Journal of Computational and Applied Mathematics 234 (2010) 812–824

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Comparison of two randomized policy M/G/1 queues with second optional service, server breakdown and startup

Kuo-Hsiung Wang [a,∗], Dong-Yuh Yang [b], W.L. Pearn [b]

[a] Department of Applied Mathematics, National Chung-Hsing University, Taichung 402, Taiwan

[b] Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 30050, Taiwan

ARTICLE INFO

Article history:
Received 16 December 2008
Received in revised form 23 January 2010

Keywords:
Comparison
Optimization
(p, N)-policy
Second optional service
(T, p)-policy

ABSTRACT

The problem addressed in this paper is to compare the minimum cost of the two randomized control policies in the M/G/1 queueing system with an unreliable server, a second optional service, and general startup times. All arrived customers demand the first required service, and only some of the arrived customers demand a second optional service. The server needs a startup time before providing the first required service until the system becomes empty. After all customers are served in the queue, the server immediately takes a vacation and the system operates the (T, p)-policy or (p, N)-policy. For those two policies, the expected cost functions are established to determine the joint optimal threshold values of (T, p) and (p, N), respectively. In addition, we obtain the explicit closed form of the joint optimal solutions for those two policies. Based on the minimal cost, we show that the optimal (p, N)-policy indeed outperforms the optimal (T, p)-policy. Numerical examples are also presented for illustrative purposes.

1. Introduction

This paper deals with the comparison of the optimal (T, p)- and the optimal (p, N)-policy M/G/1 queues with an unreliable server, a second optional service (here abbreviated as *SOS*), and general startup times. Both policies provide the exhaustive service discipline, i.e., the server is turned off when there is no customer in the system. The (T, p)-policy is characterized by the following requirements: (i) switch the server off when the system becomes empty; (ii) if the server is turned off, the server takes a vacation of time T whenever the system becomes empty. If at least one customer is present in the system, then switch the server on with probability p $(0 \leq p \leq 1)$, and leave the server off with probability $(1 - p)$. After the server is turned off, the server will take another vacation of time T until the system becomes empty; and (iii) do not switch the server on/off at other epochs. In other words, the (T, p)-policy is to control the server randomly at the beginning epoch of the service when at least one customer appears. Based on the definition of the (T, p)-policy, the $(T, 1)$-policy coincides with the T-policy introduced in [1], and the $(T, 0)$-policy is identical to the $2T$-policy.

On the other hand, the (p, N)-policy for controlling the system is defined as follows: (i) turn the server off when the system is empty; (ii) turn the server on if N $(N \geq 1)$ or more customers are present; (iii) if the server is turned off and the number of customers in the system reaches N, turn the server on with probability p and leave the server off with probability $(1 - p)$; and (iv) do not turn the server on/off at other epochs. If the server finds at least N customers present in the system, the server starts to provide service for the waiting customers whenever it completes its startup. That is to say, the (p, N)-policy is to control the server randomly at the arrival epoch if the Nth customer finds that the server is idle. If the probability p is one, then we have the N-policy introduced in [2]. In case $p = 0$, we have the $(N + 1)$-policy.

∗ Corresponding author. Tel.: +886 4 22860133; fax: +886 4 22873028.
E-mail address: khwang@amath.nchu.edu.tw (K.-H. Wang).

0377-0427/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2010.01.045

There are many real-world situations in which the server is subject to breakdowns and repairs. In the literature, controllable queueing systems with an unreliable server have been studied extensively. Exact steady state solutions of the N-policy $M/E_k/1$ and $M/H_k/1$ queueing systems subject breakdowns were developed in [3,4], respectively. Wang and Ke [5] analyzed an M/G/1 queue with server breakdowns operating under the N-policy, T-policy and Min (N, T)-policy. Later, Pearn et al. [6] obtained the analytical results for sensitivity analysis in the N-policy M/G/1 queueing system with server breakdowns. The server startup corresponds to the preparatory work of the server before starting the service. In some actual situations, the server often needs a startup time before providing service. The N-policy M/G/1 queueing system with server startup was studied by several researchers such as Minh [7], Takagi [8], Lee and Park [9], Hur and Paik [10], and so on. Ke [11] studied the $M^{[x]}/G/1$ queueing systems under bi-level policy with an unreliable server and early startup. Wang et al. [12] considered the optimal control of the N-policy M/G/1 queueing system with server breakdowns and general startup times. Ke [13] examined a modified T-policy for the M/G/1 queue with an unreliable server and startup. Ke [14] extended Ke's model [13] to the $M^{[x]}/G/1$ queue with an unreliable server, startup and closedown. Recently, Wang et al. [15] focused mainly on performing a sensitivity analysis for the T-policy M/G/1 queue with server breakdowns and general startup times.

Extensive literature exist on the M/G/1 queue with *SOS*, where all arrived customers demand the main service and only some of them require subsidiary service provided by the server. The pioneering work on the M/G/1 queue with *SOS* was studied in [16]. Such queueing situations occur in many practical applications (see for details, [16]). Medhi [17] extended Madan's model [16] that the *SOS* time follows a general distribution. Medhi's model [17] was also analyzed in [18], in which they obtained the time-dependent probability generating functions and the corresponding steady state results. Using the supplementary variable technique, Wang [19] investigated the reliability behaviour in an M/G/1 queue with *SOS* and an unreliable server. Wang and Zhao [20] examined a discrete-time Geo/G/1 retrial queue with an unreliable server and *SOS*. In their work [20], they obtained explicit formulas for the stationary distribution and some performance measures in steady state. Furthermore, Choudhury and Dekaa [21] investigated the steady state behaviour of an M/G/1 retrial queue with *SOS* and server breakdowns. Dimitriou and Langaris [22] generalized their previous work of a retrial queue with two-phase service and server vacations (Dimitriou and Langaris [23]) by considering a server breakdowns and startup times. For such a system the stability conditions and steady state analysis are also derived by Dimitriou and Langaris [22]. More recently, Choudhury et al. [24] generalized both the classical $M^X/G/1$ retrial queue with service interruption as well as the $M^X/G/1$ queue with *SOS* and service interruption.

As for the randomized server control problem, Feinberg and Kim [25] first introduced the (p, N)- and the (N, p)-policy M/G/1 queues with a removable sever. Subsequently, Kim and Moon [26] considered the queueing system with the (p, T)-policy, exploited its properties and obtained the optimal values of T and p for a constrained problem. Ke et al. [27] performed the estimation of the expected busy period for the (p, N)-policy M/G/1 queueing system by using the bootstrap methods. Wang and Huang [28] utilized the maximum entropy approach to derive analytic maximum entropy results for the (p, N)-policy queue with a removable and unreliable server. Yang et al. [29] applied the same approach to investigate the (N, p)-policy M/G/1 queue with server breakdowns and general startup times. Ke and Chu [30] optimized the operating cost of the (p, T)-policy for an M/G/1 queueing system with *SOS*. Recently, Ke and Chu [31] compared the operating cost of the two bicriterion policies, (p, T) and (p, N), for an M/G/1 queue with a reliable server and *SOS*. Such comparative work between the randomized N- and T-policy M/G/1 queues with *SOS*, server breakdown and general startup times is rarely explored in the literature. In this paper, we perform such comparative work, which may be viewed as an extension of that done in [31].

The objectives of this paper are follows. First, we present the system performances for the T- and the N-policy M/G/1 queues with *SOS*, server breakdowns and general startup times. Second, we develop the system performances for the (T, p)- and the (p, N)-policy M/G/1 queues with *SOS*, server breakdowns, and general startup times. Third, we construct cost functions for the (T, p)- and the (p, N)-policy to obtain explicit forms for the joint optimal threshold values of the (T, p) and (p, N) at the minimum cost, respectively. Finally, an analytical comparison is made between the optimum costs for those two randomized control policies. We show that the optimal (p, N)-policy outperforms the optimal (T, p)-policy.

2. The model description

In this paper, we consider the (T, p)- and the (p, N)-policy M/G/1 queues with *SOS*, server breakdowns, and general startup times. It is assumed that arrivals of customers follow a Poisson process with parameter λ. Arrived customers form a single waiting line at a server based on the order of their arrivals; that is, in a first-come, first-served (FCFS) discipline. A single server needs to serve all arrived customers for the first required service (here abbreviated as *FRS*), denoted by S_1. As soon as the *FRS* of a customer is completed, a customer may leave the system with probability $1-\theta$ or may opt for *SOS*, denoted by S_2, with probability $\theta(0 \le \theta \le 1)$, at the completion of which the customer departs from the system and the next customer, if any, is taken up for the *FRS*. The service times S_1, S_2 of two channels are independent and identically distributed (i.i.d.) random variables obeying a general distribution function $S_i(t)(t \ge 0)$, $i = 1, 2$, mean service time μ_{S_i}, $i = 1, 2$, the Laplace–Stieltjes (abbreviated LS) transform $\bar{f}_{S_i}(s)\quad i = 1, 2$, and the k-th moment $E[S_i^k]$, $k \ge 1$, $i = 1, 2$, where the sub-index $i = 1$ denotes the *FRS* and $i = 2$ denotes the *SOS*. Furthermore, the same server is assumed to serve both service channels. Thus, a total service time provided to a customer is defined as:

$$S = \begin{cases} S_1 + S_2, & \text{with probability } \theta, \\ S_1, & \text{with probability } (1-\theta), \end{cases}$$

and its LS transform $\bar{f}_S(s) = (1-\theta)\bar{f}_{S_1}(s) + \theta\bar{f}_{S_1}(s)\bar{f}_{S_2}(s)$ with the first two moments of S are

$$E[S] = E[S_1] + \theta E[S_2] = \mu_{S_1} + \theta\mu_{S_2}, \tag{1}$$

$$E[S^2] = E[S_1^2] + 2\theta E[S_1]E[S_2] + \theta E[S_2^2]. \tag{2}$$

When the server is providing the *FRS* or *SOS*, the server may meet unpredictable breakdowns at any time but is immediately repaired. We assume that a server's breakdown time has an exponential distribution with rate α_1 in the *FRS* channel. In the *SOS* channel, the server fails at an exponential rate α_2. The repair times of the *FRS* and *SOS* channels are independent general distributions with distribution functions $R_1(t), R_2(t), (t \geq 0)$, the LS transforms $\bar{f}_{R_1}(s), \bar{f}_{R_2}(s)$, the mean repair times μ_{R_1}, μ_{R_2}, and the k-th moment $E[R_1^k]$, $E[R_2^k]$, $k \geq 1$, respectively. Although no service occurs during the repair period of the server, customers continue to arrive following a Poisson process. Once the failed server is repaired, the server immediately returns to serve a customer until the system becomes empty. After completion of the *FRS* or *SOS*, the server again is turned off when there are no customers in the system. Then, the server operates the (T, p)-policy or (p, N)-policy. The server requires a startup time with random length before starting the *FRS*. Again, the startup times are independent and identically distributed random variables obeying a general distribution function $U(t)(t \geq 0)$ and the k-th moment $E[U^k]$, $k \geq 1$. As soon as the server completes startup, it begins serving the waiting customers until the system is empty. Various stochastic processes involved in the system are assumed to be independent of each other.

Conveniently, we will present those two queueing models as the (T, p)-policy and the (p, N)-policy M/G(G, G)/1 queues, respectively. It is noted that the second symbol denotes service time distributions for both *FRS* and *SOS* channels, the third symbol denotes the repair time distributions for both *FRS* and *SOS* channels and the fourth symbol is the startup time distribution.

3. *T*-policy and *N*-policy queues

Let H_1 and H_2 be random variables representing the completion time of the *FRS* and *SOS*, respectively. The completion time of a customer includes both the service time of a customer and the repair time of a server. Now, we define that $\bar{f}_{H_1}(s) = E[e^{-sH_1}]$ and $\bar{f}_{H_2}(s) = E[e^{-sH_2}]$ as the LS transforms of H_1 and H_2, respectively. Thus, we have

$$\begin{aligned}\bar{f}_{H_i}(s) &= \int_0^\infty \sum_{n=0}^{\infty} \frac{e^{-\alpha_i t}(\alpha_i t)^n}{n!} e^{-st}\left[\bar{f}_{R_i}(s)\right]^n dS_i(t) \\ &= \bar{f}_{S_i}[s + \alpha_i - \alpha_i\bar{f}_{R_i}(s)], \quad i = 1, 2.\end{aligned} \tag{3}$$

From Eq. (3), we obtain the first two moments of H_1 and H_2 as follows:

$$E[H_i] = -\frac{d}{ds}\left[\bar{f}_{H_i}(s)\right]\Big|_{s=0} = \mu_{S_i}(1 + \alpha_i\mu_{R_i}), \quad i = 1, 2, \tag{4}$$

and

$$E[H_i^2] = \frac{d^2}{ds^2}\left[\bar{f}_{H_i}(s)\right]\Big|_{s=0} = (1 + \alpha_i\mu_{R_i})^2 E[S_i^2] + \alpha_i\mu_{S_i}E[R_i^2], \quad i = 1, 2. \tag{5}$$

Let H be the total completion time, and the LS transform of H is given by

$$\bar{f}_H(s) = (1-\theta)\bar{f}_{H_1}(s) + \theta\bar{f}_{H_1}(s)\bar{f}_{H_2}(s). \tag{6}$$

The first two moments of H are found to be

$$E[H] = -\frac{d}{ds}\left[\bar{f}_H(s)\right]\Big|_{s=0} = E[H_1] + \theta E[H_2] \tag{7}$$

and

$$E[H^2] = \frac{d^2}{ds^2}\left[\bar{f}_H(s)\right]\Big|_{s=0} = E[H_1^2] + 2\theta E[H_1]E[H_2] + \theta E[H_2^2], \tag{8}$$

where $E[H_i]$, $E[H_i^2]$, $i = 1, 2$, are given in Eqs. (4) and (5), respectively.

Applying the well-known formula for the probability generating function (p.g.f.) of the number of customers in the ordinary M/G/1 queue with an reliable server and *SOS*, the p.g.f. of the number of customers in the ordinary M/G/1 queue with an unreliable server and *SOS* is given by

$$G(s) = \frac{(1-\rho_H)(1-s)\bar{f}_H(\lambda - \lambda s)}{\bar{f}_H(\lambda - \lambda s) - s}, \tag{9}$$

where $\rho_H = \lambda E[H]$. It has to be noted that ρ_H is assumed to be less than unity.

3.1. T-policy queue

3.1.1. Expected number of customers in the system

According to the results of Yang et al. [32], we obtain the p.g.f. of the number of customers found in the T-policy M/G(G, G)/1 queue as follows:

$$G_T(s) = G(s)\left[\frac{1 - W_1(z)}{W_1'(1)(1-s)}\right], \tag{10}$$

where

$G_T(s)$ = the p.g.f. of number of customers in the T-policy M/G(G, G)/1 queueing system.

$W_1(s)$ = the p.g.f. of the number of customers that arrive during the during a period length T and the startup period;
$= e^{-(1-z)\lambda T}\bar{f}_U(\lambda - \lambda s)$, where $\bar{f}_U(\cdot)$ is the LS transform of the startup time.

Let L_T be the expected number of customers in the T-policy $M/G(G, G)/1$ queue. Thus, it follows that

$$L_T = G_T'(s)\big|_{s=1} = \frac{1}{(T+\mu_U)}\left[\frac{\lambda T^2}{2} + \rho_U T + \frac{\lambda E(U^2)}{2}\right] + L_H, \tag{11}$$

where $\rho_U = \lambda\mu_U$ and

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

3.1.2. Expected length of the idle and startup periods

The idle period begins when all the customers in the system are served and no customers are waiting for service. It terminates at least one customer arrives at the period length T. We can easily see that

$$E[I_T] = T. \tag{13}$$

On the other hand, the server begins startup when there is at least one waiting customer at the end of the fixed period T in the system. We call this startup period and denote it by U_T. It follows that

$$E[U_T] = \mu_U. \tag{14}$$

3.1.3. Expected length of the busy and breakdown periods

The completion period is from the end of the startup period to no customers in the system, which occurs before the system becomes empty and can be represented as the sum of the busy period and the breakdown period. A time interval when the server is working continuously is called busy period. During the busy period, the server may break down when *FES* or *SOS* is provided and start its repair immediately. This is called the breakdown period. After the server is repaired, it returns and provides service until there are no customers in the system. Let $E[H_T]$ be the expected length of the completion period the T-policy M/G(G, G)/1 queue. Again, we know from et al. [32] that

$$E[H_T] = \frac{(T+\mu_U)\rho_H}{1-\rho_H} = \frac{(\lambda T + \rho_U)[\mu_{S_1}(1+\alpha_1\mu_{R_1}) + \theta\mu_{S_2}(1+\alpha_2\mu_{R_2})]}{1-\rho_H}. \tag{15}$$

We also denote the expected length of the busy and breakdown periods by $E[B_T]$ and $E[D_T]$, respectively. Since the completion period is composed of the busy period and the breakdown period, which implies that $E[H_T] = E[B_T] + E[D_T]$. From Eq. (15), we obtain

$$E[B_T] = \frac{(\lambda T + \rho_U)(\mu_{S_1} + \theta\mu_{S_2})}{1-\rho_H}, \tag{16}$$

and

$$E[D_T] = \frac{(\lambda T + \rho_U)(\alpha_1\mu_{S_1}\mu_{R_1} + \theta\alpha_2\mu_{S_2}\mu_{R_2})}{1-\rho_H}. \tag{17}$$

3.1.4. Expected length of the busy cycle

The expected length of busy cycle for the T-policy M/G(G, G)/1 queue is denoted by $E[C_T]$. Since the busy cycle consists of the idle period (I_T), the startup period (U_T), the busy period (B_T) and the breakdown period (D_T). Hence, it can be shown that

$$E[C_T] = E[I_T] + E[U_T] + E[B_T] + E[D_T] = \frac{T+\mu_U}{1-\rho_H}. \tag{18}$$

3.2. *N*-policy queue

3.2.1. Expected number of customers in the system

Following the result of Wang et al. [12], we obtain

$$G_N(s) = G(s)\frac{[1 - W_2(s)]}{W_2'(1)(1 - s)}, \tag{19}$$

where

$G_N(s)$ = the p.g.f. of number of customers in the N-policy $M/G(G, G)/1$ queueing system.
$W_2(s)$ = the p.g.f. of the number of customers that arrive during the turned-off period plus the startup period;
$= s^N \bar{f}_U(\lambda - \lambda s)$, where $\bar{f}_U(\cdot)$ is the LS transform of the startup time.

Let L_N denote the expected number of customers in the N-policy $M/G(G, G)/1$ queue. Thus, we obtain

$$L_N = G_N'(s)\big|_{s=0} = \frac{1}{N + \rho_U}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E[U^2]}{2}\right] + L_H, \tag{20}$$

where L_H is given in Eq. (12).

3.2.2. Expected length of the idle and complete startup periods

Let us define the expected length of the idle and complete startup periods by I_N and V_N for the N-policy M/G(G, G)/1 queue, respectively. We know that the turned-off period I_N terminates when the Nth customer arrives in system. Since the length of times between two successive arrivals are independently, identically and exponentially distributed with mean $1/\lambda$, the expected length of the turned-off period, $E[I_N]$, for the N-policy M/G/1 queueing system with server breakdowns and general startup times is given by

$$E[I_N] = \frac{N}{\lambda}. \tag{21}$$

The complete startup period is the sum of the complete period and the startup period which implies $V_N = H_N + U_N$, where H_N and U_N are the complete period and the startup period, respectively. Again, using the results of Wang et al. [12], it yields that

$$E[V_N] = \frac{(N + \lambda\mu_U)[\mu_{S_1}(1 + \alpha_1\mu_{R_1}) + \theta\mu_{S_2}(1 + \alpha_2\mu_{R_2})]}{1 - \rho_H} + \mu_U. \tag{22}$$

Since $V_N = H_N + U_N$, it follows that $E[V_N] = E[H_N] + E[U_N]$. From Eq. (22), we have

$$E[H_N] = \frac{(N + \lambda\mu_U)[\mu_{S_1}(1 + \alpha_1\mu_{R_1}) + \theta\mu_{S_2}(1 + \alpha_2\mu_{R_2})]}{1 - \rho_H}, \tag{23}$$

$$E[U_N] = \mu_U. \tag{24}$$

3.2.3. Expected length of the busy and breakdown periods

Recall that the completion period is composed of the busy period and the breakdown period, which implies that $E[H_N] = E[B_N] + E[D_N]$. From Eq. (23), it gives

$$E[B_N] = \frac{(N + \rho_U)(\mu_{S_1} + \theta\mu_{S_2})}{1 - \rho_H}, \tag{25}$$

$$E[D_N] = \frac{(N + \rho_U)(\alpha_1\mu_{S_1}\mu_{R_1} + \theta\alpha_2\mu_{S_2}\mu_{R_2})}{1 - \rho_H}. \tag{26}$$

3.2.4. Expected length of the busy cycle

The busy cycle for the N-policy M/G(G, G)/1 queue is denoted by C_N, is the length of time from the beginning of the last idle period to the beginning of the next idle period. Because the busy cycle consists of the idle period (I_N), the startup period (U_N), the busy period (B_N) and the breakdown period (D_N), we get from Eqs. (21) and (24)–(26)

$$E[C_N] = E[I_N] + E[U_N] + E[B_N] + E[D_N] = \frac{N + \rho_U}{\lambda(1 - \rho_H)}. \tag{27}$$

4. (T, p)-policy and (p, N)-policy queues

4.1. (T, p)-policy queue

The primary objective of this subsection is to develop the various system performances for the (T, p)-policy M/G(G, G)/1 queue, including (i) the expected length of the idle, startup, busy, breakdown periods and the busy cycle; and (ii) the expected number of customers in the system.

4.1.1. Expected length of the idle, startup, busy, breakdown periods and the busy cycle

We denote by $(I_{2T}, I_{T,p})$, $(U_{2T}, U_{T,p})$, $(B_{2T}, B_{T,p})$, $(D_{2T}, D_{T,p})$ the idle, startup, busy, breakdown periods for the $2T$-policy and (T, p)-policy M/G(G, G)/1 queues, respectively. Let C_{2T} and $C_{T,p}$ be the busy cycle for the $2T$-policy and (T, p)-policy M/G(G, G)/1 queues, respectively. Based on the results of Feinberg and Kim [25], the system performances for the (T, p)-policy queue are the convex combinations of the system performances for the T-policy queue and the $2T$-policy queue. Using the above formulas (13)–(14) and (16)–(18), we have

$$E[I_{T,p}] = pE[I_T] + (1-p)E[I_{2T}] = T(2-p), \tag{28}$$

$$E[U_{T,p}] = pE[U_T] + (1-p)E[U_{2T}] = \mu_U, \tag{29}$$

$$E[B_{T,p}] = pE[B_T] + (1-p)E[B_{2T}] = \frac{[\lambda T(2-p) + \rho_U](\mu_{S_1} + \theta\mu_{S_2})}{1-\rho_H}, \tag{30}$$

$$E[D_{T,p}] = pE[D_T] + (1-p)E[D_{2T}] = \frac{[\lambda T(2-p) + \rho_U](\alpha_1\mu_{S_1}\mu_{R_1} + \theta\alpha_2\mu_{S_2}\mu_{R_2})}{1-\rho_H}, \tag{31}$$

$$E[C_{T,p}] = pE[C_T] + (1-p)E[C_{2T}] = \frac{T(2-p) + \mu_U}{1-\rho_H}. \tag{32}$$

4.1.2. Expected number of customers in the system

We denote Π_T, Π_{2T} and $\Pi_{T,p}$ by the cumulative amount of time that all customers spent in the system during a busy cycle for the T-, $2T$- and (T, p)-policy M/G(G, G)/1 queues, respectively. By using the renewal-reward theorem, we obtain

$$E[\Pi_T] = L_T E[C_T] = \frac{1}{1-\rho_H}\left[\frac{\lambda T^2}{2} + T\rho_U + \frac{\lambda E(U^2)}{2} + L_H(T + \mu_U)\right], \tag{33}$$

where L_H is given in Eq. (12). It follows that

$$\begin{aligned} E[\Pi_{T,p}] &= pE[\Pi_T] + (1-p)E[\Pi_{2T}] \\ &= \frac{1}{1-\rho_H}\left[\lambda T^2\left(2 - \frac{3}{2}p\right) + T\rho_U(2-p) + \frac{\lambda E(U^2)}{2} + L_H[T(2-p) + \mu_U]\right]. \end{aligned} \tag{34}$$

Let $L_{T,p}$ denote the expected number of customers in the (T, p)-policy M/G(G, G)/1 queue. Again, form the renewal-reward theorem, we have

$$L_{T,p} = \frac{E[\Pi_{T,p}]}{E[C_{T,p}]} = \frac{1}{T(2-p) + \mu_U}\left[\lambda T^2\left(2 - \frac{3}{2}p\right) + T\rho_U(2-p) + \frac{\lambda E(U^2)}{2}\right] + L_H. \tag{35}$$

Based on the result of Feinberg and Kim [25], $L_{T,p}$ is a convex combination of L_T for a T-policy and L_{2T} for a $2T$-policy. Thus, we get

$$L_{T,p} = \Theta L_T + (1-\Theta)L_{2T}, \tag{36}$$

where

$$\Theta = \frac{p(T + \mu_U)}{(2-p)T + \mu_U}.$$

One can demonstrate that Eq. (36) is identical to Eq. (35) easily. Moreover, Eq. (35) is in accordance with the expression (3) of Wang et al. [15] if we set $p = 1$ and $\theta = 0$.

4.2. (p, N)-policy queue

We develop various system performances for the (p, N)-policy M/G(G, G)/1 queue, including (i) the expected length of the idle, startup, busy, breakdown periods and the busy cycle; and (ii) the expected number of customers in the system.

4.2.1. Expected length of the idle, startup, busy, breakdown periods and the busy cycle

We denote by $(I_{N+1}, I_{p,N})$, $(U_{N+1}, U_{p,N})$, $(B_{N+1}, B_{p,N})$, $(D_{N+1}, D_{p,N})$ the idle, startup, busy, breakdown periods for the $(N+1)$-policy and (p, N)-policy $M/G(G, G)/1$ queue, respectively. Let C_{N+1} and $C_{p,N}$ be the busy cycle for the $(N+1)$-policy and (p, N)-policy M/G(G, G)/1 queues, respectively. Based on the results of Feinberg and Kim [25], the system performances for the (p, N)-policy queue are the convex combinations of the system performances for the N-policy queue and the $(N+1)$-policy queue. Applying the above formulas (21) and (24)–(27), we have

$$E\left[I_{p,N}\right] = pE\left[I_N\right] + (1-p)E\left[I_{N+1}\right] = \frac{N+1-p}{\lambda}, \tag{37}$$

$$E\left[U_{p,N}\right] = pE\left[U_N\right] + (1-p)E\left[U_{N+1}\right] = \frac{\rho_U}{\lambda}, \tag{38}$$

$$E\left[B_{p,N}\right] = pE\left[B_N\right] + (1-p)E\left[B_{N+1}\right] = \frac{E\left[S\right](N+1-p+\rho_U)}{1-\rho_H}, \tag{39}$$

$$E\left[D_{p,N}\right] = pE\left[D_N\right] + (1-p)E\left[D_{N+1}\right] = \frac{(N+1-p+\rho_U)(\alpha_1\mu_{S_1}\mu_{R_1} + \theta\alpha_2\mu_{S_2}\mu_{R_2})}{1-\rho_H}, \tag{40}$$

$$E\left[C_{p,N}\right] = pE\left[C_N\right] + (1-p)E\left[C_{N+1}\right] = \frac{N+1-p+\rho_U}{\lambda(1-\rho_H)}. \tag{41}$$

4.2.2. Expected number of customers in the system

Let Π_N^c, Π_{N+1}^c and $\Pi_{p,N}^c$ denote the cumulative amount of time that all customers spent in the system during a busy cycle for the N-, $(N+1)$- and (p, N)-policy M/G(G, G)/1 queues, respectively. Following the results of Feinberg and Kim [25], we can obtain

$$\begin{aligned} E\left[\Pi_N^c\right] &= L_N E\left[C_N\right] \\ &= \frac{1}{\lambda(1-\rho_H)}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E(U^2)}{2}\right] + \frac{L_H(N+\rho_U)}{\lambda(1-\rho_H)}, \end{aligned} \tag{42}$$

where L_H is given in Eq. (12).

It follows that

$$\begin{aligned} E\left[\Pi_{p,N}^c\right] &= pE\left[\Pi_N^c\right] + (1-p)E\left[\Pi_{N+1}^c\right] \\ &= \frac{1}{\lambda(1-\rho_H)}\left[\frac{N(N+1-2p)}{2} + (N+1-p)\rho_U + \frac{\lambda^2 E(U^2)}{2}\right] + \frac{L_H(N+1-p+\rho_U)}{\lambda(1-\rho_H)}. \end{aligned} \tag{43}$$

Let $L_{p,N}$ denote the expected number of customers in the (p, N)-policy M/G(G, G)/1 queue. Applying the renewal-reward theorem, it yields that

$$\begin{aligned} L_{p,N} &= \frac{E[\Pi_{p,N}^c]}{E[C_{p,N}]} \\ &= \frac{1}{N+1-p+\rho_U}\left[\frac{N(N+1-2p)}{2} + (N+1-p)\rho_U + \frac{\lambda^2 E[U^2]}{2}\right] + L_H, \end{aligned} \tag{44}$$

where L_H is given in Eq. (12).

Note that $L_{p,N}$ is a convex combination of L_N for an N-policy and L_{N+1} for an $(N+1)$-policy. Thus, we have

$$L_{p,N} = \Omega L_N + (1-\Omega)L_{N+1}, \tag{45}$$

where $\Omega = p(N+\rho_U)/(N+1-p+\rho_U)$.

It is easy to show that Eq. (44) is identical to Eq. (45). Additionally, Eq. (45) is coincided with the expression (3) of Wang et al. [12] if we set $p = 1$ and $\theta = 0$.

5. The optimal (T, p)- and (p, N)-policies

5.1. Determining the optimal (T, p)-policy

We develop the expected cost function per unit time for the (T, p)-policy M/G(G, G)/1 queue, in which T and p are decision variables. Our objective is to determine the optimum threshold values$T$ and p, say T^* and p^*, to minimize this cost function.

Since $E[B_{T,p}]/E[C_{T,p}]$ and $E[D_{T,p}]/E[C_{T,p}]$ are not functions of the decision variables T and p, the operating cost and the breakdown cost for the server are ignored in the cost function. Therefore, we will restrict ourselves to select the cost elements as follows:

$C_h \equiv$ holding cost per unit time for each customer present in the system;
$C_f \equiv$ cost per unit time for keeping the server off;
$C_s \equiv$ startup cost per unit time for the preparatory work of the server before starting the service;
$C_l \equiv$ setup cost per busy cycle.

Without loss of generality, we assume that $C_s > C_f$. Utilizing the definition of each cost element listed above and the corresponding system performances, the expected cost with threshold values (T, p) is given by

$$\begin{aligned} F_1(T,p) &= C_h L_{T,p} + C_f \frac{E[I_{T,p}]}{E[C_{T,p}]} + C_s \frac{E[U_{T,p}]}{E[C_{T,p}]} + C_l \frac{1}{E[C_{T,p}]} \\ &= C_h L_H + C_h \frac{1}{T(2-p)+\mu_U}\left[\lambda T^2\left(2-\frac{3}{2}p\right) + T\rho_U(2-p) + \frac{\lambda E(U^2)}{2}\right] + C_f \frac{T(1-\rho_H)(2-p)}{T(2-p)+\mu_U} \\ &\quad + C_s \frac{\mu_U(1-\rho_H)}{T(2-p)+\mu_U} + C_l \frac{1-\rho_H}{T(2-p)+\mu_U} \end{aligned} \tag{46}$$

where $E[I_{T,p}]$, $E[U_{T,p}]$, $E[C_{T,p}]$ and $L_{T,p}$ are given in Eqs. (28), (29), (32) and (35), respectively.

In Theorem 1 in the paper of Yang et al. [32], it claims that the joint optimal threshold values (T^*, p^*) exist, which minimize the expected cost function analytically. It follows that p^* is equal to 0 or 1, and T^* can be written as

$$T^* = \Psi\left(-\mu_U + \sqrt{\sigma_U^2 + \frac{2\left[(C_s - C_f)\mu_U + C_l\right](1-\rho_H)}{\lambda C_h}}\right), \tag{47}$$

where $\Psi = \begin{cases} 1/2, & \text{if } p = 0, \\ 1, & \text{if } p = 1. \end{cases}$

Substituting p^* and T^* into Eq. (46), we have the minimal value of $F_1(T,p)$, say $\text{Min}F_1$, which is given by

$$\text{Min}F_1 = F_1(T^*, p^*) = C_h L_H + C_f(1-\rho_H) + C_h\sqrt{\lambda^2\sigma_U^2 + \frac{2\lambda\left[(C_s - C_f)\mu_U + C_l\right](1-\rho_H)}{C_h}}. \tag{48}$$

5.2. Determining the optimal (p, N)-policy

We establish the expected cost function per unit time for the (p, N)-policy M/G(G, G)/1 queue, in which p and N are decision variables. Our objective is to determine the optimum threshold values p and N, say p^* and N^*, to minimize this cost function.

Since $E[B_{p,N}]/E[C_{p,N}]$, $E[D_{p,N}]/E[C_{p,N}]$ are not functions of the decision variables p and N, the operating cost and the breakdown cost for the server are ignored in the cost function. The expected cost function is given by

$$\begin{aligned} F_2(p,N) &= C_h L_{p,N} + C_f \frac{E[I_{p,N}]}{E[C_{p,N}]} + C_s \frac{E[U_{p,N}]}{E[C_{p,N}]} + C_l \frac{1}{E[C_{T,p}]} \\ &= C_h L_H + \frac{1}{N+1-p+\rho_U}\left[\frac{N(N+1-2p)}{2} + (N+1-p)\rho_U + \frac{\lambda^2 E[U^2]}{2}\right] C_h + \frac{(N+1-p)(1-\rho_H)}{N+1-p+\rho_U} C_f \\ &\quad + \frac{\rho_U(1-\rho_H)}{N+1-p+\rho_U} C_s + \frac{\lambda(1-\rho_H)}{N+1-p+\rho_U} C_l, \end{aligned} \tag{49}$$

where $E[I_{p,N}]$, $E[U_{p,N}]$, $E[C_{p,N}]$ and $L_{p,N}$ are given in Eqs. (37), (38), (41) and (44), respectively.

With the cost structure being constructed, our objective is to determine optimal threshold values (p^*, N^*) to minimize this cost function. To this end, we first differentiate the cost function (49) with respect to p in the following:

$$\begin{aligned} \frac{\partial F_2(p,N)}{\partial p} &= \frac{C_h}{2(N+1-p+\rho_U)^2}\left\{-N(N+1+2\rho_U) + \frac{2\left[\rho_U\left(C_s - C_f\right) + \lambda C_l\right](1-\rho_H) + \lambda^2\left(\sigma_U^2 - \mu_U^2\right)C_h}{C_h}\right\} \\ &= \frac{C_h}{2(N+1-p+\rho_U)^2}\left\{\varpi - N(N+1+2\rho_U)\right\}, \end{aligned} \tag{50}$$

where

$$\omega = \frac{2\left[\rho_U\left(C_s - C_f\right) + \lambda C_l\right](1-\rho_H) + \lambda^2\left(\sigma_U^2 - \mu_U^2\right)C_h}{C_h}. \tag{51}$$

Obviously, the following results can be conducted from Eq. (50).

(1) As $\varpi > N(N+1+2\rho_U)$, $F_2(p,N)$ is an increasing function in $p \in [0,1]$. It implies that

$$\min_{0\le p\le 1} F_1(p,N) = F_1(0,N)$$
$$= C_hL_H + \frac{\left\{\left(\frac{N(N+1)}{2} + (N+1)\rho_U + \frac{\lambda^2 E[U^2]}{2}\right)C_h + \left((N+1)C_f + \rho_U C_s + \lambda C_l\right)(1-\rho_H)\right\}}{N+1+\rho_U}. \quad (52)$$

(2) As $\omega = N(N+1+2\rho_U)$, $F_2(p,N)$ is independent of $p \in [0,1]$. If there exists an $N = N_0$ such that $\omega = N_0(N_0+1+2\rho_U)$. We obtain

(i) $F_2(p,N_0) = C_hL_H + (N_0+\rho_U)C_h + (1-\rho_H)C_f$.
(ii) $\mathrm{Min}F_2 = \min_{0\le p\le 1} F_2(p,N_0) = C_hL_H + (N_0+\rho_U)C_h + (1-\rho_H)C_f$.

(3) As $\varpi < N(N+1+2\rho_U)$, $F_2(p,N)$ is a decreasing function in $p \in [0,1]$. It indicates that

$$\min_{0\le p\le 1} F_2(p,N) = F_2(1,N)$$
$$= C_hL_H + \frac{\left\{\left(\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E[U^2]}{2}\right)C_h + \left(NC_f + \rho_U C_s + \lambda C_l\right)(1-\rho_H)\right\}}{N+\rho_U}. \quad (53)$$

(4) It is noted that $\varpi > N(N+1+2\rho_U)$, which is equivalent to $0 < N < \xi$, where

$$\xi = -0.5 - \rho_U + \sqrt{0.25 + \rho_U + \rho_U^2 + \omega}$$
$$= -0.5 - \rho_U + \sqrt{0.25 + \rho_U + \lambda^2\sigma_U^2 + \frac{2\left[\rho_U\left(C_s - C_f\right) + \lambda C_l\right](1-\rho_H)}{C_h}}. \quad (54)$$

We define that $\mathrm{H}(N) = F_1(0,N)$ when $0 < N < \xi$. Hence, we have

$$\mathrm{H}(N) = C_hL_H + \frac{1}{N+1+\rho_U}\left(\frac{N(N+1)}{2} + (N+1)\rho_U + \frac{\lambda^2 E[U^2]}{2}\right)C_h + \frac{\left((N+1)C_f + \rho_U C_s + \lambda C_l\right)(1-\rho_H)}{N+1+\rho_U}. \quad (55)$$

Using Eq. (55) yields

$$\mathrm{H}(N) - \mathrm{H}(N-1) = \frac{C_hN(N+1+2\rho_U)}{2(N+1+\rho_U)(N+\rho_U)}\left(1 - \frac{\omega}{N(N+1+2\rho_U)}\right) < 0. \quad (56)$$

One can see that $\mathrm{H}(N)$ is a decreasing function of N. Thus, we get

$$\min \mathrm{H} = \mathrm{H}([\xi])$$
$$= C_hL_H + \frac{C_h\left([\xi]([\xi]+1) + 2([\xi]+1)\rho_U + \lambda^2 E[U^2]\right)}{2([\xi]+1+\rho_U)} + \frac{\left(([\xi]+1)C_f + \rho_U C_s + \lambda C_l\right)(1-\rho_H)}{[\xi]+1+\rho_U}, \quad (57)$$

where [ ] denotes the Gaussian notation, i.e., $[\xi]$ is the greatest integer less than or equal to ξ.

(5) When $\varpi < N(N+1+2\rho_U)$, it is equivalent to $N > \xi$. Let us define that $\mathrm{M}(N) = F_1(1,N)$ when $N > \xi$, and $\mathrm{M}(N)$ can be expressed as

$$\mathrm{M}(N) = C_hL_H + \frac{1}{N+\rho_U}\left\{\left(\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E[U^2]}{2}\right)C_h + \left(NC_f + \rho_U C_s + \lambda C_l\right)(1-\rho_H)\right\}. \quad (58)$$

Using Eq. (58), we have

$$\mathrm{M}(N+1) - \mathrm{M}(N) = \frac{C_hN(N+1+2\rho_U)}{2(N+1+\rho_U)(N+\rho_U)}\left(1 - \frac{\omega}{N(N+1+2\rho_U)}\right) > 0. \quad (59)$$

Hence, $\mathrm{M}(N)$ is an increasing function of N with the following result.

$$\min \mathrm{M} = \mathrm{M}([\xi]+1)$$
$$= C_hL_H + \frac{C_h\left\{[\xi]([\xi]+1) + 2([\xi]+1)\rho_U + \lambda^2 E[U^2]\right\}}{2([\xi]+1+\rho_U)} + \frac{\left\{([\xi]+1)C_f + \rho_U C_s + \lambda C_l\right\}(1-\rho_H)}{[\xi]+1+\rho_U}. \quad (60)$$

Overall, we conclude that

(i) $\mathrm{Min}F_2 = F_2(p,N_0) = C_hL_H + (N_0+\rho_U)C_h + (1-\rho_H)C_f$ if $\omega = N_0(N_0+1+2\rho_U)$ for some integer N_0.
(ii) $\mathrm{Min}F_2 = F_2(0,[\xi]) = \min \mathrm{H} = F_2(1,[\xi]+1) = \min \mathrm{M}$ if $\omega \ne N(N+1+2\rho_U)$ for all integers N.

Generally, an integer N_0 does not exist to satisfy $\omega = N_0(N_0 + 1 + 2\rho_U)$. According to the above listed results, we obtain the following theorem.

Theorem 1. *Let (p^*, N^*) be the joint optimal threshold values that minimize the expected cost function in Eq. (49), i.e., $F_2(p^*, N^*)$ is the minimizer of $F_2(p, N)$. Let us define that*

$$\delta = \begin{cases} 0, & \text{if } p = 0, \\ 1, & \text{if } p = 1. \end{cases}$$

Then p^* is equal to zero or one and

$$N^* = \left[-0.5 - \rho_U + \sqrt{0.25 + \rho_U + \lambda^2\sigma_U^2 + \frac{2\left(\rho_U\left(C_s - C_f\right) + \lambda C_l\right)(1 - \rho_H)}{C_h}} \right] + \delta, \tag{61}$$

where [ ] is the Gaussian notation.

6. Comparison between the optimal (T, p)- and (p, N)-policies

The purpose of this section is to compare $F_1(T^*, p^*)$ for the (T, p)-policy and $F_2(p^*, N^*)$ for the (p, N)-policy. Based on Eq. (48), the optimal value of the expected cost function for the (T, p)-policy is given by

$$\begin{aligned} \text{Min}F_1 &= F_1(T^*, p^*) = C_h L_H + C_f(1 - \rho_H) + C_h\sqrt{\lambda^2\sigma_U^2 + \frac{2\lambda\left[(C_s - C_f)\mu_U + C_l\right](1 - \rho_H)}{C_h}} \\ &= C_h L_H + C_f(1 - \rho_H) + C_h\sqrt{\omega + \rho_U^2}, \end{aligned} \tag{62}$$

where ω is given in Eq. (51).

From Eqs. (57) and (60), the optimal value of the expected cost function for the (p, N)-policy yields

$$\begin{aligned} \text{Min}F_2 &= F_2(p^*, N^*) = C_h L_H + \frac{C_h\left\{[\xi]([\xi] + 1) + 2([\xi] + 1)\rho_U + \lambda^2 E[U^2]\right\}}{2([\xi] + 1 + \rho_U)} \\ &\quad + \frac{\left\{([\xi] + 1)C_f + \rho_U C_s + \lambda C_l\right\}(1 - \rho_H)}{[\xi] + 1 + \rho_U}. \end{aligned} \tag{63}$$

The result summarized in the following theorem show that the optimal (p, N)-policy is superior to the optimal (T, p)-policy.

Theorem 2. *Fixing the same values of cost elements and system parameters in the (T, p)- and the (p, N)-policy* M/G(G, G)/1 *queues. Then the minimum expected cost per unit time of the (p, N)-policy is less than that of (T, p)-policy, that is,* $\text{Min}F_2 < \text{Min}F_1$.

Proof. Since Eq. (54) is equivalent to $\xi(\xi + 1 + 2\rho_U) = \omega$, we have

$$2(\xi + \rho_U)([\xi] + 1 + \rho_U) - [\xi]([\xi] + 1) - \omega - 2([\xi] + 1 + \rho_U)\rho_U = ([\xi] + 1 - \xi)(\xi - [\xi]) > 0. \tag{64}$$

We know from Eq. (64) that

$$2(\xi + \rho_U)([\xi] + 1 + \rho_U) > [\xi]([\xi] + 1) + \omega + 2([\xi] + 1 + \rho_U)\rho_U. \tag{65}$$

From Eq. (54), again, one can easily verify that

$$\sqrt{\omega + \rho_U^2} = \sqrt{\xi^2 + \xi + 2\xi\rho_U + \rho_U^2} > \xi + \rho_U. \tag{66}$$

Combining Eqs. (65) and (66), it implies that

$$[\xi]([\xi] + 1) + \omega + 2([\xi] + 1 + \rho_U)\rho_U < 2\sqrt{\omega + \rho_U^2}([\xi] + 1 + \rho_U). \tag{67}$$

Eq. (67) leads the following result:

$$\frac{[\xi]([\xi] + 1) + \omega + 2([\xi] + 1 + \rho_U)\rho_U}{\sqrt{\omega + \rho_U^2}([\xi] + 1 + \rho_U)} < 2. \tag{68}$$

Table 1
Comparison of the minimum costs and the optimal threshold values for two different policies. ($C_h = 5$, $C_s = 300$, $C_l = 400$, $C_f = 100$, $E[H] = 0.3$, $E[H^2] = 0.5$).

(λ, γ)		(1.0, 3.0)	(1.5, 3.0)	(2.0, 3.0)	(2.0, 4.0)	(2.0, 5.0)	(2.0, 6.0)
$p^* = 0$	T^*	5.55	3.97	2.89	2.88	2.87	2.86
$p^* = 1$	T^*	11.10	7.94	5.78	5.75	5.73	5.72
$F_1(T^*, p^*)$		130.449	124.429	116.631	115.517	114.841	114.386
$p^* = 0$	N^*	10	11	11	11	10	10
$p^* = 1$	N^*	11	12	12	12	11	11
$F_2(p^*, N^*)$		128.024	122.030	114.301	113.167	112.477	112.003

Table 2
Comparison of the minimum costs and the optimal threshold values for two different policies. ($C_l = 400$, $C_f = 100$, $\lambda = 1.0$, $\gamma = 3.0$, $E[H] = 0.3$, $E[H^2] = 0.5$).

(C_h, C_s)		(100, 200)	(50, 200)	(10, 200)	(10, 400)	(10, 600)	(10, 800)
$p^* = 0$	T^*	1.07	1.58	3.73	4.02	4.29	4.54
$p^* = 1$	T^*	2.14	3.16	7.46	8.04	8.58	9.08
$F_1(T^*, p^*)$		382.771	277.287	154.484	160.260	165.661	170.754
$p^* = 0$	N^*	1	2	6	7	8	8
$p^* = 1$	N^*	2	3	7	8	9	9
$F_2(p^*, N^*)$		340.317	254.968	149.854	155.460	160.937	165.937

Table 3
Comparison of the minimum costs and the optimal threshold values for two different policies. ($C_h = 5$, $C_s = 300$, $\lambda = 1.0$, $\gamma = 3.0$, $E[H] = 0.3$, $E[H^2] = 0.5$).

(C_l, C_f)		(100, 100)	(300, 100)	(500, 100)	(400, 150)	(400, 200)	(400, 250)
$p^* = 0$	T^*	3.25	4.90	6.13	5.45	5.34	5.23
$p^* = 1$	T^*	6.50	9.80	12.26	10.89	10.68	10.47
$F_1(T^*, p^*)$		107.456	123.957	136.275	164.419	198.370	232.301
$p^* = 0$	N^*	6	9	11	10	10	9
$p^* = 1$	N^*	7	10	12	11	11	10
$F_2(p^*, N^*)$		105.154	121.547	133.856	161.995	195.966	229.935

After some simplification efforts, it finally yields that

$$\begin{aligned}\mathrm{Min}F_1 - \mathrm{Min}F_2 &= C_f(1-\rho_H) + C_h\sqrt{\omega+\rho_U^2} - \frac{C_h\left\{[\xi]\left([\xi]+1\right)+2([\xi]+1)\rho_U+\lambda^2 E[U^2]\right\}}{2\left([\xi]+1+\rho_U\right)} \\ &\quad - \frac{\left\{([\xi]+1)\,C_f+\rho_U C_s+\lambda C_l\right\}(1-\rho_H)}{[\xi]+1+\rho_U} \\ &= C_h\sqrt{\omega+\rho_U^2}\left\{1-\frac{1}{2}\left(\frac{([\xi])\,([\xi]+1)+2\rho_U\,([\xi]+1+\rho_U)+\omega}{\sqrt{\omega+\rho_U^2}\,([\xi]+1+\rho_U)}\right)\right\} > 0. \end{aligned} \tag{69}$$

Therefore, it follows that $\mathrm{Min}F_2 < \mathrm{Min}F_1$, that is, the minimum expected cost per unit time of the (p, N)-policy is less than that of (T, p)-policy. □

Numerical examples

In order to illustrate Theorem 2, we perform a numerical comparison between the optimal (T, p)-policy and the optimal (p, N)-policy. For convenience of computations, we assume that the distribution of the startup time to be E_3 (3-stage Erlang distribution), i.e., $E[U] = 1/\gamma$ and $E[U^2] = 4/(3\gamma^2)$. First, we fix cost elements $C_h = 5$, $C_s = 300$, $C_l = 400$, $C_f = 100$, select $E[H] = 0.3$, $E[H^2] = 0.5$, and change in specific values (λ, γ). The numerical results of the joint optimal threshold values and the minimum cost are summarized in the Tables 1–4. From Table 1, it indicates that (i) T^* decreases as λ increases; (ii) N^* increases as λ increases; (iii) T^* and N^* are insensitive to γ; (iv) $F_1(T^*, p^*)$ and $F_2(p^*, N^*)$ decrease as λ or γ increases; and $(v)F_2(p^*, N^*)$ is less than $F_1(T^*, p^*)$, that is, $\mathrm{Min}F_2 < \mathrm{Min}F_1$.

Next, we fix cost elements $C_l = 400$, $C_f = 100$, select $\lambda = 1.0$, $\gamma = 3.0$, $E[H] = 0.3$, $E[H^2] = 0.5$, and change in specific values of (C_h, C_s). Table 2 shows that the numerical results of the joint optimal threshold values and the minimum cost. From Table 2, we observe that (i) T^* and N^* increase as C_h decreases; (ii) T^* and N^* increase as C_s increases; (iii) $F_1(T^*, p^*)$ and $F_2(p^*, N^*)$ increase as C_h or C_s increases; and (iv) $F_2(p^*, N^*)$ is less than $F_1(T^*, p^*)$.

Table 4
Comparison of the minimum costs and the optimal threshold values for two different policies. ($C_h = 5, C_s = 300, C_l = 400, C_f = 100, \lambda = 1.0, \gamma = 3.0$).

$(E[H], E[H^2])$		(0.2, 0.4)	(0.5, 0.4)	(0.8, 0.4)	(0.5, 0.2)	(0.5, 0.5)	(0.5, 1.0)
$p^* = 0$	T^*	5.94	4.66	2.89	4.66	4.67	4.66
$p^* = 1$	T^*	11.89	9.33	5.78	9.33	9.33	9.33
$F_1(T^*, p^*)$		143.359	102.814	59.566	101.814	103.314	105.814
$p^* = 0$	N^*	11	8	5	8	8	8
$p^* = 1$	N^*	12	9	6	9	9	9
$F_2(p^*, N^*)$		140.929	100.433	57.216	99.433	100.933	103.433

Furthermore, we fix cost elements $C_h = 5$, $C_s = 300$, select $\lambda = 1.0$, $\gamma = 3.0$, $E[H] = 0.3$, $E[H^2] = 0.5$, and change in specific values of (C_l, C_f). The numerical results of the joint optimal threshold values and the minimum cost are summarized in the Table 3. One can see from Table 3 that (i) T^* and N^* increase as C_l increases; (ii) T^* and N^* decrease as C_f increases; (iii) $F_1(T^*, p^*)$ and $F_2(p^*, N^*)$ increase as C_l or C_f increases; and (iv) $F_2(p^*, N^*)$ is less than $F_1(T^*, p^*)$.

Finally, we fix cost elements $C_h = 5$, $C_s = 300$, $C_l = 400$, $C_f = 100$, select $\lambda = 1.0$, $\gamma = 3.0$, and change in specific values of $(E[H], E[H^2])$. Table 4 shows that the numerical results of the joint optimal threshold values and the minimum cost. From Table 4, it appears that (i) T^* and N^* decrease as $E[H]$ increases; (ii) T^* and N^* are unchanged as $E[H^2]$ changes; and (iii) $F_1(T^*, p^*)$ and $F_2(p^*, N^*)$ decrease as $E[H]$ increases or $E[H^2]$ decreases; and (iv) $F_2(p^*, N^*)$ is less than $F_1(T^*, p^*)$, that is $\text{Min}F_2 < \text{Min}F_1$.

7. Conclusions

In this paper, we investigated the optimal (T, p)-policy and the optimal (p, N)-policy M/G/1 queues with *SOS*, server breakdown and general startup times. We first developed various system performances for those two policies. We then established cost functions to determine the joint optimal threshold values (T, p) and (p, N). The explicit closed forms of the joint optimal solutions for those two policies were obtained. In particular, we showed that under the optimal operating conditions, the (p, N)-policy indeed has less cost than the (T, p)-policy. Numerical comparisons are provided to illustrate that the optimal (p, N)-policy outperforms the optimal (T, p)-policy.

References

[1] D.P. Heyman, The *T* policy for the M/G/1 queue, Management Science 23 (1977) 775–778.
[2] M. Yadin, P. Naor, Queueing systems with a removable service station, Operational Research Quarterly 14 (1963) 393–405.
[3] K.-H. Wang, Optimal control of an $M/E_k/1$ queueing system with removable service station subject to breakdowns, Journal of the Operational Research Society 48 (1997) 936–942.
[4] 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, Quality Technology & Quantitative Management 1 (2004) 325–339.
[5] K.-H. Wang, J.-C. Ke, Control policies of an M/G/1 queueing system with a removable and non-reliable server, International Transactions in Operational Research 9 (2002) 195–212.
[6] W.L. Pearn, J.-C. Ke, Y.C. Chang, Sensitivity analysis of optimal management policy for a queueing system with a removable and non-reliable server, Computers & Industrial Engineering 46 (2004) 87–99.
[7] D.L. Minh, Transient solutions for some exhaustive M/G/1 queues with generalized independent vacations, European Journal of Operational Research 36 (1988) 197–201.
[8] H. Takagi, $M/G/1/K$ queues with *N*-policy and setup times, Queueing Systems 14 (1993) 79–98.
[9] H.W. Lee, J.O. Park, Optimal strategy in *N*-policy production system with early set-up, Journal of Operational Research Society 48 (1997) 306–313.
[10] S. Hur, S.J. Paik, The effect of different arrival rates on the *N*-policy of M/G/1 with server setup, Applied Mathematical Modelling 23 (1999) 289–299.
[11] J.-C. Ke, Bi-level control for batch arrival queues with an early startup and un-reliable server, Applied Mathematical Modelling 28 (2004) 469–485.
[12] K.-H. Wang, T.-Y. Wang, W.L. Pearn, Optimal control of the *N* policy M/G/1 queueing system with server breakdowns and general startup times, Applied Mathematical Modelling 31 (2007) 2199–2212.
[13] J.-C. Ke, Modified *T* vacation policy for an M/G/1 queueing system with an unreliable server and startup, Mathematical and Computer Modeling 41 (2005) 1267–1277.
[14] J.-C. Ke, An $M^{[x]}/G/1$ system with startup server and *J* additional options for service, Applied Mathematical Modelling 32 (2008) 443–458.
[15] K.-H. Wang, T.-Y. Wang, W.L. Pearn, Optimization of the *T* policy M/G/1 queue with server breakdowns and general startup times, Journal of Computational and Applied Mathematics 228 (2009) 270–278.
[16] K.C. Madan, An M/G/1 queue with second optional service, Queueing Systems 34 (2000) 37–46.
[17] J. Medhi, A single server Poisson input queue with a second optional channel, Queueing Systems 42 (2002) 239–242.
[18] J. Al-Jararha, K.C. Madan, An M/G/1 queue with second optional service with general service time distribution, Information and Management Sciences 14 (2003) 47–56.
[19] J. Wang, An M/G/1 queue with second optional service and server breakdowns, Computers and Mathematics with Applications 47 (2004) 1713–1723.
[20] J. Wang, Q. Zhao, A discrete-time Geo/G/1 retrial queue with starting failures and second optional service, Computers and Mathematics with Applications 53 (2007) 115–127.
[21] G. Choudhury, K. Dekaa, An M/G/1 retrial queueing system with two phases of service subject to the server breakdown and repair, Performance Evaluation 65 (2008) 714–724.
[22] I. Dimitriou, C. Langaris, A repairable queueing model with two-phase service, start-uptimes and retrial customers, Computers & Operations Research 37 (2010) 1181–1190.
[23] I. Dimitriou, C. Langaris, Analysis of a retrial queue with two-phase service and server vacations, Queueing Systems 60 (2008) 111–129.
[24] G. Choudhurya, L. Tadj, K. Dekaa, A batch arrival retrial queueing system with two phases of service and service interruption, Computers and Mathematics with Applications 59 (2010) 437–450.

[25] E.A. Feinberg, D.J. Kim, Bicriterion optimization of an M/G/1 queue with a removable server, Probability in the Engineering and Informational Sciences 10 (1996) 57–73.
[26] D.-J. Kim, S.-A. Moon, Randomized control of T-policy for an M/G/1 system, Computers & Industrial Engineering 51 (2006) 684–692.
[27] J.-C. Ke, M.Y. Ko, S.-H. Sheu, Estimation comparison on busy period for a controllable M/G/1 system with bicriterion policy, Simulation Modelling Practice and Theory 16 (2008) 645–655.
[28] K.-H. Wang, K.-B. Huang, A maximum entropy approach for the $\langle p, N \rangle$-policy M/G/1 queue with a removable and unreliable server, Applied Mathematical Modelling 33 (2009) 2024–2034.
[29] D.Y. Yang, K.-H. Wang, W.L. Pearn, Steady-State probability of the randomized server control system with second optional service, server breakdowns and startup, Journal of Applied Mathematics and Computing 32 (1) (2009) 39–58.
[30] J.-C. Ke, Y.-K. Chu, Notes of M/G/1 system under the $\langle p, T \rangle$-policy with second optional service, Central European Journal of Operations Research 17 (2009) 425–431.
[31] J.-C. Ke, Y.-K. Chu, Optimization on bicriterion policies for M/G/1 system with second optional service, Journal of Zhejiang University - Science A 9 (7) (2008) 1437–1445.
[32] D.-Y. Yang, K.-H. Wang, J.-C. Ke, W.L. Pearn, 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 25 (2008) 783–800.