho)}+\frac{D^*(\lambda)\{N(N-1)+2N\lambda E[U]+\lambda^2E[U^2]\}}{2\lambda[D^*(\lambda)(N+\lambda E[U])+1-D^*(\lambda)]},$$

which is identical to (2.63b) given in Takagi [5, Section 2.2].

Remark 1. Note that when $f_0=0$, the results of *Model* 2 are the same as those of *Model* 1.

6. Conclusions

In this paper, we analyzed the system size distribution of an M/G/1 queueing system with an unreliable server and generalized vacations. Using the analytical results, we derived the LSTs of various system characteristics for two different kinds of NT policy M/G/1 queueing system with a startup and closedown time possibly considered. This research presents an extension of the generalized vacation model theory and the analysis of the model will provide a useful performance evaluation tool for more general situations arising in practical applications (see [32]).

The research work related to the numerical results and the further comparison of the optimized system based on this paper are more practical but complicated analytically. In the future, the work can be considering other variant dyadic policies (such as *ND* policy, *TD* policy, ...) or triadic policies. Another interesting extension of this work is to consider a control vacation policy where the server reactivates when there are more than N customers in the queue, or the server reactivates with a possible probability when the first customer has waited for at least T units of time and the queue size doesn't reach the predetermined threshold N.

References

[1] B.T. Doshi, Queueing system with vacations—a survey, Queue. Syst. 1 (1986) 29–66.
[2] M. Yadin, P. Naor, Queueing systems with a removable service station, Operation. Res. Quart. 14 (1963) 393–405.
[3] D.P. Heyman, The T policy for the M/G/1 queue, Manage. Sci. 23 (1977) 775–778.
[4] J. Teghem Jr., Optimal control of a removable server in an M/G/1 queue with finite capacity, Eur. J. Operat. Res. 31 (1987) 358–367.
[5] H. Takagi, Queueing Analysis: a foundation of performance evaluationVacation and Priority Systems, Part I, vol. I, North-Holland, Amsterdam, 1991.
[6] D.P. Heyman, Optimal operating policies for M/G/1 queueing system, Operat. Res. 16 (1968) 362–382.
[7] H.C. Tijms, Stochastic Modelling and Analysis, Wiley, New York, 1986.
[8] K.G. Gakis, H.K. Rhee, B.D. Sivazlian, Distributions and first moments of the busy and idle periods in controllable M/G/1 queueing models with simple and dyadic policies, Stochast. Anal. Applic. 13 (1) (1995) 47–81.
[9] J.R. Artalejo, Some results on the M/G/1 queue with N-policy, Asia-Pacific J. Operat. Res. 15 (1998) 147–157.
[10] K.-H. Wang, J.-C. Ke, A recursive method to the optimal control of an M/G/1 queueing system with finite capacity and infinite capacity, Appl. Math. Model. 24 (2000) 899–914.
[11] K.-H. Wang, K.-L. Yen, Optimal control of an $M/H_k/1$ queueing system with a removable server, Math. Methods Operat. Res. 57 (2003) 255–262.
[12] K.-H. Wang, Optimal control of a Markovian queueing system with a removable and non-reliable server, Microelectron. Reliab. 35 (1995) 1131–1136.
[13] K.-H. Wang, Optimal control of an $M/E_k/1$ queueing system with removable service station subject to breakdowns, J. Operat. Res. Soc. 48 (1997) 936–942.
[14] K.-H. Wang, K.-W. Chang, B.D. Sivazlian, Optimal control of a removable and non-reliable server in an infinite and a finite $M/H_2/1$ queueing system, Appl. Math. Model. 23 (1999) 651–666.
[15] K.-H. Wang, J.-C. Ke, Control policies of an M/G/1 queueing system with a removable and non-reliable server, Int. Trans. Operat. Res. 9 (2002) 195–212.
[16] K.-H. Wang, H.-T. Kao, G. Chen, Optimal management of a removable and non-reliable server in an infinite and a finite $M/H_k/1$ queueing system, Qual. Technol. Quant. Manage. 1 (2) (2004) 325–339.
[17] Y. Levy, U. Yechiali, Utilization of idle time in an M/G/1 queueing system, Manage. Sci. 22 (1975) 202–211.
[18] L. Tadj, A quorum queueing system under T-policy, J. Operat. Res. Soc. 54 (2003) 466–471.
[19] O. Kella, The threshold policy in the M/G/1 queue with server vacations, Naval Res. Logist. 36 (1989) 111–123.
[20] H.S. Lee, M.M. Srinivasan, Control policies for the $M^{[x]}/G/1$ queueing system, Manage. Sci. 35 (1989) 708–721.
[21] S.S. Lee, H.W. Lee, K.C. Chae, On a batch arrival queue with N policy and single vacation, Comput. Operat. Res. 22 (1995) 173–189.
[22] H.W. Lee, S.S. Lee, J.O. Park, K.C. Chae, Analysis of $M^{[x]}/G/1$ queue with N policy and multiple vacations, J. Appl. Prob. 31 (1994) 467–496.
[23] S.W. Fuhrmann, R.B. Cooper, Stochastic decompositions in the M/G/1 queue with generalized vacations, Operat. Res. 33 (1985) 1117–1129.

[24] J.-C. Ke, The optimal control of an M/G/1 queueing system with server vacations, startup and breakdowns, Comput. Ind. Eng. 44 (2003) 567–579.
[25] D.L. Minh, Transient solutions for some exhaustive M/G/1 queues with generalized independent vacations, Eur. J. Operat. Res. 36 (1988) 197–201.
[26] J. Medhi, J.G.C. Templeton, A Poisson input queue under *N*-policy and with a general start up time, Comput. Operat. Res. 19 (1) (1992) 35–41.
[27] H. Takagi, M/G/1/K queues with *N*-policy and setup times, Queue. Syst. 14 (1993) 79–98.
[28] H.W. Lee, J.O. Park, Optimal strategy in *N*-policy production system with early set-up, J. Operat. Res. Soc. 48 (1997) 306–313.
[29] S. Hur, S.J. Paik, The effect of different arrival rates on the *N*-policy of M/G/1 with server setup, Appl. Math. Model. 23 (1999) 289–299.
[30] Z. Niu, Y. Takahashi, A finite-capacity queue with exhaustive vacation/close-down/setup times and Markovian arrival processes, Queue. Syst. 31 (1999) 1–23.
[31] Y.N. Doganata, *NT*-vacation policy for M/G/1 queue with starter, in: E. Arikan (Ed.), Communication, Control, and Signal Processing, Elsevier Science, Amsterdam, 1990, pp. 1663–1669.
[32] A.S. Alfa, I. Frigui, Discrete *NT*-policy single server queue with Markovian arrival process and phase type service, Eur. J. Operat. Res. 88 (1996) 599–613.
[33] A.S. Alfa, W. Li, Optimal (N,T)-policy for M/G/1 system with cost structures, Perform. Evaluat. 42 (2000) 265–277.
[34] W. Li, A.S. Alfa, Optimal policies for M/M/m queue with two different kinds of (N,T)-policies, Naval Res. Logist. 47 (2000) 240–258.
[35] S. Hur, J. Kim, C. Kang, An analysis of the M/G/1 system with *N* and *T* policy, Appl. Math. Model. 27 (2003) 665–675.
[36] Z.G. Zhang, R. Vickson, E. Love, The optimal service policies in an M/G/1 queueing system with multiple vacation types, INFOR 39 (2001) 357–366.
[37] D.P. Gaver, A waiting line with interrupted service, including priorities, J. R. Statist. Soc., Ser. B 24 (1962) 73–96.
[38] Y.H. Tang, A single-server M/G/1 queueing system subject to breakdowns-some reliability and queueing problem, Microelectron. Reliab. 37 (2) (1997) 315–321.
[39] L. Kleinrock, Queueing Systems: Theory, vol. I, John Wiley & Sons, Inc., New York, 1975.
[40] H. Takagi, Time-dependent process of M/G/1 vacation models with exhaustive service, J. Appl. Prob. 29 (1992) 418–429.