Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Vol. 8, Issue 2 (December 2013), pp. 346 – 365

Applications and Applied
Mathematics:
An International Journal
(AAM)

The $M^X/G/1$ Queue with Unreliable Server, Delayed Repairs, and Bernoulli Vacation Schedule under T-Policy

L. Tadj
Department of Information Systems and Decision Sciences
Fairleigh Dickinson University
Vancouver, BC, V6B 2P6 Canada
ltadj@fdu.edu

G. Choudhury
Mathematical Sciences Division
Institute of Advanced Study in Science and Technology
Guwahati – 781035 Assam, India
gc.iasst@gmail.com

Received: August 27, 2013; Accepted: November 6, 2013

Abstract

In this paper we study a batch arrival queuing system. The server may break down while delivering service. However, repair is not provided immediately, rather it is delayed for a random amount of time. At the end of service, the server may process the next customer if any are available, or may take a vacation to execute some other job. Finally, the server implements the T-policy. We describe for this system an optimal management policy. Numerical examples are provided.

Keywords: Queue, breakdown, vacation, optimal management policy

AMS-MSC 2010 No.: 60K10, 60K25

1. Introduction

One of the fundamental concerns of queuing systems managers is to gain control over their system. Such control is difficult since many elements of a queuing system are uncertain. The area of queuing theory concerned with the design of an optimal management policy has received much attention from researchers, and a huge number of models have been studied. This is because applications are widely encountered in diverse practical situations such as telecommunication and computer networks, manufacturing systems, transportations systems, etc.

The purpose of this paper is to design an optimal management policy for a system where the server implements the $\boldsymbol{T}$-policy. To cater for situations where the server cannot monitor its queue continuously, Heyman (1977) introduces, in the context of an $M/G/1$ queue, the $\boldsymbol{T}$-policy which activates the server $\boldsymbol{T}$ time units after the end of the last busy period. Under a linear cost structure, Heyman (1977) obtained the optimal value of $\boldsymbol{T}$ and the optimal cost rate.

Tadj (2003) studies a bulk service queue under $\boldsymbol{T}$-policy. Given a fixed threshold $\boldsymbol{r}$ $(\boldsymbol{r \geq 1})$, the server checks the number of customers in the queue at the end of a service. If more than $\boldsymbol{r}$ customers are present, the server picks a batch of $\boldsymbol{r}$ customers to process simultaneously. Otherwise, the server leaves and scans the queue $\boldsymbol{T}$ time units after the end of the previous service. He keeps scanning the queue every $\boldsymbol{T}$ units of time and resumes a bulk service only when $\boldsymbol{r}$ customers are available in the queue. Tadj (2003a) derives the steady-state distribution for the system size and some performance measures. Tadj (2003b) generalizes Tadj (2003a) to the case where arrivals are batch Poisson and the server capacity is bounded by some other threshold $R, (R \geq r \geq 1)$.

Analysis of the $\boldsymbol{T}$-policy $M/G/1$ queue is extended by Wang et al. (2009a) to the case where the server is subject to breakdowns, and startup times are required at the beginning of a busy period. They obtain various system performance measures and determine the optimal threshold $\boldsymbol{T}$. The same system is analyzed by Wang et al. (2009b) using the maximum entropy approach. They show that this approach is accurate enough for practical purposes and is a useful method for solving complex queuing systems by using the first moment instead of the second moment of the number of customers in the system. Zhang et al. (2011) clarify the concept of regeneration cycle used in evaluating the average operating cost of the $M/G/1$ queue with $\boldsymbol{T}$-policy. Two ways of defining the regeneration cycle are compared and advantages and disadvantages of each way are pointed out. Finally, Choudhury and Tadj (2013) study the unreliable $M/G/1$ queue under $\boldsymbol{T}$-policy and Bernoulli vacation schedule. Also, they assume that, when a breakdown occurs, repair of the server is not instantaneous. An optimal management policy is formulated for the system.

Those are the only papers that deal exclusively with the T-policy. A set of papers deals with what are called *randomized policies*. These are policies where one of two possible scenarios may happen:

(i) control the server randomly at the beginning epoch of the service, when at least one customer appears; or
(ii) at the end of each busy period, the server can be either switched off or kept on with some positive probability.

Randomized policies were considered by Kim and Moon (2006), Ke and Chu (2008), Yang et al. (2008), Ke and Chu (2009), Wang et al. (2010), and Wang et al. (2012).

Another set of papers combines the N-policy and the T-policy into what is called the *NT-policy*. Two scenarios also have been considered:

(i) The server goes on vacation as soon as he/she completes service and the system is empty and termination of the vacation period is controlled by two threshold parameters N and T, i.e. the server terminates his/her vacation as soon as the number waiting reaches N or the waiting time of the leading customer reaches T units. This type of NT-policy was considered by Alfa and Frigui (1996), Alfa and Li (2000), and Li and Alfa (2000).

(ii) The server's vacation terminates if either N 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, whichever occurs first. This type of NT-policy was considered by Doganata (1990), Gakis et al. (1995), Wang and Ke (2002), Hur et al. (2003), Ke (2006a), Ke (2006b), Ke et al. (2009), Jiang et al. (2010), and Feyaerts et al. (2012).

Finally, some variants of the $\boldsymbol{T}$-policy have been proposed by some researchers, in which the threshold $\boldsymbol{T}$ is combined with some other threshold level for a better control over the service system. For examples of such papers, see Sen and Gupta (1994), Ke (2005), and Ke (2008).

The purpose of this paper is to study the $\boldsymbol{T}$-policy in a bulk arrival queue. Batch arrival models have been extensively used in an uncountable number of papers. A topical textbook is that of Chaudhury and Templeton (1983).

We also assume that server is unreliable. Indeed, a server, such as a machine, may break down while providing service. The service of the customer being served is then interrupted and cannot resume until the server is repaired. This is a very realistic assumption that models real-life situations. Madan et al. (2003) study two models in which the service facility suffers time homogeneous random breakdowns from time to time. For a review of the research on this topic the reader is referred to the broad survey of Tadj et al. (2012).

There are also situations where an out-of-order machine cannot be repaired immediately, and repair takes place only after some random time. For this reason, we are assuming that once service is interrupted due to failure, repair of the server is delayed and takes place only after some random time.

Finally, we are assuming that the server may take vacations. For a review of vacation queuing systems, see the comprehensive surveys of Doshi (1986, 1990).

In this class of queueing models, we find the models with Bernoulli vacation schedule where, if the queue is empty at a service completion, then the server becomes inactive and begins an idle period. However, if the queue is not empty the server will choose randomly to either take a vacation or process the next customer. If a vacation is chosen, it is followed by the service of the customer at the head of the queue. Various aspects of Bernoulli vacation models have been discussed by a number of authors; see the survey of Ke et al. (2010). In real-life, the vacation period may be used by the server to perform some other activities such as a maintenance operation, a quality control, or even attend another queue of customers.

Therefore, in this paper, we will be considering the compound Poisson arrival queueing system where the server is prone to failure, repairs preceding repairs are delayed, and the server implements a Bernoulli vacation schedule and a T-policy. This model generalizes Choudhury and Tadj (2013) to the case of batch arrivals.

Note that many real-world service systems have batch arrivals. For an example, we cite Haghighi and Mishev (2013) who use a compound Poisson process to model the arrival of job applications in a hiring process at a large company.

The rest of the paper proceeds as follows. The next section reviews the literature relevant to our queueing system. In the following two sections, we first review the analysis of the classical $M^X/G/1$ queue under $\boldsymbol{T}$-policy in Section 2. Then, in Section 3, we extend the results obtained to the $M^X/G/1$ queue with unreliable server, delayed repair, and Bernoulli vacation schedule under $\boldsymbol{T}$-policy. We design an optimal management policy in Section 4, and in Section 5 we present some numerical examples. We conclude the paper in Section 6.

2. The $\boldsymbol{M^X/G/1}$ Queueing System under $\boldsymbol{T}$-Policy

We briefly summarize the analysis of the $M^X/G/1$ queueing system under T-policy. We note that these results have not been reported in the literature.

The input process is a compound Poisson arrival process with positive rate λ. Let X represent the size of an arriving group of customers. We will denote by $a(z) = E[z^X]$ the probability generating function (PGF) of X and by a_1 and a_2 its first and second moments, respectively. The waiting room is infinite and the service discipline is first-in, first-out (FIFO). The single server takes a vacation if a customer departure leaves the system empty, in accordance with the T-policy. Thus, the server scans the queue T time units after the end of the last busy period (the beginning of the idle period). If at least one unit is found in the queue, a busy period starts and service is exhaustive. Otherwise, if no unit is found in the queue, another vacation starts, and a new scan is made after T units of time, and so on.

Let B denote the duration of the service provided by the server to a customer. The service times are assumed to be a sequence of independent, identically distributed random variables with cumulative distribution function (CDF) $B(t)$, Laplace-Stieltjes Transform (LST) $B^*(\theta) = E\left[e^{-\theta B}\right]$, and first and second moments b_1 and b_2, respectively.

2.1. Preliminaries

Suppose that after $\nu - 1$ vacations, the server finds no customer in the queue and after ν vacations, he finds Y_ν customers in the queue. Let $q(z) = e^{-\lambda T(1-a(z))}$ denote the PGF of a compound Poisson random variable. The distributions of the random variables ν and Y_ν are needed for the rest of the analysis and are derived in the following

Lemma 1.

The PGFs of ν and Y_ν are given by

$$\eta(z) := E[z^\nu] = \frac{z[1-q(0)]}{1-zq(0)}. \tag{1}$$

$$L(z) := E[z^{Y_\nu}] = \frac{q(z)-q(0)}{1-q(0)}. \tag{2}$$

It follows from (1) and (2) that the mean number of vacations the server has to take before finding at least one customer in the queue and the mean number of customer arrivals during the last vacation of length T are respectively given by

$$E[\nu] = \left.\frac{d\eta(z)}{dz}\right|_{z=1} = \frac{1}{1-e^{-\lambda T}}. \tag{3}$$

$$E[Y_\nu] = \left.\frac{dL(z)}{dz}\right|_{z=1} = \frac{\lambda T a_1}{1-e^{-\lambda T}}. \tag{4}$$

2.2. Queue Size Distribution at a Departure Epoch

Now, let $Q(t)$ denote the number of customers in the system at an arbitrary instant of time t and let Q_n be the number of customers in the system at the completion epoch T_n of the nth service, that is, $Q_n = Q(T_n^+)$. Also, let V_{n+1} represent the number of customer arrivals during the $(n+1)$st service. Then, Q_n satisfies the recursion

$$Q_{n+1} = \begin{cases} Q_n - 1 + V_{n+1}, & \text{if} \quad Q_n > 0, \\ Y_\nu - 1 + V_{n+1}, & \text{if} \quad Q_n = 0. \end{cases} \tag{5}$$

Therefore, the process $\{Q_n;\ n = 0,1,\cdots\}$ is a homogeneous Markov chain. Its transition probability matrix (TPM) is a Δ–matrix. Abolnikov and Dukhovny (1991) have shown that a Markov chain, whose TPM is a Δ–matrix, is ergodic if and only if

$$\left.\frac{dK(z)}{dz}\right|_{z=0} < 1, \tag{6}$$

where

$$K(z) := E[z^{V_{n+1}}] = B^*\big(\lambda - \lambda a(z)\big), \tag{7}$$

represents the PGF of the number of customer arrivals during the service time B. Denote by $p_i = \lim_{n\to\infty} P\{Q_n = i\}$ the steady-state distribution and by $P_T(z) = E[z^{Q_n}]$ the PGF of $\{Q_n;\ n = 0,1,\cdots\}$ in the steady-state.

Theorem 1.

Let $\rho_T = \lambda a_1 b_1$ denote the server utilisation. Then, the Markov chain $\{Q_n;\ n = 0,1,\cdots\}$ is ergodic if and only if

$$\rho_T < 1, \tag{8}$$

and the PGF $P_T(z)$ is given by

$$P_T(z) = \frac{1-\lambda a_1 b_1}{\lambda T a_1} \times \frac{B^*\big(\lambda - \lambda a(z)\big)\big[e^{-\lambda T(1-a(z))} - 1\big]}{z - B^*\big(\lambda - \lambda a(z)\big)}. \tag{9}$$

Proof:

Condition (8) is equivalent to condition (6). Using the recursive relation (5), we have

$$P_T(z) = \frac{K(z)[L(z)-1]p_0}{z - K(z)} \tag{10}$$

where $L(z)$ is given by (2) and $K(z)$ by (7). Finally, L'Hôpital's rule yields

$$p_0 = \frac{1-\lambda a_1 b_1}{E[Y_v]}, \tag{11}$$

where $E[Y_v]$ is given by (4). Substituting for the different expressions yields the result (9).

□

Therefore, in the rest of this section, we will assume that condition (8) is satisfied so that the Markov chain $\{Q_n; n = 0,1,\cdots\}$ is ergodic and the steady-state distribution for the number of customers in the system exists.

2.3. Special Case, Decomposition Property

Note that for an $M^X/G/1$ queueing system without T-policy, see for example Chaudhry and Templeton (1983), the PGF $P_{(M^X/G/1)}(z)$ is given by

$$P_{(M^X/G/1)}(z) = \frac{1-\lambda a_1 b_1}{a_1} \times \frac{B^*\big(\lambda - \lambda a(z)\big)[a(z)-1]}{z - B^*\big(\lambda - \lambda a(z)\big)}, \tag{12}$$

so that the expression (9) can be decomposed as

$$P_T(z) = P_{(M^X/G/1)}(z) \times \frac{e^{-\lambda T[1-a(z)]} - 1}{\lambda T[a(z)-1]}. \tag{13}$$

On one hand, this decomposition confirms the decomposition property stated by Fuhrmann and Cooper (1985). On the other hand, taking the limit as $T \to 0$, the second term on the right hand tends to 1, so that the result (9) agrees with expression (3.4.9) of Chaudhury and Templeton (1983).

2.4. Queue Size Distribution at an Arbitrary Point in Time

The process $\{Q(t);\ t \geq 0\}$ is a semi-regenerative process relative to the sequence $\{T_n;\ n = 0,1,\cdots\}$ of service completions epochs. Denote by $\pi_T(z)$ the PGF in the steady-state of $\{Q(t);\ t \geq 0\}$ and by $\pi_i = \lim_{t\to\infty} P\{Q(t) = i\}$ the steady-state distribution.

Theorem 2.

The semi-regenerative process $\{Q(t);\ t \geq 0\}$ is ergodic if and only if

$$\rho_T < 1, \tag{14}$$

and the PGF $\pi(z)$ is given by

$$\pi_T(z) = a_1 \frac{1-z}{1-a(z)} P_T(z). \tag{15}$$

Proof:

Direct application of the main convergence theorem for semi-regenerative processes, see e.g., Çinlar (1975, p. 347). □

In particular, relation (15) yields

$$\pi_0 = a_1 p_0. \tag{16}$$

2.5. Special case, decomposition property

Note that for an $M^X/G/1$ queueing system without T-policy, the PGF $\pi_{(M^X/G/1)}(z)$ is given by

$$\pi_{(M^X/G/1)}(z) = a_1 \frac{1-z}{1-a(z)} P_{(M^X/G/1)}(z). \tag{17}$$

Combining (13), (15), and (17) we have the decomposition

$$\pi_T(z) = \pi_{(M^X/G/1)}(z) \times \frac{e^{-\lambda T[1-a(z)]} - 1}{\lambda T[a(z) - 1]}. \tag{18}$$

Again, this decomposition confirms the decomposition property stated by Fuhrmann and Cooper (1985) and shows that as $T \to 0$, expression $\pi_T(z)$ coincides with $\pi_{(M^X/G/1)}(z)$, as it ought to be.

2.6. Idle Period Distribution

Let I_T and $I_T^*(\theta)$ be the idle period random variable and its LST, respectively. Then, from the theory of fluctuating processes, see Abolnikov et al. (1994), it can be shown that

$$I_T^*(\theta) = \frac{e(\theta)\left(1 - e^{-\lambda T}\right)}{1 - e(\theta)e^{-\lambda T}}. \tag{19}$$

Here $e(\theta)$ represents the LST of a vacation period. Since the server takes vacations of constant length T, then

$$e(\theta) = e^{-\theta T}. \tag{20}$$

2.7. Busy Period Distribution

We define a busy period as the interval of time that keeps the server busy and this goes on until the first subsequent instant when the system becomes empty. Let B_T and $B_T^*(\theta)$ be the busy period random variable and its LST respectively. Then, from expression (2), we have

$$B_T^*(\theta) = L\left(\varpi^*(\theta)\right)\frac{e^{-\lambda T\left[1 - a\left(\varpi^*(\theta)\right)\right]} - e^{-\lambda T}}{1 - e^{-\lambda T}}, \tag{21}$$

where $\varpi^*(\theta)$ is the LST of the busy period of an ordinary $M^X/G/1$ queue satisfying Takács's functional equation

$$\varpi^*(\theta) = B^*\left(\theta + \lambda - \lambda a\left(\varpi^*(\theta)\right)\right). \tag{22}$$

2.8. Waiting Time Distribution

Let $W_Q^*(\theta)$ be the LST of the waiting time distribution, then by applying the stochastic decomposition property for waiting time distribution, we may write

$$W_Q^*(\theta) = W_0^*(\theta)W_{(M^X/G/1)}^*(\theta). \tag{23}$$

where $W_0^*(\theta)$ is the LST of the residual time of an idle period, i.e., the amount of time spent in the queue for the batches (behaves like a single customer) which arrive during the idle period. This is equivalent to

$$W_0^*(\theta) = \frac{1 - e^{-\theta T}}{\theta T}, \tag{24}$$

and $W^*_{(M^X/G/1)}(\theta)$ is the LST of the waiting time distribution for an $M^X/G/1$ system without vacation. This is equivalent to (see Chaudhry and Templeton (1983), Chapter 3)

$$W^*_{(M^X/G/1)}(\theta) = \frac{(1-\rho)\theta}{\theta - \lambda - \lambda a\big(B^*(\theta)\big)} \times \frac{1 - a\big(B^*(\theta)\big)}{a_1[1 - B^*(\theta)]}, \tag{25}$$

Hence, from the above equations (23)-(25), we obtain

$$W^*_Q(\theta) = \frac{(1-\rho_T)\big(1-e^{-\theta T}\big)\big[1 - a\big(B^*(\theta)\big)\big]}{a_1 T\big[\theta - \lambda + \lambda a\big(B^*(\theta)\big)\big][1 - B^*(\theta)]}. \tag{26}$$

2.9. System Performance Measures

We derive here various characteristics for the $M^X/G/1$ queueing system under T-policy.

(i) The expected number of customers in the system at an arbitrary instant of time $L_T = \sum_{i=1}^{\infty} i\pi_i$ is obtained by taking the first derivative of the PGF $\pi_T(z)$, given in (15), with respect to z and setting $z = 1$. We obtain

$$L_T = L_{(M^X/G/1)} + \frac{\lambda T a_1}{2} - \frac{a_2 - a_1}{2}, \tag{27}$$

where $L_{(M^X/G/1)}$ is the expected number of customers in the classical $M^X/G/1$ queueing system (i.e., without T-policy) and is given by

$$\begin{aligned} L_{(M^X/G/1)} = \frac{1}{2(1-\lambda a_1 b_1)}\Big\{&(2\lambda a_1^2 b_1 + a_2 - a_1)\left(\frac{1-\lambda a_1 b_1}{a_1}\right) \\ &+ \lambda[(a_2 - a_1)b_1 + \lambda a_1^2 b_2]\Big\}. \end{aligned} \tag{28}$$

(ii) Similarly, using (19), the mean idle period is given by

$$E[I_T] = -\frac{dI^*_T(\theta)}{d\theta}\bigg|_{\theta=0} = \frac{T}{1-e^{-\lambda T}}. \tag{29}$$

(iii) Also, using (21) and (22), the mean delayed busy period is found to be

$$E[B_T] = -\frac{dB^*_T(\theta)}{d\theta}\bigg|_{\theta=0} = \frac{\rho_T}{1-\rho_T} \times \frac{T}{1-e^{-\lambda T}}. \tag{30}$$

(iv) Finally, the mean busy cycle $E[C_T] = E[I_T] + E[B_T]$ is found to be

$$E[C_T] = \frac{1}{1-\rho_T} \times \frac{T}{(1-e^{-\lambda T})}. \tag{31}$$

(v) Using (29)-(31), we find that the probabilities P_I and P_B that the server is idle and busy, are respectively given by

$$\begin{aligned} P_I &= \frac{E[I_T]}{E[C_T]} = 1-\rho_T, \\ P_B &= \frac{E[B_T]}{E[C_T]} = \rho_T. \end{aligned} \tag{32}$$

(vi) An interesting performance measure is the system intensity denoted by $\mathcal{I}$ and defined by

$$\mathcal{I} = \lambda a_1 \boldsymbol{P\beta}, \tag{33}$$

where $\boldsymbol{P} = (p_0, p_1, \cdots)$ denotes the vector of the system state steady-state probabilities and the vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots)$ where $\beta_i = E[T_{n+1} - T_n \mid Q_n = i]$ represents the conditional expected duration between two consecutive service completion epochs. The quantity $\boldsymbol{P\beta}$ is called the mean service cycle. Since

$$\beta_i = \begin{cases} TE[v] + b_1, & \text{if} \quad i = 0, \\ \quad b_1, & \text{if} \quad i > 0, \end{cases} \tag{34}$$

then, the mean service cycle is given by

$$\boldsymbol{P\beta} = \frac{1}{\lambda a_1}, \tag{35}$$

so that the system intensity (30) becomes

$$\mathcal{I} = 1, \tag{36}$$

which is equal to the server load.

3. The $M^X/G/1$ queue with Unreliable Server, Delayed Repair, and Bernoulli Vacation Schedule under T-Policy

It is interesting to note that, using a modified service time distribution, it is possible to obtain all the results of the $M^X/G/1$ queue with unreliable server, delayed repairs, and Bernoulli vacation schedule under $\boldsymbol{T}$-policy from the results of the classical $M^X/G/1$ system under $\boldsymbol{T}$-policy.

3.1. Notation

Before stating the main results, let us introduce some notation to describe the new system. As in the previous model, the service times are assumed to be a sequence of independent, identically distributed random variables with CDF $\boldsymbol{B(t)}$, LST $\boldsymbol{B^*(\theta)}$, and first and second moments $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$.

Now we are assuming that the server is unreliable, and the time to failure of the server has an exponential distribution with positive rate α. When service interruption occurs, a delay period takes place, as repair may not be undertaken immediately. A repair follows the delay period, and as soon as the repair is completed, the server immediately returns to provide his service. We are assuming throughout that a pre-empted service is resumed when the service interruption is over.

Let D denote the duration of the delay time. The delay times form a sequence of independent, identically distributed random variables with CDF $D(t)$, LST $D^*(\theta)$, and first and second moments d_1 and d_2.

Let G denote the duration of the repair time. The repair times are assumed to be a sequence of independent, identically distributed random variables with CDF $G(t)$, LST $G^*(\theta)$, and first and second moments g_1 and g_2.

Let H denote the duration of the generalized service of a customer, which includes the actual service time and possible delays and repairs. Then, the LST $H^*(\theta)$ of H is given by

$$\begin{aligned} H^*(\theta) &= \sum_{n=0}^{\infty}\int_0^{\infty} e^{-\theta x}e^{-\alpha x}\frac{(\alpha x)^n}{n!}[D^*(\theta)G^*(\theta)]^n dB(x) \\ &= B^*\left(\theta+\alpha\left(1-D^*(\theta)G^*(\theta)\right)\right). \end{aligned} \tag{37}$$

Also, the first and second moments h_1 and h_2 of H are given by

$$\begin{aligned} h_1 &= b_1[1+\alpha(d_1+g_1)], \\ h_2 &= b_2[1+\alpha(d_1+g_1)]^2+\alpha b_1(d_2+2d_1g_1+g_2). \end{aligned} \tag{38}$$

Finally, the server implements the Bernoulli vacation schedule, so that at the end of a service, the server may take a vacation with some positive probability p or may remain working with the complementary probability $q = 1-p$. Let F denote the duration of the *modified service time* of a customer, which includes the generalized service time and a possible vacation, then

$$F = \begin{cases} H+T, & \text{with probability} \quad p, \\ H, & \text{with probability} \quad q=1-p. \end{cases} \tag{39}$$

The LST $F^*(\theta)$ of F satisfies

$$F^*(\theta) = \left(q+pe^{-\theta T}\right)H^*(\theta), \tag{40}$$

while the first and second moments f_1 and f_2 of F are given by

$$\begin{aligned} f_1 &= h_1+pT, \\ f_2 &= h_2+2h_1pT+pT^2. \end{aligned} \tag{41}$$

3.2. Main Results

The process $\{Q_n;\ n = 0,1,\cdots\}$ satisfies again the recursion (5), however, now V_{n+1} represents the number of arrivals during the modified service time F, so that

$$E[z^{V_{n+1}}] = F^*\big(\lambda - \lambda a(z)\big). \tag{42}$$

Denote by $P(z)$ and $\pi(z)$ the PGFs of $\{Q_n;\ n = 0,1,\cdots\}$ and $\{Q(t);\ t \geq 0\}$ in the steady-state. We present the generalized results in the following

Theorem 3.

Let $\rho = \lambda a_1\{b_1[1 + \alpha(d_1 + g_1)] + pT\}$.

(i) The Markov chain $\{Q_n;\ n = 0,1,\cdots\}$ and the semi-regenerative process $\{Q(t);\ t \geq 0\}$ are ergodic if and only if

$$\rho < 1, \tag{43}$$

and the PGFs $P(z)$ and $\pi(z)$ are given by

$$P(z) = \frac{1 - \lambda a_1 f_1}{\lambda T f_1} \times \frac{[q + pe^{-\lambda(z)T}]B^*\Big(\lambda(z) + \alpha\Big(1 - D^*\big(\lambda(z)\big)G^*\big(\lambda(z)\big)\Big)\Big)\left[e^{-\lambda(z)T} - 1\right]}{z - [q + pe^{-\lambda(z)T}]B^*\Big(\lambda(z) + \alpha\Big(1 - D^*\big(\lambda(z)\big)G^*\big(\lambda(z)\big)\Big)\Big)}, \tag{44}$$

$$\pi(z) = a_1 \frac{1 - z}{1 - a(z)} P(z), \tag{45}$$

where $\lambda(z) = \lambda - \lambda a(z)$. Also,

$$p_0 = \frac{1 - \lambda a_1\{b_1[1 + \alpha(d_1 + g_1)] + pT\}}{\lambda T a_1} \times \big(1 - e^{-\lambda T}\big), \tag{46}$$

$$\pi_0 = a_1 p_0. \tag{47}$$

(ii) Let I and B represent the idle period and busy period random variables. Then, their LSTs $I^*(\theta)$ and $B^*(\theta)$ are given by

$$I^*(\theta) = \frac{e^{-\theta T}\big(1 - e^{-\lambda T}\big)}{1 - e^{-(\theta - \lambda)T}}, \tag{48}$$

$$B_T^*(\theta) \quad = \quad \frac{e^{-\lambda T[1-a(\varpi^*(\theta))]} - e^{-\lambda T}}{1-e^{-\lambda T}}, \tag{49}$$

where $\varpi^*(\theta)$ is the LST of the busy period of an ordinary $M^X/G/1$ queue with one unit by modified service time as original service time i.e., under Bernoulli vacation schedule, which satisfies the well-known Tacak's functional equation

$$\varpi^*(\theta) = (q + pe^{-\theta T})H^*\left(\theta + \lambda - \lambda a\big(\varpi^*(\theta)\big)\right). \tag{50}$$

(iii) Let $W_Q^*(\theta)$ be the LST of the waiting time distribution. Then we have

$$W_Q^*(\theta) = \frac{(1-\rho_T)\big(1-e^{-\theta T}\big)\big[1-a\big(F^*(\theta)\big)\big]}{a_1 T\big[\theta - \lambda + \lambda a\big(F^*(\theta)\big)\big][1-F^*(\theta)]}, \tag{51}$$

where $F^*(\theta) = \big(q + pe^{-\theta T}\big)H^*(\theta)$. This can be obtained by applying the stochastic decomposition property of the waiting time distribution for this type of model.

3.3. System Performance Measures

We derive here various characteristics for the $M^X/G/1$ queueing system with unreliable server and Bernoulli vacation schedule under T-policy.

(i) The expected number of customers in the system at an arbitrary instant of time L is found to be

$$\begin{aligned} L = \frac{1}{2(1-\lambda a_1 f_1)}&\left\{(2\lambda a_1^2 f_1 + a_2 - a_1)\left(\frac{1-\lambda a_1 f_1}{a_1}\right)\right. \\ &\left. + \lambda[(a_2 - a_1)f_1 + \lambda a_1^2 f_2]\right\} + \frac{\lambda T a_1}{2} - \frac{a_2 - a_1}{2}, \end{aligned} \tag{52}$$

where f_1 and f_2 are given by (41).

(ii) The mean idle period is given by

$$E[I] = -\frac{dI^*(\theta)}{d\theta}\bigg|_{\theta=0} = \frac{T}{1-e^{-\lambda T}}. \tag{53}$$

(iii) The mean delayed busy period is found to be

$$E[B] = -\frac{dB^*(\theta)}{d\theta}\bigg|_{\theta=0} = \frac{[\rho_T + \lambda a_1 b_1(d_1 + g_1)]T}{(1-\rho)(1-e^{-\lambda T})}. \tag{54}$$

(iv) The mean busy cycle $E[C] = E[I] + E[B]$ is given by

$$E[C] \quad = \quad \frac{T}{(1-\rho)(1-e^{-\lambda T})}. \tag{55}$$

(v) The probabilities P_I and P_B that the server is idle and busy, are respectively given by

$$\begin{aligned} P_I &= \frac{E[I]}{E[C]} = 1 - \rho_T - \lambda a_1 b_1 (d_1 + g_1), \\ P_B &= \frac{E[B]}{E[C]} = \rho_T + \lambda a_1 b_1 (d_1 + g_1). \end{aligned} \tag{56}$$

(vi) The mean service cycle is given by

$$\boldsymbol{P\beta} = \frac{1}{\lambda a_1}. \tag{57}$$

(vii) The system intensity $\mathcal{I}$ equals the server load

$$\mathcal{I} = 1. \tag{58}$$

4. Optimal Management Policy

Having obtained the required performance measures, we now are ready to develop a management policy for the system. The length T of the vacation period will be our decision variable. Denote by $TC(T)$ the total expected cost per unit of time. Then,

$$TC(T) = c_h L + c_o P_B + c_a P_I + c_s \frac{1}{E[C]}, \tag{59}$$

where c_h is the holding cost per unit for each customer present in the system, c_o is the cost per unit time for keeping the server on and in operations, c_a is the startup cost per unit time for the preparatory work of the server before starting service, and c_s is the setup cost per busy cycle. Using the previous expressions and the measures derived in the previous section, we have

$$\begin{aligned} TC(T) = c_h \Bigg(& \frac{1}{2[1-\lambda a_1(h_1+pT)]} \Bigg\{ [2\lambda a_1^2 (h_1 + pT) + a_2 \\ & - a_1] \left[\frac{1 - \lambda a_1 (h_1 + pT)}{a_1} \right] \\ & + \lambda[(a_2 - a_1)(h_1 + pT) + \lambda a_1^2 (h_2 + 2h_1 pT + pT^2)] \Bigg\} + \frac{\lambda T a_1}{2} \\ & - \frac{a_2 - a_1}{2} \Bigg) + c_o[\rho_T + \lambda a_1 b_1 (d_1 + g_1)] \\ & + c_a[1 - \rho_T - \lambda a_1 b_1 (d_1 + g_1)] \\ & + c_s \frac{\{1 - [\lambda a_1 (h_1 + pT)]\}(1 - e^{-\lambda T})}{T}. \end{aligned} \tag{60}$$

The optimal value T^* of T is obtained by solving

$$\frac{dTC(T)}{dT} = 0.$$

We also need to check the second order optimality condition

$$\left.\frac{d^2TC(T)}{dT^2}\right|_{T=T^*} > 0.$$

5. Numerical Example

We present some illustrative examples in the section. Note that the total expected cost per unit of time only requires first and second moments of the different distributions involved.

Let us assume, for example, that the service time, delay time, and repair time follow the exponential distribution, so that $b_2 = 2b_1^2, d_2 = 2d_1^2$, and $g_2 = 2g_1^2$ Also, let us assume that customers arrive to the system in batches whose size follows the shifted geometric distribution $a_k = P(X = k) = (1-c)\varepsilon^{n-1}, n \geq 1$, so that $a(z) = E[z^X] = \frac{(1-\varepsilon)z}{1-\varepsilon z}$, $a_1 = \frac{1}{1-\varepsilon}$, and $a_2 = \frac{1+\varepsilon}{(1-\varepsilon)^2}$.

For the purpose of calculations, we take the following values for the non-monetary system parameters: $\lambda = 0.01, \varepsilon = 0.2, \alpha = 0.05, p = 0.15, b_1 = 0.7, d_1 = 0.3$, and $g_1 = 0.5$; and the following values for the nonmonetary system parameters: $c_h = 2, c_o = 50, c_a = 50$, and $c_s = 300$. The variations of the total expected cost per unit of time are depicted in Figure 1 below.

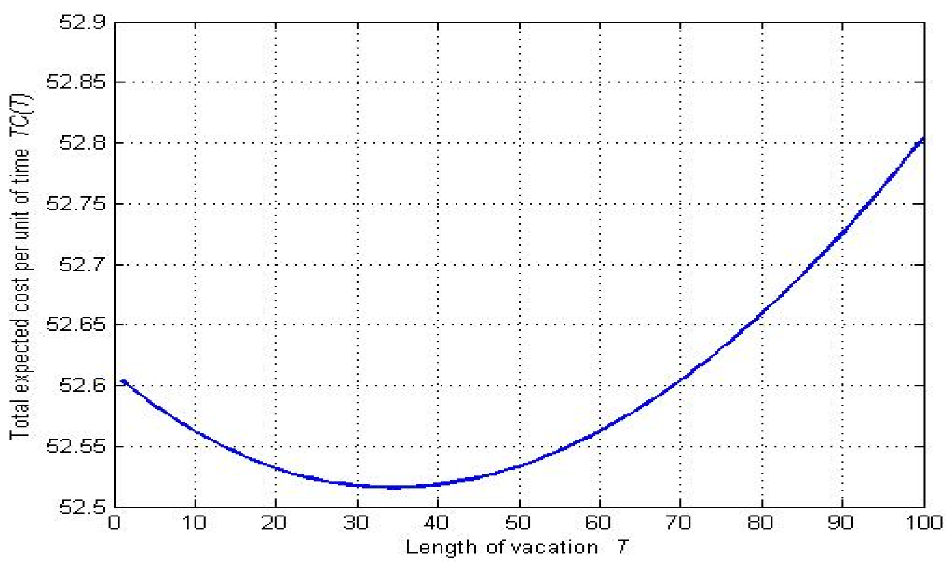

Figure 1: Variations of $TC(T)$ for shifted geometric arrival batch.

The optimal cost is $TC(T^*) = 52.5158$ and it is reached at the optimal threshold level $T^* = 34$, which represents the length of the vacation that the server should take performing auxiliary work each time a vacation is taken.

The next question one might ask is how sensitive is the optimal policy to the system parameters. Specifically, we have assumed that the design parameters can be determined by statistical methods to such a precision that the insecurity about the actual value is negligible. To compensate for the possibility of misjudgement of the actual value of the parameter, sensitivity analysis is carried out so that should the model be sensitive with respect to the value of the parameter, a more thorough statistical analysis of the parameter is demanded.

We first look at the monetary system parameters. Keeping all parameters at their original values, we changed each unit cost, one at a time, to obtain Tables 1 – 4 below.

Table 1: Effect of c_h on optimal policy

c_h	2	3	4	5	6	7	8	9	10
T^*	34	1	1	1	1	1	1	1	1
$TC(T^*)$	52.52	52.43	52.26	52.08	51.91	51.73	51.56	51.38	51.21

Table 2: Effect of c_o on optimal policy

c_o	50	100	150	200	250	300	350	400	450	500
T^*	34	34	34	34	34	34	34	34	34	34
$TC(T^*)$	52.52	53.30	54.09	54.88	55.67	56.45	57.24	58.03	58.82	59.60

Table 3: Effect of c_a on optimal policy

c_a	50	100	150	200	250	300	350	400	450	500
T^*	34	34	34	34	34	34	34	34	34	34
$TC(T^*)$	52.52	101.73	150.94	200.15	249.37	298.58	347.79	397.00	446.22	495.43

Table 4: Effect of c_s on optimal policy

c_s	100	200	300	400	500	600	700	800	900	1000
T^*	1	1	34	65	89	108	124	138	150	161
$TC(T^*)$	50.64	51.62	52.52	53.22	53.81	54.32	54.78	55.19	55.58	55.93

We note that only the increase of c_h causes the optimal cost to decrease. Otherwise, $TC(T^*)$ increases as either of the other 3 unit costs increases. Also, the effect of c_o on $TC(T^*)$ is much more significant than the other unit costs. Furthermore, c_o and c_a have no effect on the optimal threshold level T^*, while an increase in c_h causes T^* to decrease, and an increase in c_s causes T^* to increase.

Concerning the nonmonetary parameters, our computations showed that α, p, d_1, and b_1 had a negligible effect on either T^* or $TC(T^*)$, and for this reason they are not reproduced here.

However, the arrival process parameters have a significant effect on the optimal policy, as shown in Table 5 and 6 below.

Table 5: Effect of λ on optimal policy

λ	0.010	0.011	0.012	0.013	0.014	0.015	0.016	0.017	0.018	0.019	0.020
T^*	34	40	45	48	50	51	52	53	53	53	53
$TC(T^*)$	52.52	52.74	52.95	53.15	53.34	53.52	53.69	53.85	54.01	54.16	54.31

Table 6: Effect of ε on optimal policy

ε	0.10	0.17	0.24	0.31	0.38	0.45	0.52	0.59	0.66
T^*	47	38	29	19	9	1	1	1	1
$TC(T^*)$	52.69	52.58	52.40	52.10	51.61	50.83	49.56	47.42	43.54

This is rather interesting since the intent of this paper was to generalize the paper of Choudhury and Tadj (2013) to the case of batch arrivals. We note that as the arrival rate λ increases, the optimal value of T increases, although slightly. Also, as the parameter ε increases, i.e., as the mean size a_1 of an arriving group increases, the optimal value of T decreases while the optimal cost decreases.

To further show the effect of the arrival process on the optimal policy, we tried the following other distributions for the size of the batches: single arrivals $a(z) = z$, arrivals in groups of 2 units $a(z) = z^2$, and arrivals in groups of either one or two units $a(z) = \delta z + (1-\delta)z^2$. We used the same original data as above, and we also took $\delta = 0.2$. Figure 2 below shows the expected total cost per unit of time for all four types of distributions for $a(z)$.

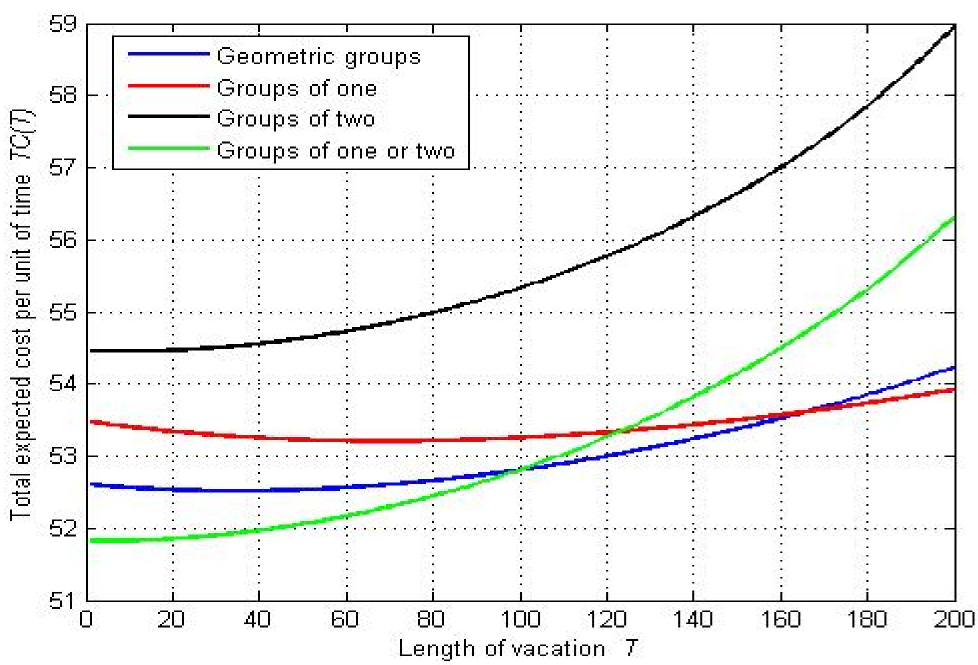

Figure 2: Variations of $TC(T)$ for various arrival batches.

Table 7 shows the optimal value of T and the optimal expected total cost per unit of time when different distributions are assumed for the size of arriving groups of units.

Table 7: Effect of batch size distribution on optimal policy

	$a(z) = \frac{(1-\varepsilon)z}{1-\varepsilon z}$	$a(z) = z$	$a(z) = z^2$	$a(z) = \delta z + (1-\delta)z^2$
T^*	34	69	10	4
$TC(T^*)$	52.52	53.21	54.45	51.82

The lowest optimal occurs when units arrive to the system in groups of either 1 or two. The worst scenario happens when units arrive always in groups of two. This type of analysis is useful when the decision maker has control over the input process.

6. Conclusion

We developed in this paper an optimal management policy for a compound Poisson arrivals queueing system with unreliable server. Repairs take place following a failure; however they are delayed, as could happen in a real situation. The server implements both a Bernoulli schedule and a T-policy. Numerical examples are run. It would be interesting to investigate this same system assuming the other applications of the T-policy cited in the Introduction section: randomized T-policy, NT-policy, and modified T-policies.

Acknowledgement

The authors would like to thank the referees and editor-in-chief for their useful comments and suggestions which considerably improved the presentation of the paper.

REFERENCES

Abolnikov, L. and Dukhovny, A. (1991). Markov chains with transition delta matrix: Ergodicity conditions, invariant probability measures and applications, Journal of Applied Mathematics Stochastic Analysis, Vol. 4, No. 4, 333–356.

Abolnikov, L., Dshalalow, J.H. and Dukhovny, A.M. (1994). First passage processes in queuing system $M^X/G^r/1$ with service delay discipline, International Journal of Mathematics and Mathematical Sciences, Vol. 17, No. 3, 571-586.

Alfa, A.S. and Frigui, I. (1996). Discrete NT-policy single server queue with Markovian arrival process and phase type service, European Journal of Operational Research, Vol. 88, 599-613.

Alfa, A.S. and Li, W. (2000). Optimal (N,T)-policy for $M/G/1$ system with cost structures, Performance Evaluation, Vol. 42, 265-277.

Chaudhry, M.L. and Templeton, J.G.C. (1983). A First Course in Bulk Queues, John Wiley & Sons, Inc., New York, New York.

Çinlar, E. (1975). Introduction to Stochastic Processes, Prentice-Hall, Englewood Cliffs, New Jersey.

Choudhury, G. and Tadj, L. (2013). Optimal policy for an unreliable service system. In Encyclopedia of Business Analytics and Optimization (John Wang, Editor), IGI Global under the imprint "Information Science Reference" (to appear).

Doganata, Y.N. (1990). NT-vacation policy for $M/G/1$ queue with starter. In Communication, Control, and Signal Processing (E. Arikan, Editor), 1663–1669.

Doshi, B.T. (1990). Single-server queues with vacations. In Stochastic Analysis of Computer and Communication Systems (H. Takagi, Editor), 217-265, North-Holland, Amsterdam.

Doshi, B.T. (1986). Queueing systems with vacations: A survey, Queueing Systems, Vol. 1, 29-66.

Feyaerts, B., De Vuyst, S., Wittevrongel, S. and Bruneel, H. (2012). Analysis of a discrete-time NT-policy queue with general service times, 7th International Conference on Queueing Theory and Network Applications (QTNA - 2012).

Fuhrmann, S.W. and Cooper, R.B. (1985). Stochastic decomposition in the $M/G/1$ queue with generalized vacations. Operations Research, Vol. 36, 566-569.

Gakis, K.G., Rhee, H.K., and Sivazlian, B.D. (1995). Distributions and first moments of the busy and idle periods in controllable $M/G/1$ queueing models with simple and dyadic policies, Stochastic Analysis and Applications, Vol. 13, 47-81.

Haghighi, A.M. and Mishev, D.P. (2013). Stochastic three-stage hiring model as a tandem queueing process with bulk arrivals and Erlang phase-type selection, International Journal of Mathematics in Operations Research, Vol. 5, No. 5, 571-603.

Heyman, D.P. (1977). The T-policy for the $M/G/1$ queue, Management Science, Vol. 23, No. 7, 775-778.

Hur, S., Kim, J. and Kang, C. (2003). An analysis of the $M/G/1$ system with N and T policy, Applied Mathematical Modelling, Vol. 27, 665-675.

Jiang, F.-C., Huang, D.-C., Yang, C.-T. and Wang, K.-H. (2010). Design framework to optimize power consumption and latency delay for sensor nodes using Min(N, T) policy $M/G/1$ queuing models, Future Information Technology (FutureTech), 5th International Conference on , pp.1,8, 21-23 May 2010.

Ke, J.-C. (2008). Two thresholds of a batch arrival queueing system under modified T vacation policy with startup and closedown, Mathematical Methods in the Applied Sciences, Vol. 31, No. 2, 229-247.

Ke, J.-C. (2006a). On $M/G/1$ system under NT policies with breakdowns, startup and closedown, Applied Mathematical Modelling, Vol. 30, 49-66.

Ke, J.-C. (2006b). Optimal NT policies for $M/G/1$ system with a startup and unreliable server, Computers & Industrial Engineering, Vol. 50, 248-262.

Ke, J.-C. (2005). Modified T vacation policy for an $M/G/1$ queueing system with an unreliable server and startup, Mathematical and Computer Modelling, Vol. 41, 1267-1277.

Ke, J.-C. and Chu, Y.-K. (2009). Notes of $M/G/1$ system under the $< p, T >$ policy with second optional service, Central European Journal of Operations Research, Vol. 17, 425-431.

Ke, J.-C. and Chu, Y.-K. (2008). Optimization on bicriterion policies for $M/G/1$ system with second optional service, Journal of Zhejiang University Science A, Vol. 9, No. 10, 1437-1445.

Ke, J.-C., Lin, C.-H. and Huang, H.-I. (2009). Entropy maximization and NT vacation $M/G/1$ model with a startup and unreliable server: Comparative analysis on the first two moments of system size, International Journal of Computer Mathematics, Vol. 86, 703-718.

Ke, J.-C., Wu, C.-H. and Zhang, Z.G. (2010). Recent developments in vacation queueing models: A short survey, International Journal of Operations Research, Vol. 7, No. 4, 3-8.

Kim, D.-J. and Moon, S.-A. (2006), Randomized control of T-policy for an $M/G/1$ system, Computers & Industrial Engineering, Vol. 51, pp. 684-92.

Li, W. and Alfa, A.S. (2000). Optimal policies for $M/M/m$ queue with two different kinds of (N,T)-policies, Naval Research Logistics, Vol. 47, Issue 3, 240-258.

Madan, K.C., Abu-Dayyeh, W. and Gharaibeh, M. (2003). Steady state analysis of two $M^X/M^{a,b}/1$ queue models with random breakdowns, Information and Management Science, Vol. 14, No. 3, 37-51.

Sen, K. and Gupta, R. (1994). Transient analysis of threshold T-policy $M/M/1$ queue with server control, Sankhyā: The Indian Journal of Statistics, Vol. 56, Series B, 39-51.

Tadj, L. (2003a). On a quorum queueing system under T-policy, Journal of the Operational Research Society, Vol. 54, 466-471.

Tadj, L. (2003b). On a bilevel bulk queueing system under T-policy, Journal of Statistical Research, Vol. 37, No. 2, 127-144.

Tadj, L., Choudhury, G. and Rekab, K. (2012). A two-phase quorum queueing system with Bernoulli vacation schedule, setup, and N-policy for an unreliable server with delaying repair, International Journal of Services and Operations Management, Vol. 12, No. 2, 139-164.

Wang, K.-H. and Ke, J.-C. (2002). Control policies of an $M/G/1$ queueing system with a removable and non-reliable server, International Transactions in Operational Research, Vol. 9, 195-212.

Wang, T.-Y., Wang, K.-H. and Pearn, W.L. (2009a). Optimization of the T policy $M/G/1$ queue with server breakdowns and general startup times, Journal of Computational and Applied Mathematics, Vol. 228, 270-278.

Wang, T.-Y., Wang, K.-H. and Pearn, W.L. (2009b). Maximum entropy analysis to the T policy $M/G/1$ queue with server breakdowns and startup times, International Journal of Information and Management Sciences, Vol. 20, 395-414.

Wang, K.-H., Yang; D.-Y. and Pearn, W.L. (2012). Analytical method for accuracy analysis of the randomized T-policy queue, International Journal of Innovative Computing, Information and Control, Vol. 8, No. 3(A), 1717-1730.

Wang, K.-H., Yang, D.-Y. and Pearn, W.L. (2010). Comparison of two randomized policy $M/G/1$ queues with second optional service, server breakdown and startup, Journal of Computational and Applied Mathematics, Vol. 234, 812-824.

Yang, D.-Y. Wang, K.-H., Ke, J.-C. and Pearn, W.L. (2008). Optimal randomized control policy of an unreliable server system with second optional service and startup, Engineering Computations: International Journal for Computer-Aided Engineering and Software, Vol. 25, No. 8, 783-800.

Zhang, Z.G., Tadj, L. and Bounkhel, M. (2011). Cost evaluation in $M/G/1$ queue with T-policy revisited: Technical Note, European Journal of Operational Research, Vol. 214, Issue 3, 814-817.