Expert Systems with Applications 38 (2011) 9461–9471

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Comparative analysis of a randomized N-policy queue: An improved maximum entropy method

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

[a] Department of Business Administration, Asia University, Wufeng, Taichung 41354, Taiwan
[b] Institute of Information Science and Management, National Taipei College of Business,Taipei 100, Taiwan
[c] Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 30050, Taiwan

ARTICLE INFO

Keywords:
Comparative analysis
Improved maximum entropy
$\langle p,N\rangle$-policy
Sever breakdowns
Second optional service
Startup

ABSTRACT

We analyze a single removable and unreliable server in an M/G/1 queueing system operating under the $\langle p,N\rangle$-policy. As soon as the system size is greater than N, turn the server on with probability p and leave the server off with probability $(1 - p)$. All arriving customers demand the first essential service, where only some of them demand the second optional service. He needs a startup time before providing first essential service until there are no customers in the system. The server is subject to break down according to a Poisson process and his repair time obeys a general distribution. In this queueing system, the steady-state probabilities cannot be derived explicitly. Thus, we employ an improved maximum entropy method with several well-known constraints to estimate the probability distributions of system size and the expected waiting time in the system. By a comparative analysis between the exact and approximate results, we may demonstrate that the improved maximum entropy method is accurate enough for practical purpose, and it is a useful method for solving complex queueing systems.

1. Introduction

In this paper, we consider an unreliable server in an M/G/1 queue operating under the $\langle p,N\rangle$-policy with a second optional service (here abbreviated as *SOS*) and general startup times. An unreliable server means that the server is typically subject to unpredictable breakdowns. We elaborate an information theoretic technique based on the principle of maximum entropy to give an alternative solution for deriving probability distributions in this queueing model. We call that the policy is a $\langle p,N\rangle$-policy if it prescribes the following conditions: (i) turn the server off when the system is empty, (ii) turn the server on if there are $N(N \geqslant 1)$ 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 at other epochs. If the server finds at least N customers present in the system, it starts to provide first essential service (here abbreviated as *FES*) for the waiting customers whenever he completes its startup. In other words, the $\langle p,N\rangle$-policy is to control the server randomly at the arrival epoch of the Nth customer finds that the server is idle. If the probability p is one, then we have N-policy introduced by Yadin and Naor (1963). In case $p = 0$, we have the $(N + 1)$-policy. An M/G/1 queue involving the randomized server control problem has been treated by Feinberg and Kim (1996). They considered either $\langle p,N\rangle$- or $\langle N,p\rangle$-policy M/G/1 queue with a removable sever at first and performed the optimal control policy is of the randomized form. Subsequently, Kim and Moon (2006) considered the system with the $\langle p, T\rangle$-policy, exploit its properties and found the optimal values of T and p for a constrained problem. Lately, Ke, Ko, and Sheu (2008) utilized bootstrap methods to investigate the estimation of the expected busy period of an M/G/1 queueing system under $\langle p,N\rangle$-policy.

One of the most significant regions of queueing problem is the control of queue, and have studied extensive by many researchers. Yadin and Naor (1963) first introduced the concept of an N-policy which turns the server on whenever N $(N \geqslant 1)$ or more customers are present, turns the server off only when the system is empty. 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. Exact steady-state solutions of the N policy M/M/1 queue with exponential startup times were first derived by Baker (1973). Borthakur, Medhi, and Gohain (1987) extended Baker's model to general startup times. Wang (2003) developed the exact steady-state solutions of the N policy M/M/1 queue with server breakdowns and exponential startup times. The N-policy M/G/1 queue with startup times was

* Corresponding author.
E-mail address: khwang@amath.nchu.edu.tw (K.-H. Wang).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.01.153

investigated by several authors such as Medhi and Templeton (1992), Takagi (1993), Lee and Park (1997), etc. Ke (2003) analyzed the N policy M/G/1 queueing system with server vacations, startup and breakdowns. He developed the probability generating function of the queue size when the server begins performing startup and also derived important system characteristics. Recently, Wang, Wang, and Pearn (2007) focused mainly on performing a sensitivity analysis for the N-policy with server breakdowns and general startup times.

In many real service systems, one encounters numerous examples of the queueing situation where all arrivals require the main service and only some may require the subsidiary service provided by the server. Madan (2000) was the first to study an M/G/1 queue with *SOS* in which the first essential service time obeys a general distribution but second optional service time follows an exponential distribution. He also cited some important examples in daily life. Medhi (2002) extended Madan's model (Madan, 2000) that the second optional service time follows a general distribution. Al-Jararha and Madan (2003) generalized Madan's work in the sense that they assumed that both first essential service time and second optional service time are general with different distribution functions. Based on the supplementary variable technique, Wang (2004) studied the reliability behavior in an M/G/1 queue with *SOS* and an unreliable server. Recently, Wang and Zhao (2007) considered a discrete-time Geo/G/1 retrial queue with an unreliable server and *SOS*. Some performance measures of the system in steady state and explicit formulae for the stationary distribution are developed in their work.

In a stochastic context, little is known analytically about the behaviors of queue length distributions of a randomized server control queueing system. When exact methods of solution are not known, we frequently make use of numerical solution methods. One elegant approach for this is given by an information theoretic technique, which based on the principle of maximum entropy, to provide a self-contained method of inference for obtaining an unknown and unique probability distribution. In other word, this method is applied to estimate probability distributions, which consists of maximizing entropy function subject to the available mean constraints. El-Affendi and Kouvatsos (1983) presented the maximum entropy formalism to analyze the M/G/1 and G/M/1 queues. Based on the maximum entropy principle, Artalejo and Lopez-Herrero (2004) investigated the probability density function of busy period under some controllable M/G/1 queueing models. Wang, Wang, and Pearn (2005) used maximum entropy analysis to study the N policy M/G/1 queueing system with server breakdowns and general startup times. Recently, Ke and Lin (2006) studied the $M^{[x]}$/G/1 queueing system with an unreliable server and delaying vacations. They derived the approximate steady-state probability distribution of the queue length as well. To the best of our knowledge, that there has been no research that investigates a randomized controllable queueing system with *SOS* and startup times by the maximum entropy principle. Our work is motivated by such works and employ maximum entropy method to estimate the queue length distribution for the $\langle p,N\rangle$-policy M/G/1 queue with server breakdowns, *SOS* and startup times.

The purpose of this paper is fourfold. Firstly, we develop some exact and important system performance measures for the $\langle p,N\rangle$-policy M/G/1 queue with server breakdowns, *SOS* and startup times. Secondly, we construct an improved maximum entropy function for this queueing system. Thirdly, the improved maximum entropy solutions are developed through the Lagrange's method. Thirdly, we obtain the approximate expected waiting time in the system and the exact expected waiting time in the system. Finally, we perform a comparative analysis between approximate results obtained through the improved maximum entropy method and exact results obtained from the convex combination property.

2. The mathematical model

In this paper, we consider the $\langle p,N\rangle$ M/G/1 queue with the following specifications. It is assumed that customers arrive according to a Poisson process with rate λ. Arriving 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 is required to serve all arriving customers for the first essential service (*FES*), denoted by S_1. As soon as *FES* 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 θ $(0 \leqslant \theta \leqslant 1)$, at the completion of which the customer departs from the system and the next customer, if any, from the queue is taken up for his *FES*. 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 $S_i(t)$ $(t \geqslant 0)$, $i = 1, 2$, mean service time $\mu_{S_i}, i=1,2$, Laplace-Stieltjes transforms (LST) $\bar{f}_{S_i}(s) i=1,2$, and the kth moment $E\left[S_i^k\right], k \geqslant 1, i=1,2$, where the sub-index $i = 1$ (respectively, $i = 2$) denote the *FES* (respectively the *SOS*). Further, the same server is assumed to serve both service channels. Therefore, a total service time is 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 LST $\bar{f}_S(s) = (1-\theta)\bar{f}_{S_1}(s) + \theta \bar{f}_{S_1}(s)\bar{f}_{S_2}(s)$ with the first moments of S are

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

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

When the server is working, it may meet unpredictable breakdowns but is immediately repaired. We assume that a server's breakdown time has an exponential distribution with rate α_1 in the *FES* channel. In the *SOS* channel, the server fails at an exponential rate α_2. When the server fails, it is immediately repaired at a repair facility. The repair times of *FES* and *SOS* channels are independent general distributions with distribution functions $R_1(t)$, $R_2(t)$, $(t \geqslant 0)$, mean repair times μ_{R_1}, μ_{R_2} and the kth moment $E\left[R_1^k\right], E\left[R_2^k\right], k \geqslant 1$, 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, it immediately returns to serve a customer until the system is empty.

The idle server operates the $\langle p,N\rangle$-policy when there are N customers accumulated in the system. He requires a startup time with random length before starting *FES*. Again, the startup times are independent and identically distributed random variables obeying a general distribution function $U(t)$ $(t \geqslant 0)$, mean startup time μ_U and the kth moment $E[U^k]$, $k \geqslant 1$. As soon as the server completes startup, it begins serving the waiting customers until the system is empty. Let us suggest to the usual independence assumptions between inter-arrival times, service times, inter-breakdown times, startup times and repair times. Conveniently, We will present this queueing model as the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue, where the second and third symbols denote service time distributions for *FES* and *SOS* channels, respectively. The fourth and fifth symbols denote the repair time distributions for *FES* and *SOS* channels, respectively. The sixth symbol is the startup time distribution.

3. System performance measures

Let H_1 and H_2 be a random variable representing the completion time of *FES* and *SOS*, respectively. The completion time of a customer includes both the service time of a customer and the repair time of a server. Using the known results of Wang and Ke (2002), we get the first two moments of the completion time distribution for the first essential channel and second optional channel:

$$E[H_i] = \mu_{S_i}\left(1 + \alpha_i \mu_{R_i}\right), \quad i = 1, 2, \tag{3}$$

$$E\left[H_i^2\right] = \left(1 + \alpha_i \mu_{R_i}\right)^2 E\left[S_i^2\right] + \alpha_i \mu_{S_i} E\left[R_i^2\right], \quad i = 1, 2. \tag{4}$$

We denote by I_N, U_N, B_N and D_N, idle, startup, busy, breakdown periods for the *N*-policy M/(G,G), (G,G), G/1 queue, respectively. Suppose that C_N is a busy cycle, which is a sum of idle, startup, busy, breakdown periods. Applying the results of Wang et al. (2007), we have:

$$E[I_N] = \frac{N}{\lambda}, \tag{5}$$

$$E[U_N] = \frac{\rho_U}{\lambda}, \tag{6}$$

$$E[B_N] = \frac{E[S](N + \rho_U)}{1 - \rho_H}, \tag{7}$$

$$E[D_N] = \frac{\left(\alpha_1 \mu_{S_1} \mu_{R_1} + \theta \alpha_2 \mu_{S_2} \mu_{R_2}\right)(N + \rho_U)}{1 - \rho_H}, \tag{8}$$

$$E[C_N] = E[I_N] + E[U_N] + E[B_N] + E[D_N] = \frac{N + \rho_U}{\lambda(1 - \rho_H)}, \tag{9}$$

where $\rho_H = \lambda(E[H_1] + \theta E[H_2])$ is the traffic intensity, it should be assumed to be less than unity and $\rho_U = \lambda \mu_U$.

Let L_N denote the expected number of customers in the *N* policy M/(G,G), (G,G), G/1 queue. From the results of Wang et al. (2007), it gives that:

$$L_N = \frac{1}{N + \rho_U}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E[U^2]}{2}\right] + L_H, \tag{10}$$

where $E[H^2] = E\left[H_1^2\right] + 2\theta E[H_1]E[H_2] + \theta E\left[H_2^2\right]$, which can be represented as:

$$\begin{aligned} E\left[H^2\right] = &\left(1 + \alpha_1 \mu_{R_1}\right)^2 E\left[S_1^2\right] + \alpha_1 \mu_{S_1} E\left[R_1^2\right] \\ &+ 2\theta \mu_{S_1} \mu_{S_2}\left(1 + \alpha_1 \mu_{R_1}\right)\left(1 + \alpha_2 \mu_{R_2}\right) \\ &+ \theta\left(1 + \alpha_2 \mu_{R_2}\right)^2 E\left[S_2^2\right] + \theta \alpha_2 \mu_{S_2} E\left[R_2^2\right], \end{aligned} \tag{11}$$

$$L_H = \rho_H + \frac{\lambda^2 E\left[H^2\right]}{2[1 - \rho_H]}. \tag{12}$$

We denote by $I_{p,N}$, $U_{p,N}$, $B_{p,N}$ and $D_{p,N}$ idle, startup, busy, breakdown periods for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue. And let $C_{p,N}$ be a busy cycle for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue. Based on the arguments of Feinberg and Kim (1996), it shows that the system performance measures for the $\langle p,N\rangle$-policy queue is a convex combination of the performance measures for the *N*-policy queue and the performance measures for the $(N+1)$-policy queue. Using the above formulae (5)–(9), we can obtain:

$$E\left[I_{p,N}\right] = pE[I_N] + (1-p)E[I_{N+1}] = \frac{N + 1 - p}{\lambda}, \tag{13}$$

$$E\left[U_{p,N}\right] = pE[U_N] + (1-p)E[U_{N+1}] = \frac{\rho_U}{\lambda}, \tag{14}$$

$$E\left[B_{p,N}\right] = pE[B_N] + (1-p)E[B_{N+1}] = \frac{E[S](N + 1 - p + \rho_U)}{1 - \rho_H}, \tag{15}$$

$$\begin{aligned} E\left[D_{p,N}\right] &= pE[D_N] + (1-p)E[D_{N+1}] \\ &= \frac{(N + 1 - p + \rho_U)\left(\alpha_1 \mu_{S_1} \mu_{R_1} + \theta \alpha_2 \mu_{S_2} \mu_{R_2}\right)}{1 - \rho_H}, \end{aligned} \tag{16}$$

$$E\left[C_{p,N}\right] = pE[C_N] + (1-p)E[C_{N+1}] = \frac{N + 1 - p + \rho_U}{\lambda(1 - \rho_H)}. \tag{17}$$

3.1. The long-run fraction of time measures

We will develop the maximum entropy solutions for steady-state probabilities of the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue. Steady-state probabilities $P_I(n)$, $P_S(n)$, $P_1(n)$, $P_2(n)$, $Q_1(n)$ and $Q_2(n)$ for the entropy formalism are defined as follows:

$P_I(n) \equiv$ probability that there are n customers in the system when the serve is turned off, where $n = 0, 1, 2, \ldots, N-1, N$

$P_S(n) \equiv$ probability that there are n customers in the system when the serve is startup, where $n = N, N+1, \ldots$

$P_1(n) \equiv$ probability that there are n customers in the queue excluding the one being provided *FES*, and the server is in operation, where $n = 1, 2, 3, \ldots$

$P_2(n) \equiv$ probability that there are n customers in the queue excluding the one being provided *SOS*, and the server is in operation, where $n = 1, 2, 3, \ldots$

$Q_1(n) \equiv$ probability that there are n customers in the queue excluding the one being provided *FES*, and the server is in operation but found to be broken down, where $n = 1, 2, 3, \ldots$

$Q_2(n) \equiv$ probability that there are n customers in the queue excluding the one being provided *SOS*, and the server is in operation but found to be broken down, where $n = 1, 2, 3, \ldots$

From Eqs. (13)–(17), we can easily obtain the following probabilities for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue.

The probability that the server is idle given by

$$\sum_{n=0}^{N} P_I(n) = \frac{E\left[I_{p,N}\right]}{E\left[C_{p,N}\right]} = \frac{(N + 1 - p)(1 - \rho_H)}{N + 1 - p + \rho_U}. \tag{18}$$

The probability that the server is startup given by

$$\sum_{n=N}^{\infty} P_S(n) = \frac{E\left[U_{p,N}\right]}{E\left[C_{p,N}\right]} = \frac{\rho_U(1 - \rho_H)}{N + 1 - p + \rho_U}. \tag{19}$$

The probability that the server is busy given by

$$\sum_{n=1}^{\infty} P_1(n) + \sum_{n=1}^{\infty} P_2(n) = \frac{E\left[B_{p,N}\right]}{E\left[C_{p,N}\right]} = \lambda E[S] = \lambda \mu_{S_1} + \theta \lambda \mu_{S_2}. \tag{20}$$

The probability that the server is breakdown given by

$$\sum_{n=1}^{\infty} Q_1(n) + \sum_{n=1}^{\infty} Q_2(n) = \frac{E\left[D_{p,N}\right]}{E\left[C_{p,N}\right]} = \lambda \alpha_1 \mu_{S_1} \mu_{R_1} + \theta \lambda \alpha_2 \mu_{S_2} \mu_{R_2}. \tag{21}$$

For a start, we note that the long-run fraction of time the server is busy when *FES* or *SOS* is provided, and can be represented as $\lambda \mu_{S_1}$ and $\theta \lambda \mu_{S_2}$, respectively. Next, it is noticed that the long-run fraction of time the server is broken down when the *FES* or *SOS* provided, which can also be represented as $\lambda \mu_{S_1} \alpha_1 \mu_{R_1}$ and $\theta \lambda \mu_{S_2} \alpha_2 \mu_{R_2}$, respectively.

3.2. The expected number of customers in the system

Let T_N^c, T_{N+1}^c and $T_{p,N}^c$ denote the cumulative amount of time that all customers spent in the system during a busy cycle for the *N*-, $(N+1)$- and $\langle p,N\rangle$-policies M/(G,G), (G,G), G/1 queue. Following the results of Feinberg and Kim (1996), we can obtain:

$$E\left[T_N^c\right] = L_N E[C_N] = \frac{1}{\lambda(1-\rho_H)}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda^2 E\left(U^2\right)}{2}\right] + \frac{L_H(N+\rho_U)}{\lambda(1-\rho_H)}, \quad (22)$$

where L_H is given in Eq. (12).

It follows that:

$$E\left[T_{p,N}^c\right] = pE\left[T_N^c\right] + (1-p)E\left[T_{N+1}^c\right] = \frac{1}{\lambda(1-\rho_H)}\left[\frac{N(N+1-2p)}{2} + (N+1-p)\rho_U + \frac{\lambda^2 E\left(U^2\right)}{2}\right] + \frac{L_H(N+1-p+\rho_U)}{\lambda(1-\rho_H)}. \quad (23)$$

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

$$L_{p,N} = \frac{E\left[T_{p,N}^c\right]}{E\left[C_{p,N}\right]} = \frac{1}{N+1-p+\rho_U}\left[\frac{N(N+1-2p)}{2} + (N+1-p)\rho_U + \frac{\lambda^2 E\left[U^2\right]}{2}\right] + L_H, \quad (24)$$

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:

$$L_{p,N} = \Theta L_N + (1-\Theta)L_{N+1}, \quad (25)$$

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

It is easy to demonstrate that Eq. (25) is identical Eq. (24). Additionally, Eq. (24) is in accordance with expression (3) of Wang et al. (2007)if we set $p = 1$ and $\theta = 0$.

3.3. Some known steady-state probabilities

Applying a convex combination property, the probability $P_I(0)$ for a $\langle p,N\rangle$-policy is a convex combination of the probability $P_I^N(0)$ for an N-policy and the $P_I^{N+1}(0)$ for an $(N+1)$-policy, where $P_I^N(0) = (1-\rho_H)/(N+\rho_U)$ and $P_I^{N+1}(0) = (1-\rho_H)/(N+1+\rho_U)$. Thus, we have:

$$P_I(0) = \Theta P_I^N(0) + (1-\Theta)P_I^{N+1}(0) = \frac{1-\rho_H}{N+1-p+\rho_U}. \quad (26)$$

From Eq. (18) and the results of Wang et al. (2007), we have:

$$\sum_{n=0}^{N} P_I(n) = NP_I(0) + P_I(N) = \frac{(N+1-p)(1-\rho_H)}{N+1-p+\rho_U}.$$

This gives:

$$P_I(N) = \frac{(1-p)(1-\rho_H)}{N+1-p+\rho_U}. \quad (27)$$

4. Improved maximum entropy results

Exact probability distributions of the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue have not been found. Therefore, we employ the improved maximum entropy principle to estimate probability distributions of the number of customers given several known results. In this section, we will develop the improved maximum entropy solutions for the steady-state probabilities of the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue.

4.1. The improved maximum entropy model

In order to derive the approximate steady-state probabilities $P_S(n)$, $P_i(n)$ $(i = 1,2)$, $Q_i(n)$ $(i = 1,2)$, we formulate the maximum entropy model in the following. Because that the exact results for $P_I(0)$ and $P_I(N)$ are known, the improved entropy function Y of the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue can be formed as:

$$Y = -\sum_{n=N}^{\infty} P_S(n)\ln P_S(n) - \sum_{n=1}^{\infty} P_1(n)\ln P_1(n) - \sum_{n=1}^{\infty} P_2(n)\ln P_2(n) - \sum_{n=1}^{\infty} Q_1(n)\ln Q_1(n) - \sum_{n=1}^{\infty} Q_2(n)\ln Q_2(n). \quad (28)$$

The improved maximum entropy solutions for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue are obtained by maximizing Eq. (28) subject to the following six constraints, written as:

1. The probability that the server is startup:

$$\sum_{n=N}^{\infty} P_S(n) = \frac{\rho_U(1-\rho_H)}{N+1-p+\rho_U} = \Pi\rho_U(1-\rho_H), \quad (29)$$

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

2. The probability that the server is busy of providing *FES*:

$$\sum_{n=1}^{\infty} P_1(n) = \lambda\mu_{S_1} = \rho_1. \quad (30)$$

3. The probability that the server is busy of providing *SOS*:

$$\sum_{n=1}^{\infty} P_2(n) = \theta\lambda\mu_{S_2} = \theta\rho_2. \quad (31)$$

4. The probability that the server is broken down when *FES* is provided:

$$\sum_{n=1}^{\infty} Q_1(n) = \rho_1\alpha_1\mu_{R_1}. \quad (32)$$

5. The probability that the server is broken down when *SOS* is provided:

$$\sum_{n=1}^{\infty} Q_2(n) = \theta\rho_2\alpha_2\mu_{R_2}. \quad (33)$$

6. The expected number of customers in the system when the server is not idle:

$$\sum_{n=N}^{\infty} nP_S + \sum_{n=1}^{\infty} nP_1(n) + \sum_{n=1}^{\infty} nP_2(n) + \sum_{n=1}^{\infty} nQ_1(n) + \sum_{n=1}^{\infty} nQ_2(n) = L_{p,N} - L_I, \quad (34)$$

where L_I is the expected length of customers as the server is idle and can be expressed as follows:

$$L_I = \frac{N(N-1)}{2}P_I(0) + NP_I(N) = \frac{N(N+1-2p)(1-\rho_H)}{2(N+1-p+\rho_U)}. \quad (35)$$

In Eqs. (29)–(34), Eq. (29) is multiplied by δ_1, Eq. (30) is multiplied by δ_2, Eq. (31) is multiplied by δ_3, Eq. (32) is multiplied by δ_4, Eq. (33) is multiplied by δ_5, Eq. (34) is multiplied by δ_6, Thus, the Lagrangian function y is given by

$$\begin{aligned} y = &-\sum_{n=N}^{\infty} P_S(n)\ln P_S(n) - \sum_{n=1}^{\infty} P_1(n)\ln P_1(n) \\ &- \sum_{n=1}^{\infty} P_2(n)\ln P_2(n) - \sum_{n=1}^{\infty} Q_1(n)\ln Q_1(n) \\ &- \sum_{n=1}^{\infty} Q_2(n)\ln Q_2(n) - \delta_1\left[\sum_{n=N}^{\infty} P_S(n) - \Pi\rho_U(1-\rho_H)\right] \\ &- \delta_2\left[\sum_{n=1}^{\infty} P_1(n) - \rho_1\right] - \delta_3\left[\sum_{n=1}^{\infty} P_2(n) - \theta\rho_2\right] \\ &- \delta_4\left[\sum_{n=1}^{\infty} Q_1(n) - \rho_1\alpha_1\mu_{R_1}\right] - \delta_5\left[\sum_{n=1}^{\infty} Q_2(n) - \theta\rho_2\alpha_2\mu_{R_2}\right] \\ &- \delta_6\left[\sum_{n=N}^{\infty} nP_S(n) + \sum_{n=1}^{\infty} nP_1(n) + \sum_{n=1}^{\infty} nP_2(n) + \sum_{n=1}^{\infty} nQ_1(n)\right. \\ &\left. + \sum_{n=1}^{\infty} nQ_2(n) - L_{p,N} + L_I\right], \end{aligned} \tag{36}$$

where $\delta_1 - \delta_6$ are the Lagrangian multipliers corresponding to constrains (29)–(34), respectively.

4.2. The improved maximum entropy solutions

To find the improved maximum entropy solutions $P_S(n)$, $P_i(n)$ ($i = 1,2$) and $Q_i(n)$ ($i = 1,2$), maximizing in (28) subject to constrains (29)–(34) is equivalent to maximizing (36). The improved maximum entropy solutions are obtained by taking the partial derivatives of y with respect to $P_S(n)$, $P_i(n)(i = 1,2)$, $Q_i(n)$ ($i = 1,2$) and setting the results equal to zero, namely:

$$\frac{\partial y}{\partial P_S(n)} = -\ln P_S(n) - 1 - \delta_1 - \delta_6 n = 0, \quad n = N, N+1, \ldots \tag{37}$$

$$\frac{\partial y}{\partial P_1(n)} = -\ln P_1(n) - 1 - \delta_2 - \delta_6 n = 0, \quad n = 1,2,\ldots \tag{38}$$

$$\frac{\partial y}{\partial P_2(n)} = -\ln P_2(n) - 1 - \delta_3 - \delta_6 n = 0, \quad n = 1,2,\ldots \tag{39}$$

$$\frac{\partial y}{\partial Q_1(n)} = -\ln Q_1(n) - 1 - \delta_4 - \delta_6 n = 0, \quad n = 1,2,\ldots \tag{40}$$

$$\frac{\partial y}{\partial Q_2(n)} = -\ln Q_2(n) - 1 - \delta_5 - \delta_6 n = 0, \quad n = 1,2,\ldots. \tag{41}$$

It follows from Eqs. (37)–(41) that:

$$P_S(n) = e^{-(1+\delta_1)}e^{-\delta_6 n}, \quad n = N, N+1, \ldots \tag{42}$$

$$P_1(n) = e^{-(1+\delta_2)}e^{-\delta_6 n}, \quad n = 1,2,\ldots \tag{43}$$

$$P_2(n) = e^{-(1+\delta_3)}e^{-\delta_6 n}, \quad n = 1,2,\ldots \tag{44}$$

$$Q_1(n) = e^{-(1+\delta_4)}e^{-\delta_6 n}, \quad n = 1,2,\ldots \tag{45}$$

$$Q_2(n) = e^{-(1+\delta_5)}e^{-\delta_6 n}, \quad n = 1,2,\ldots. \tag{46}$$

Let $\omega_i = e^{-(1+\delta_i)}$ for $1 \leqslant i \leqslant 5$, and $\omega_6 = e^{-\delta_6}$. We transform Eqs. (42)–(46) in terms of ω_i ($1 \leqslant i \leqslant 6$) given by

$$P_S(n) = \omega_1\omega_6^n, \quad n = N, N+1, \ldots \tag{47}$$

$$P_1(n) = \omega_2\omega_6^n, \quad n = 1,2,\ldots \tag{48}$$

$$P_2(n) = \omega_3\omega_6^n, \quad n = 1,2,\ldots \tag{49}$$

$$Q_1(n) = \omega_4\omega_6^n, \quad n = 1,2,\ldots \tag{50}$$

$$Q_2(n) = \omega_5\omega_6^n, \quad n = 1,2,\ldots. \tag{51}$$

Substituting Eqs. (47)–(51) into Eqs. (29)–(33), respectively, yields:

$$\omega_1 = \frac{\Pi\rho_U(1-\rho_H)(1-\omega_6)}{\omega_6^N}, \tag{52}$$

$$\omega_2 = \frac{\rho_1(1-\omega_6)}{\omega_6}, \tag{53}$$

$$\omega_3 = \frac{\theta\rho_2(1-\omega_6)}{\omega_6}, \tag{54}$$

$$\omega_4 = \frac{\rho_1\alpha_1\mu_{R_1}(1-\omega_6)}{\omega_6}, \tag{55}$$

$$\omega_5 = \frac{\theta\rho_2\alpha_2\mu_{R_2}(1-\omega_6)}{\omega_6}. \tag{56}$$

Substituting Eqs. (47)–(51) into Eq. (34) and taking the algebraic manipulations, we obtain:

$$\omega_6 = 1 - \frac{\Pi\rho_U(1-\rho_H) + \rho_H}{L_{p,N} - \Pi(1-\rho_H)\left[\frac{N(N+1-2p)}{2} + (N-1)\rho_U\right]}. \tag{57}$$

Substituting Eqs. (52)–(57) into Eqs. (47)–(51), respectively, we finally get:

$$P_S(n) = \Pi\rho_U(1-\rho_H)(1-\omega_6)\omega_6^{n-N}, \quad n = N, N+1, \ldots \tag{58}$$

$$P_1(n) = \rho_1(1-\omega_6)\omega_6^{n-1}, \quad n = 1,2,\ldots \tag{59}$$

$$P_2(n) = \theta\rho_2(1-\omega_6)\omega_6^{n-1}, \quad n = 1,2,\ldots \tag{60}$$

$$Q_1(n) = \rho_1\alpha_1\mu_{R_1}(1-\omega_6)\omega_6^{n-1}, \quad n = 1,2,\ldots \tag{61}$$

$$Q_2(n) = \theta\rho_2\alpha_2\mu_{R_2}(1-\omega_6)\omega_6^{n-1}, \quad n = 1,2,\ldots. \tag{62}$$

5. The exact and approximate expected waiting time in the system

In this section, we first derive the exact expected waiting time in the system by using Little's formula. Through the maximum entropy principle, the approximate formulae of the expected waiting time in the system for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue is developed.

5.1. The exact expected waiting time in the system

Let $W_S(N)$, $W_S(N+1)$ and $W_S(p,N)$ denote the exact expected waiting time in the system for the N-, $(N+1)$- and $\langle p,N\rangle$-policies, respectively. Using Eqs. (10) and (24) in Little's formula, we see that:

$$W_S(N) = \frac{L_N}{\lambda} = \frac{1}{N+\rho_U}\left[\frac{N(N-1)}{2\lambda} + N\mu_U + \frac{\lambda E[U^2]}{2}\right] + \frac{L_H}{\lambda}, \tag{63}$$

$$W_S(N+1) = \frac{L_N}{\lambda} = \frac{1}{N+1+\rho_U}\left[\frac{N(N+1)}{2\lambda} + (N+1)\mu_U + \frac{\lambda E[U^2]}{2}\right] + \frac{L_H}{\lambda}, \tag{64}$$

$$W_S(p,N) = \frac{L_{p,N}}{\lambda} = \frac{1}{N+1-p+\rho_U} \times \left[\frac{N(N+1-2p)}{2\lambda} + (N+1-p)\mu_U + \frac{\lambda E[U^2]}{2}\right] + \frac{L_H}{\lambda}. \tag{65}$$

From Feinberg and Kim (1996), we know that $W_S(p,N)$ is a convex combination of $W_S(N)$ and $W_S(N+1)$. It follows that:

$$W_S(p,N) = \frac{p(N+\rho_U)}{N+1-p+\rho_U}W_S(N) + \left[1 - \frac{p(N+\rho_U)}{N+1-p+\rho_U}\right]W_S(N+1). \tag{66}$$

Substituting Eqs. (63) and (64) into Eq. (66), we have the same result shown in Eq. (65). Thus, we demonstrate that the relationships given by Eqs. (65) and (66) are seen to hold.

5.2. The approximate expected waiting time in the system

The idle state, the startup state, the busy state, and the repair state are defined as follows:

(1) Idle state 1 denoted by I_1: the server is turned off, and the number of customers waiting in the system is less than or equal to $N-1$.
(2) Idle state 2 denoted by I_2: the server is turned off, and the number of customers waiting in the system is equal to$N$.
(3) Startup state denoted by U: the server begins startup, and the number of customers waiting in the system is greater than or equal to N.
(4) Busy state when *FES* is provided denoted by B_1: the server is busy and provides *FES* to a customer.
(5) Busy state when *SOS* is provided denoted by B_2: the server is busy and provides *SOS* to a customer.
(6) Repair state when *FES* is provided denoted by R_1: the server is broken down when *FES* is provided and being repaired.
(7) Repair state when *SOS* is provided denoted by R_2: the server is broken down when *SOS* is provided and being repaired.

We wish to find the expected waiting time of an arbitrary customer C at the state I_1, I_2, U, B_1, B_2, R_1 and R_2. Suppose an arbitrary customer C finds n customers waiting in the queue for service in front of him, while the system is at any one of the states I_1, I_2, U, B_1, B_2, R_1 and R_2 are described, respectively, as follows:

(1) In idle state I_1: Note that the idle state immediately is switched to startup state after an arbitrary customer C arrives and n customers in front of him are waiting for service. The server will begin startup after $(N-n-1)$ customers arrive with probability p or after $(N-n)$ customers arrive with probability $(1-p)$ in the system. Thus customer C will be served until $(N-n-1)$ customers arrive with probability p or $(N-n)$ customers arrive with probability $(1-p)$, and n customers in front of him are waiting for service. Hence, customer C must wait (i) the mean residual idle time, (ii) the service time of n customers in the system and (iii) the startup time before providing *FES*. From the inferences of (i)–(iii), the expected waiting time of customer C at the idle state I_1 is
$$\frac{(N-n-1)p}{\lambda}+\frac{(N-n)(1-p)}{\lambda}+\mu_U+nE[S] =\frac{N-n-p}{\lambda}+\mu_U+nE[S].$$
(2) In idle state I_2: The server will begin startup when there are N customers present in the system. Thus customer C will be served when no customers in front of him waiting for service. The expected waiting time of customer C at the idle state I_2 is $\mu_U + NE[S]$.
(3) In startup state U: We derive the expected waiting time of customer C at the startup state in the following. Let us define:
$U_r(t) \equiv$ remaining startup time for the server begin startup.
Following Borthakur et al. (1987), the cumulative distribution function (c.d.f.) of $U_r(t)$ is given by
$$F_{U_r}(t)=\Pr\{U_r(t)\leqslant t\}=\frac{1}{\mu_U}\int_0^t[1-D(x)]dx,$$
where $D(x)$ is the c.d.f. of startup time. Let $E[U_r]$ be the mean remaining startup time. It implies that $E[U_r] = E[U^2]/2\mu_U$. Thus we obtain the expected waiting time of customer C at the startup state is $nE[S] + E[U^2]/2\mu_U$.
(4) In busy states B_1 and B_2: Since the server is busy and keeps working, the customer C only waits n customers who demand the server in front of him. The expected waiting time at the busy states B_1 and B_2 are $nE[S]$, respectively.
(5) In repair states R_1 and R_2: According to the same argument as (3), we have the expected waiting time of an arbitrary customer C at the repair states R_1 and R_2 are $nE[S]+E\left[R_1^2\right]/2\mu_{R_1}$ and $nE[S]+E\left[R_2^2\right]/2\mu_{R_2}$, respectively.

Utilizing the listed above results, we obtain the approximate expected waiting time in the queue, $W_q^*(p,N)$, given by
$$\begin{aligned}W_q^*(p,N)=&\sum_{n=0}^{N-1}\left(\frac{N-n-p}{\lambda}+\mu_U+nE[S]\right)P_I(0)\\&+(\mu_U+NE[S])P_I(N)+\sum_{n=N}^{\infty}\left(nE[S]+\frac{E[U^2]}{2\mu_U}\right)P_S(n)\\&+\sum_{n=1}^{\infty}(nE[S])P_1(n)+\sum_{n=1}^{\infty}(nE[S])P_2(n)\\&+\sum_{n=1}^{\infty}\left(nE[S]+E\left[R_1^2\right]/2\mu_{R_1}\right)Q_1(n)\\&+\sum_{n=1}^{\infty}\left(nE[S]+E\left[R_2^2\right]/2\mu_{R_2}\right)Q_2(n).\end{aligned}\tag{67}$$

Substituting Eqs. (26), (27), ()()()()(58)–(62) into Expression (66), the approximate expected waiting in the queue is given by
$$\begin{aligned}W_q^*(p,N)=&\frac{N(N+1-2p)(1-\rho_H)}{2\lambda(N+1-p+\rho_U)}\\&+\frac{\left(2\mu_U(N+1-p)+\lambda E\left[U^2\right]\right)(1-\rho_H)}{2(N+1-p+\rho_U)}\\&+\rho W_S(p,N)+\frac{E\left[R_1^2\right]\rho_1\alpha_1}{2}+\frac{\theta E\left[R_2^2\right]\rho_2\alpha_2}{2},\end{aligned}\tag{68}$$
where the derivation of Eq. (68) is shown in Appendix. Consequently, we again use Little's formula to obtain the approximate expected waiting time in the system as follows:
$$\begin{aligned}W_S^*(p,N)=&\frac{N(N+1-2p)(1-\rho_H)}{2\lambda(N+1-p+\rho_U)}\\&+\frac{\left(2\mu_U(N+1-p)+\lambda E\left[U^2\right]\right)(1-\rho_H)}{2(N+1-p+\rho_U)}\\&+\rho W_S(p,N)+\frac{E\left[R_1^2\right]\rho_1\alpha_1}{2}+\frac{\theta E\left[R_2^2\right]\rho_2\alpha_2}{2}+E[H].\end{aligned}\tag{69}$$

6. Comparative analysis between exact and approximate results

This section aims to examine the accuracy of the approximate results based on the improved maximum entropy principle. We provide numerical comparisons between the exact results and the approximate results, including various service time, startup time and repair time distribution functions. There are three subsections in the following:

(1) Comparative analysis for the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D),$M/1$ queue.

(2) Comparative analysis for the $\langle p,N\rangle$-policy $M/(M,D),(E_2,E_3),D/1$ queue.
(3) Comparative analysis for the $\langle p,N\rangle$-policy $M/(E_2,M),(D,E_4),E_3/1$ queue.
Here, M is an exponential distribution, D is a deterministic distribution and E_k is a k-stage Erlang distribution.

6.1. Comparative analysis for the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D),$M/1$ queue

We perform a comparative analysis between the exact $W_S(p,N)$ and the approximate $W_S^*(p,N)$ for the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D),$M/1$ queue. For this queueing system, we have:

Table 1
The relative error percentage for the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D), $M/1$ queue ($\lambda = 0.5, \mu_1 = 1.0, \mu_2 = 2.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.10, \beta_1 = 3.0, \beta_2 = 4.0, \theta = 0.4$).

N	p						
	0.01	0.1	0.3	0.5	0.7	0.9	0.99
2	0.584	0.604	0.656	0.720	0.800	0.904	0.961
4	0.122	0.134	0.165	0.200	0.240	0.288	0.312
6	0.141	0.133	0.114	0.092	0.068	0.042	0.029
8	0.311	0.305	0.292	0.277	0.261	0.244	0.236
10	0.428	0.424	0.414	0.404	0.393	0.381	0.375
12	0.515	0.512	0.504	0.497	0.488	0.480	0.476
14	0.581	0.579	0.573	0.567	0.561	0.554	0.551
16	0.634	0.632	0.627	0.622	0.617	0.612	0.610
18	0.676	0.675	0.671	0.667	0.663	0.659	0.657
20	0.712	0.710	0.707	0.704	0.700	0.697	0.695

Table 3
The relative error percentage for the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3), $D/1$ queue ($\lambda = 0.5, \mu_1 = 1.0, \mu_2 = 2.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.10, \beta_1 = 3.0, \beta_2 = 4.0, \theta = 0.4$).

N	p						
	0.01	0.1	0.3	0.5	0.7	0.9	0.99
2	0.855	0.879	0.941	1.016	1.110	1.233	1.302
4	0.314	0.329	0.364	0.405	0.453	0.508	0.536
6	0.007	0.017	0.040	0.065	0.092	0.124	0.139
8	0.189	0.183	0.167	0.150	0.132	0.112	0.103
10	0.326	0.321	0.310	0.298	0.285	0.271	0.265
12	0.426	0.423	0.414	0.405	0.396	0.385	0.381
14	0.503	0.500	0.494	0.487	0.479	0.472	0.468
16	0.564	0.562	0.556	0.551	0.545	0.539	0.536
18	0.613	0.611	0.607	0.603	0.598	0.593	0.590
20	0.654	0.653	0.649	0.645	0.641	0.637	0.635

Table 2
Comparison of exact $W_S(p,N)$ and approximate W_S^* for the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D), $M/1$ queue ($N = 8$).

	$W_S(p,N)$			$W_S^*(p,N)$			$RE(\%)$		
	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$
λ	Case 1: $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1, \alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
0.1	40.650	39.206	37.658	40.579	39.138	37.593	0.174	0.173	0.171
0.2	21.305	20.583	19.809	21.247	20.528	19.758	0.272	0.267	0.260
0.4	12.220	11.859	11.473	12.183	11.825	11.442	0.305	0.288	0.268
0.6	10.907	10.667	10.410	10.869	10.632	10.378	0.356	0.335	0.311
0.8	53.133	52.953	52.761	52.349	52.173	51.983	1.475	1.474	1.473
(μ_1,μ_2)	Case 2: $\lambda = 0.5$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.8,1.0)	17.393	17.105	16.796	17.312	17.028	16.724	0.467	0.449	0.428
(1.0,1.0)	12.524	12.236	11.927	12.492	12.208	11.903	0.256	0.231	0.203
(1.0,2.0)	10.933	10.645	10.336	10.900	10.615	10.309	0.298	0.277	0.253
(1.0,3.0)	10.593	10.305	9.996	10.548	10.263	9.957	0.421	0.405	0.387
(1.5,3.0)	9.319	9.031	8.722	9.290	9.004	8.697	0.313	0.300	0.285
(2.0,3.0)	8.921	8.633	8.324	8.896	8.609	8.302	0.286	0.276	0.264
(α_1,α_2)	Case 3: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.05,0.10)	10.933	10.645	10.336	10.900	10.615	10.309	0.298	0.277	0.253
(0.05,0.20)	10.961	10.673	10.364	10.903	10.618	10.313	0.535	0.514	0.489
(0.10,0.05)	11.029	10.740	10.432	10.918	10.635	10.331	1.006	0.985	0.961
(0.10,0.20)	11.073	10.785	10.476	10.922	10.640	10.338	1.361	1.340	1.315
(0.20,0.05)	11.264	10.976	10.667	10.964	10.686	10.388	2.659	2.638	2.614
(0.20,0.10)	11.280	10.992	10.683	10.967	10.689	10.391	2.777	2.756	2.732
(β_1,β_2)	Case 4: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $\gamma = 3.0$, $\theta = 0.4$								
(3.0,2.0)	10.963	10.675	10.366	10.904	10.620	10.315	0.536	0.514	0.490
(3.0,4.0)	10.933	10.645	10.336	10.900	10.615	10.309	0.298	0.277	0.253
(3.0,6.0)	10.923	10.635	10.326	10.899	10.614	10.308	0.219	0.198	0.174
(6.0,2.0)	10.907	10.619	10.310	10.894	10.608	10.302	0.123	0.101	0.077
(6.0,4.0)	10.877	10.589	10.280	10.890	10.604	10.297	0.114	0.136	0.160
(6.0,6.0)	10.868	10.580	10.271	10.889	10.603	10.295	0.193	0.215	0.239
γ	Case 5: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $\beta_1,\beta_2 = (3.0,4.0)$, $\theta = 0.4$								
2.0	11.029	10.741	10.433	10.995	10.710	10.405	0.305	0.284	0.261
3.0	10.933	10.645	10.336	10.900	10.615	10.309	0.298	0.277	0.253
4.0	10.885	10.597	10.288	10.853	10.568	10.262	0.295	0.273	0.249
5.0	10.857	10.568	10.259	10.825	10.540	10.234	0.293	0.271	0.247
6.0	10.838	10.550	10.240	10.807	10.521	10.215	0.291	0.270	0.245
θ	Case 6: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, β_1, $\beta_2 = (3.0,4.0)$, $\gamma = 3.0$								
0.2	10.469	10.180	9.871	10.406	10.120	9.814	0.599	0.588	0.577
0.4	10.933	10.645	10.334	10.900	10.615	10.309	0.298	0.277	0.253
0.6	11.507	11.219	10.910	11.512	11.228	10.922	0.047	0.080	0.116
0.8	12.249	11.961	11.652	12.300	12.015	11.711	0.411	0.453	0.500
1.0	13.271	12.983	12.674	13.371	13.087	12.782	0.752	0.801	0.856

$$E[S_1]=\frac{1}{\mu_1},\quad E\left[S_1^2\right]=\frac{2}{\mu_1^2},\quad E[S_2]=\frac{1}{\mu_2},\quad E\left[S_2^2\right]=\frac{3}{2\mu_2^2},$$
$$E[R_1]=\frac{1}{\beta_1},\quad E\left[R_1^2\right]=\frac{2}{\beta_1^2},\quad E[R_2]=\frac{1}{\beta_2},\quad E\left[R_2^2\right]=\frac{1}{\beta_2^2},$$
$$E[U]=\frac{1}{\gamma},\quad E[U^2]=\frac{2}{\gamma^2}.$$

Firstly, we fix $\lambda = 0.5$, $\mu_1 = 1.0$, $\mu_2 = 2.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.10$, $\beta_1 = 3.0$, $\beta_2 = 4.0$, $\gamma = 3.0$, $\theta = 0.4$, and choose various values of (p,N). The accuracy of the approximate values is assessed by the relative error:

$$RE=\left|\frac{W_S(p,N)-W_S^*(p,N)}{W_S(p,N)}\right|\times 100\%.$$

The relative error percentage for the $\langle p,N\rangle$-policy $M/(M,E_2),(M,D),M/1$ queue under various values p and N are shown in Table 1. We observe from Table 1 that (i) for fix p, the relative error percentage decreases when N ranges from 2 to 6 and increases when N ranges from 8 to 20; (ii) if N is from 2 to 4 and fixed it, the relative error percentage increases in p; (iii) if N is from 6 to 20 and fixed it, the relative error percentage decreases in p; (iii) the relative error percentage in Table 1 is below 1%.

Next, we set $N = 8$ and consider the different values $p = 0.2$, 0.5 and 0.8. Choosing the various values of λ, (μ_1,μ_2), (α_1,α_2), (β_1,β_2), γ and θ. The numerical results are obtained by considering the following six cases:

Case 1: We fix $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$ and vary λ from 0.1 to 0.8.

Case 2: We fix $\lambda = 0.5$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$ and consider various values of $(\mu_1,\mu_2) = (0.8,1.0)$, $(1.0,1.0)$, $(1.0,2.0)$, $(1.0,3.0)$, $(1.5,3.0)$, $(2.0,3.0)$.

Table 5
The relative error percentage for the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4), $E_3/1$ queue ($\lambda = 0.5, \mu_1 = 1.0, \mu_2 = 2.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.10, \beta_1 = 3.0, \beta_2 = 4.0, \theta = 0.4$).

N	p						
	0.01	0.1	0.3	0.5	0.7	0.9	0.99
2	3.088	3.142	3.281	3.452	3.669	3.953	4.111
4	1.878	1.910	1.989	2.080	2.185	2.308	2.370
6	1.209	1.229	1.279	1.333	1.393	1.461	1.495
8	0.786	0.800	0.833	0.869	0.908	0.951	0.972
10	0.494	0.504	0.528	0.554	0.581	0.611	0.625
12	0.281	0.289	0.307	0.326	0.346	0.368	0.378
14	0.119	0.125	0.139	0.154	0.169	0.186	0.193
16	0.009	0.004	0.007	0.019	0.031	0.044	0.050
18	0.112	0.108	0.099	0.090	0.080	0.069	0.064
20	0.197	0.194	0.186	0.178	0.170	0.162	0.158

Table 4
Comparison of exact $W_S(p,N)$ and approximate $W_S^*(p,N)$ for the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3),$D/1$ queue ($N = 8$).

	$W_S(p,N)$			$W_S^*(p,N)$			$RE(\%)$		
	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$
λ	Case 1: $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
0.1	40.646	39.202	37.654	40.578	39.137	37.592	0.168	0.166	0.164
0.2	21.296	20.575	19.801	21.243	19.755	19.755	0.247	0.241	0.233
0.4	12.196	11.835	11.449	12.169	11.182	11.429	0.218	0.198	0.175
0.6	10.842	10.602	10.344	10.820	10.583	10.329	0.205	0.180	0.152
0.8	52.160	51.979	51.787	51.414	51.237	51.047	1.430	1.429	1.428
(μ_1,μ_2)	Case 2: $\lambda = 0.5$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.8, 1.0)	17.056	16.768	16.459	17.033	16.749	16.445	0.137	0.113	0.086
(1.0, 1.0)	12.336	12.048	11.739	12.359	12.075	11.770	0.186	0.222	0.263
(1.0, 2.0)	10.894	10.606	10.297	10.875	10.590	10.284	0.175	0.150	0.122
(1.0, 3.0)	10.574	10.286	9.977	10.536	10.250	9.944	0.365	0.347	0.327
(1.5, 3.0)	9.306	9.017	8.708	9.282	8.996	8.689	0.250	0.234	0.217
(2.0, 3.0)	8.909	8.621	8.312	8.890	8.603	8.296	0.220	0.207	0.194
(α_1,α_2)	Case 3: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.05, 0.10)	10.894	10.606	10.297	10.875	10.590	10.284	0.175	0.150	0.122
(0.05, 0.20)	10.921	10.633	10.324	10.877	10.592	10.287	0.406	0.381	0.353
(0.10, 0.05)	10.989	10.700	10.391	10.891	10.608	10.305	0.886	0.862	0.834
(0.10, 0.20)	11.030	10.742	10.433	10.894	10.612	10.310	1.233	1.208	1.180
(0.20, 0.05)	11.218	10.930	10.621	10.933	10.655	10.356	2.541	2.517	2.490
(0.20, 0.10)	11.233	10.945	10.636	10.935	10.657	10.359	2.657	2.632	2.605
(β_1,β_2)	Case 4: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $\gamma = 3.0$, $\theta = 0.4$								
(3.0, 2.0)	10.923	10.635	10.326	10.879	10.594	10.290	0.407	0.381	0.353
(3.0, 4.0)	10.894	10.606	10.297	10.875	10.590	10.284	0.175	0.150	0.122
(3.0, 6.0)	10.885	10.597	10.288	10.875	10.589	10.283	0.098	0.073	0.046
(6.0, 2.0)	10.869	10.581	10.272	10.870	10.584	10.278	0.007	0.032	0.060
(6.0, 4.0)	10.841	10.552	10.243	10.867	10.580	10.273	0.238	0.263	0.291
(6.0, 6.0)	10.832	10.544	10.235	10.866	10.579	10.272	0.315	0.340	0.368
γ	Case 5: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\theta = 0.4$								
2.0	10.987	10.698	10.390	10.967	10.682	10.376	0.183	0.158	0.131
3.0	10.894	10.606	10.297	10.875	10.590	10.284	0.175	0.150	0.122
4.0	10.848	10.560	10.251	10.830	10.544	10.238	0.171	0.146	0.118
5.0	10.821	10.532	10.223	10.802	10.517	10.211	0.169	0.144	0.115
6.0	10.802	10.514	10.204	10.784	10.499	10.192	0.167	0.142	0.114
θ	Case 6: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$								
0.2	10.449	10.161	9.852	10.393	10.108	9.801	0.535	0.523	0.509
0.4	10.894	10.606	10.297	10.875	10.590	10.284	0.175	0.150	0.122
0.6	11.444	11.155	10.846	11.470	11.185	10.880	0.226	0.264	0.305
0.8	12.153	11.865	11.556	12.231	11.946	11.642	0.641	0.689	0.743
1.0	13.126	12.838	12.529	13.261	12.977	12.672	1.027	1.082	1.145

Case 3: We fix $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0, 2.0)$, $(\beta_1,\beta_2) = (3.0, 4.0)$, $\gamma = 3.0$, $\theta = 0.4$ and consider various values of $(\alpha_1,\alpha_2) = (0.05, 0.10)$, $(0.05, 0.20)$, $(0.10, 0.05)$, $(0.10, 0.20)$, $(0.20, 0.05)$, $(0.20, 0.10)$.

Case 4: We fix $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0, 2.0)$, $(\alpha_1,\alpha_2) = (0.05, 0.10)$, $\gamma = 3.0$, $\theta = 0.4$ and consider various values of $(\beta_1,\beta_2) = (3.0, 2.0)$, $(3.0, 4.0)$, $(3.0, 6.0)$, $(6.0, 2.0)$, $(6.0, 4.0)$, $(6.0, 6.0)$.

Case 5: We fix $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0, 2.0)$, $(\alpha_1,\alpha_2) = (0.05, 0.10)$, $(\beta_1,\beta_2) = (3.0, 4.0)$, $\theta = 0.4$ and vary γ from 2.0 to 6.0.

Case 6: We fix $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0, 2.0)$, $(\alpha_1,\alpha_2) = (0.05, 0.10)$, $(\beta_1,\beta_2) = (3.0, 4.0)$, $\gamma = 3.0$ and vary θ from 0.2 to 1.0.

Numerical results of the $\langle p,N\rangle$-policy $M/(M,E_2)$, (M,D), $M/1$ queue are shown in Table 2. It can be found that the approximations are good because that the relative error percentages are very small (0–2.8%).

6.2. Comparative analysis for the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3), $D/1$ queue

We perform a comparative analysis between the exact $W_S(p,N)$ and the approximate $W_S^*(p,N)$ for the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3), $D/1$ queue. For this queueing system, we have:

$$E[S_1] = \frac{1}{\mu_1},\quad E\left[S_1^2\right] = \frac{2}{\mu_1^2},\quad E[S_2] = \frac{1}{\mu_2},\quad E\left[S_2^2\right] = \frac{1}{\mu_2^2},$$
$$E[R_1] = \frac{1}{\beta_1},\quad E\left[R_1^2\right] = \frac{3}{2\beta_1^2},\quad E[R_2] = \frac{1}{\beta_2},\quad E\left[R_2^2\right] = \frac{4}{3\beta_2^2},$$
$$E[U] = \frac{1}{\gamma},\quad E[U^2] = \frac{1}{\gamma^2}.$$

The relative error percentage for the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3), $D/1$ queue under various values p and N are shown in Table 3. It reveals that (i) for fix p, the relative error percentage decreases when N ranges from 2 to 6 and increases when N ranges from 8 to 20; (ii) if N is from 2 to 6 and fixed it, the relative error percentage increases in p; (iii) if N is from 8 to 20 and fixed it, the relative error percentage decreases in p; (iv) the relative error percentage in Table 3 is below 1.4%.

Numerical results of the $\langle p,N\rangle$-policy $M/(M,D)$, (E_2,E_3), $D/1$ queue summarized in Table 4 for the above six cases. Table 4 indicates that the relative error percentages are very small (0–2.7%).

6.3. Comparative analysis for the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4), $E_3/1$ queue

We perform a comparative analysis between the exact $W_S(p,N)$ and the approximate $W_S^*(p,N)$ for the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4), $E_3/1$ queue. For this queueing system, we have:

Table 6
Comparison of exact $W_S(p,N)$ and approximate $W_S^*(p,N)$ for the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4), $E_3/1$ queue ($N = 8$).

	$W_S(p,N)$			$W_S^*(p,N)$			$RE(\%)$		
	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$	$P = 0.2$	$P = 0.5$	$P = 0.8$
λ	Case 1: $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
0.1	40.622	39.179	37.630	40.575	39.134	37.589	0.117	0.114	0.109
0.2	21.242	20.520	19.746	21.230	20.512	19.742	0.053	0.040	0.024
0.4	12.035	11.674	11.288	12.092	11.734	11.351	0.475	0.517	0.564
0.6	10.377	10.137	9.880	10.485	10.248	9.994	1.035	1.090	1.152
0.8	44.844	44.664	44.471	44.390	44.213	44.024	1.013	1.010	1.007
(μ_1,μ_2)	Case 2: $\lambda = 0.5$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.8, 1.0)	16.445	16.157	15.848	16.528	16.244	15.940	0.503	0.540	0.581
(1.0, 1.0)	12.251	11.962	11.653	12.299	12.014	11.709	0.393	0.435	0.482
(1.0, 2.0)	10.629	10.341	10.032	10.716	10.431	10.125	0.816	0.869	0.929
(1.0, 3.0)	10.296	10.008	9.699	10.378	10.092	9.787	0.794	0.845	0.904
(1.5, 3.0)	9.229	8.940	8.631	9.252	8.966	8.659	0.250	0.281	0.317
(2.0, 3.0)	8.8792	8.591	8.282	8.881	8.594	8.287	0.014	0.034	0.057
(α_1,α_2)	Case 3: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\beta_1,\beta_2) = (3.0,4.0)$, $\gamma = 3.0$, $\theta = 0.4$								
(0.05, 0.10)	10.629	10.341	10.032	10.716	10.431	10.125	0.816	0.869	0.929
(0.05, 0.20)	10.657	10.369	10.060	10.718	10.434	10.129	0.570	0.623	0.683
(0.10, 0.05)	10.704	10.415	10.106	10.719	10.436	10.133	0.145	0.199	0.260
(0.10, 0.20)	10.748	10.459	10.150	10.724	10.422	10.139	0.223	0.169	0.109
(0.20, 0.05)	10.892	10.604	10.295	10.735	10.457	10.159	1.443	1.389	1.327
(0.20, 0.10)	10.908	10.620	10.311	10.737	10.459	10.161	1.566	1.511	1.449
(β_1,β_2)	Case 4: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, $\gamma = 3.0$, $\theta = 0.4$								
(3.0, 2.0)	10.659	10.371	10.062	10.720	10.436	10.131	0.570	0.623	0.683
(3.0, 4.0)	10.629	10.341	10.032	10.716	10.431	10.125	0.816	0.869	0.929
(3.0, 6.0)	10.620	10.331	10.022	10.715	10.429	10.124	0.898	0.951	1.011
(6.0, 2.0)	10.615	10.327	10.018	10.717	10.432	10.126	0.966	1.019	1.079
(6.0, 4.0)	10.585	10.297	9.988	10.713	10.427	10.120	1.213	1.265	1.325
(6.0, 6.0)	10.576	10.287	9.978	10.713	10.426	10.119	1.295	1.347	1.407
γ	Case 5: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, β_1, $\beta_2 = (3.0,4.0)$, $\theta = 0.4$								
2.0	10.723	10.435	10.126	10.808	10.523	10.218	0.800	0.852	0.910
3.0	10.629	10.341	10.032	10.716	10.431	10.125	0.816	0.869	0.929
4.0	10.582	10.294	9.985	10.670	10.384	10.079	0.824	0.878	0.939
5.0	10.555	10.266	9.957	10.642	10.357	10.051	0.829	0.883	0.944
6.0	10.536	10.247	9.938	10.624	10.338	10.032	0.833	0.887	0.948
θ	Case 6: $\lambda = 0.5$, $(\mu_1,\mu_2) = (1.0,2.0)$, $(\alpha_1,\alpha_2) = (0.05,0.10)$, β_1, $\beta_2 = (3.0,4.0)$, $\gamma = 3.0$								
0.2	10.185	9.897	9.588	10.248	9.962	9.656	0.615	0.661	0.714
0.4	10.629	10.341	10.032	10.716	10.431	10.125	0.816	0.869	0.929
0.6	11.177	10.889	10.580	11.296	11.011	10.706	1.064	1.124	1.192
0.8	11.884	11.596	11.287	12.042	11.758	11.453	1.331	1.397	1.472
1.0	12.854	12.566	12.257	13.057	12.772	12.468	1.574	1.643	1.721

$$E[S_1]=\frac{1}{\mu_1},\quad E\left[S_1^2\right]=\frac{3}{2\mu_1^2},\quad E[S_2]=\frac{1}{\mu_2},\quad E\left[S_2^2\right]=\frac{2}{\mu_2^2},$$
$$E[R_1]=\frac{1}{\beta_1},\quad E\left[R_1^2\right]=\frac{1}{\beta_1^2},\quad E[R_2]=\frac{1}{\beta_2},\quad E\left[R_2^2\right]=\frac{5}{4\beta_2^2},$$
$$E[U]=\frac{1}{\gamma},\quad E[U^2]=\frac{4}{3\gamma^2}.$$

The relative error percentage for the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4),$E_3/1$ queue under various values p and N are shown in Table 5. One can easily see that (i) for fix p, the relative error percentage decreases when N ranges from 2 to 16 and increases when N ranges from 18 to 20; (ii) if N is from 2 to 16 and fixed it, the relative error percentage increases in p; (iii) if N is from 18 to 20 and fixed it, the relative error percentage decreases in p; (iv) the relative error percentage in Table 5 is below 4.2 %.

Numerical results of the $\langle p,N\rangle$-policy $M/(E_2,M)$, (D,E_4),$E_3/1$ queue summarized in Table 6 for the above six cases. Again, it shows that the relative error percentages are very small (0–1.8%).

7. Conclusion

In this paper, we developed some important system performance measures for the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue. An elegant approach, the maximum entropy principle, is used to derive the approximate formulae for the steady-state probability distributions of the queue length. Our numerical investigations show that it is feasible to use the probability of various server states and the expected number of customers in the system when the server is not idle. The numerical results also indicate that the relative error percentages are very small (below 4.2%). As expected, it is sufficiently accuracy to obtain the approximate estimations. Finally, based on the improved maximum entropy principle, we demonstrate that the $\langle p,N\rangle$-policy M/(G,G), (G,G), G/1 queue is really robust to the variations of service time distribution, repair time distribution and startup time distribution functions. Consequently, this improved maximum entropy method is a useful analytic tool for approximating the solution of complex queueing systems.

Appendix

$$\begin{aligned}
W_q^*(p,N) &= \sum_{n=0}^{N-1}\left(\frac{N-n-p}{\lambda}+\mu_U+nE[S]\right)P_I(0)+\left(\mu_U+NE[S]\right)P_I(n)\\
&\quad+\sum_{n=N}^{\infty}\left(nE[S]+\frac{E\left[U^2\right]}{2\mu_U}\right)P_S(n)+\sum_{n=1}^{\infty}(nE[S])P_1(n)\\
&\quad+\sum_{n=1}^{\infty}(nE[S])P_2(n)+\sum_{n=1}^{\infty}\left(nE[S]+\frac{E\left[R_1^2\right]}{2\mu_{R_1}}\right)Q_1(n)\\
&\quad+\sum_{n=1}^{\infty}\left(nE[S]+\frac{E\left[R_2^2\right]}{2\mu_{R_2}}\right)Q_2(n)\\
&=\sum_{n=0}^{N-1}\left(\frac{N-n-p}{\lambda}+\mu_U+nE[S]\right)\frac{1-\rho_H}{N+1-p+\rho_U}\\
&\quad+\left(\mu_U+NE[S]\right)\frac{(1-p)(1-\rho_H)}{N+1-p+\rho_U}+\sum_{n=N}^{\infty}(nE[S])P_S(n)\\
&\quad+\sum_{n=N}^{\infty}\frac{E\left[U^2\right]}{2\mu_U}P_S(n)+E[S]\sum_{n=1}^{\infty}(nP_1(n)+nP_2(n)+nQ_1(n)\\
&\quad+nQ_2(n))+\frac{E\left[R_1^2\right]}{2\mu_{R_1}}\sum_{n=1}^{\infty}Q_1(n)+\frac{E\left[R_2^2\right]}{2\mu_{R_2}}\sum_{n=1}^{\infty}Q_2(n)\\
&=\frac{1-\rho_H}{\lambda(N+1-p+\rho_U)}\left[N(N-p)-\frac{N(N-1)}{2}\right.\\
&\quad\left.+N\rho_U+\frac{N(N-1)}{2}\rho\right]+\frac{(1-p)(1-\rho_H)\mu_U}{N+1-p+\rho_U}\\
&\quad+\frac{N(1-p)(1-\rho_H)E[S]}{N+1-p+\rho_U}+\frac{\lambda(1-\rho_H)E\left[U^2\right]}{2(N+1-p+\rho_U)}\\
&\quad+E[S]\left[L_{p,N}-\frac{N(N+1-2p)(1-\rho_H)}{2(N+1-p+\rho_U)}\right]\\
&\quad+\frac{E\left[R_1^2\right]}{2\mu_{R_1}}\sum_{n=1}^{\infty}Q_1(n)+\frac{E\left[R_2^2\right]}{2\mu_{R_2}}\sum_{n=1}^{\infty}Q_2(n)\\
&=\frac{N(N+1-2p)(1-\rho_H)}{2\lambda(N+1-p+\rho_U)}+\frac{N\mu_U(1-\rho_H)}{N+1-p+\rho_U}\\
&\quad+E[S]\frac{N(N+1-2p)(1-\rho_H)}{2(N+1-p+\rho_U)}+\frac{\lambda E\left[U^2\right](1-\rho_H)}{2(N+1-p+\rho_U)}\\
&\quad+\frac{(1-p)(1-\rho_H)\mu_U}{N+1-p+\rho_U}+\rho W_S(p,N)-E[S]\\
&\quad\times\frac{N(N+1-2p)(1-\rho_H)}{2(N+1-p+\rho_U)}+\frac{E\left[R_1^2\right]\rho_1\alpha_1}{2}+\frac{\theta E\left[R_2^2\right]\rho_2\alpha_2}{2}\\
&=\frac{N(N+1-2p)(1-\rho_H)}{2\lambda(N+1-p+\rho_U)}\\
&\quad+\frac{\left(2\mu_U(N+1-p)+\lambda E\left[U^2\right]\right)(1-\rho_H)}{2(N+1-p+\rho_U)}+\rho W_S(p,N)\\
&\quad+\frac{E\left[R_1^2\right]\rho_1\alpha_1}{2}+\frac{\theta E\left[R_2^2\right]\rho_2\alpha_2}{2}
\end{aligned}$$

References

Al-Jararha, J., & Madan, K. C. (2003). An M/G/1 queue with second optional service with general service time distribution. *Information and Management Sciences, 14*(2), 47–56.

Artalejo, J. R., & Lopez-Herrero, M. J. (2004). Entropy maximization and the busy period of some single-server vacation models. *RAIRO Operations Research*, 195–213.

Baker, K. R. (1973). A note on operating policies for the queue M/M/1 with exponential startup. *INFOR, 11*, 71–72.

Borthakur, A., Medhi, J., & Gohain, R. (1987). Poisson input queueing systems with startup time and under control operating policy. *Computers & Operations Research, 14*, 33–40.

El-Affendi, M. A., & Kouvatsos, D. D. (1983). A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems at equilibrium. *Acta Information, 19*, 339–355.

Feinberg, E. A., & Kim, D. J. (1996). Bicriterion optimization of an M/G/1 queue with a removable server. *Probability in the Engineering and Informational Sciences*, 57–73.

Ke, J.-C. (2003). The optimal control of an M/G/1 queueing system with server vacations, startup and breakdowns. *Computers & Industrial Engineering, 44*, 567–579.

Ke, J.-C., Ko, M. Y., & Sheu, S.-H. (2008). Estimation comparison on busy period for a controllable M/G/1 system with bicriterion policy. *Simulation Modelling Practice and Theory, 16*, 645–655.

Ke, J.-C., & Lin, C.-H. (2006). Maximum entropy solutions for batch arrival queue with an un-reliable server and delaying vacations. *Applied Mathematics and Computation, 183*, 1328–1340.

Kim, D.-J., & Moon, S.-A. (2006). Randomized control of *T*-policy for an M/G/1 system. *Computers & Industrial Engineering, 51*, 684–692.

Lee, H. W., & Park, J. O. (1997). Optimal strategy in *N*-policy production system with early set-up. *Journal of the Operational Research Society, 48*, 306–313.

Madan, K. C. (2000). An M/G/1 queue with second optional service. *Queueing Systems, 34*, 37–46.

Medhi, J. (2002). A single server Poisson input queue with a second optional channel. *Queueing Systems, 42*, 239–242.

Medhi, J., & Templeton, J. G. C. (1992). A Poisson input queue under *N*-policy and with a general start up time. *Computers & Operations Research, 19*, 35–41.

Takagi, H. (1993). A M/G/1/K queues with *N*-policy and setup times. *Queueing Systems, 14*, 79–98.

Wang, K.-H. (2003). Optimal control of a removable and non-reliable server in an M/M/1 queueing system with exponential startup time. *Mathematical Methods of Operations Research, 58*, 29–39.

Wang, J. (2004). An M/G/1 queue with second optional service and server breakdowns. *Computers and Mathematics with Applications, 47*, 1713–1723.

Wang, K.-H., & 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*, 195–212.

Wang, K.-H., Wang, T.-Y., & Pearn, W. L. (2005). Maximum entropy analysis to the *N* policy M/G/1 queueing system with server breakdowns and general startup times. *Applied Mathematics and Computation, 165*, 45–61.

Wang, K.-H., Wang, T.-Y., & Pearn, W. L. (2007). Optimal control of the *N* policy M/G/1 queueing system with server breakdowns and general startup times. *Applied Mathematical Modelling, 31*, 2199–2212.

Wang, J., & Zhao, Q. (2007). A discrete-time Geo/G/1 retrial queue with starting failures and second optional service. *Computers and Mathematics with Applications, 53*, 115–127.

Yadin, M., & Naor, P. (1963). Queueing systems with a removable service station. *Operational Research Quarterly, 14*, 393–405.