ANALYTICAL METHOD FOR ACCURACY ANALYSIS OF THE RANDOMIZED T-POLICY QUEUE

KUO-HSIUNG WANG1, DONG-YUH YANG2,* AND W. L. PEARN3

1Department of Business Administration
Asia University
No. 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan
khwang@amath.nchu.edu.tw

2Institute of Information and Decision Sciences
National Taipei College of Business
No. 321, Sec. 1, Jinan Rd., Taipei 100, Taiwan
*Corresponding author: yangdy@webmail.ntcb.edu.tw

3Department of Industrial Engineering and Management
National Chiao Tung University
No. 1001, University Rd., Hsinchu 30050, Taiwan
wlpearn@mail.nctu.edu.tw

Received December 2010; revised April 2011

Abstract. We consider a randomized policy to control the M/G/1 queueing system with an unreliable server, second optional service and general startup times. The server is subject to breaking down according to a Poisson process, and the repair time obeys a general distribution. All arrived customers demand the first required service, and only some of the arrived customers demand the second optional service. After all the customers are served in the system, the server immediately takes a vacation and operates the (T, p)-policy. For this queueing system, we employ maximum entropy approach with several constraints to develop the probability distributions of the system size and the expected waiting time in the queue. Based on the accuracy comparison between the exact and approximate methods, we show that the maximum entropy approach is quite accurate for practical purpose, which is a useful method for solving complex queueing systems.

Keywords: Accuracy comparison, Maximum entropy, Sever breakdowns, Second optional service, Startup, (T, p)-policy

1. Introduction. Queueing theory is a helpful tool to deal with real cases (some literature can be found in Ma et al. [1] and Nakashima [2]). However, in many queueing problems, the exact steady-state solutions for the service times or repair times or startup times of the general type have not been disclosed. It is rather difficult to obtain the explicit formulae such as the steady-state probability of the number of customers. In this paper, we consider the (T, p)-policy M/G/1 queue with an unreliable server, a second optional service (here abbreviated as SOS) and general startup times. We elaborate an information theoretic technique based on the maximum entropy principle to give a feasible solution for deriving probability distributions in this queueing model. 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 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 \((T, p)\)-policy, the \((T, 1)\)-policy coincides with the \(T\)-policy introduced by Heyman [3], and the \((T, 0)\)-policy is identical to the 27-policy.

In real world applications, one encounters numerous examples of the queueing situations where all arrivals require the main service and only some may require the subsidiary service provided by the server. There is an extensive literature on the M/G/1 queue in which the server may provide a second phase of service. From the practical point of view, this system can be modeled to build a scheduling problem, where all ships arriving at a port may need unloading service but only some of them may require reloading service soon after the unloading. Madan [4] was the first to study the M/G/1 queue with \(\text{SOS}\) in which the first essential service time obeys a general distribution but second optional service time follows an exponential distribution. Also Madan [4] referred some practical applications of this model. Medhi [5] extended Madan’s model [4] that the second optional service time follows a general distribution. Using the supplementary variable technique, Wang [6] investigated the reliability behavior in an M/G/1 queue with \(\text{SOS}\) and an unreliable server. Recently, Wang and Zhao [7] examined a discrete-time Geo/G/1 retrial queue with an unreliable server and \(\text{SOS}\). The main results in Wang and Zhao [7] are to obtain explicit formulae for the stationary distribution and some performance measures of the system in steady state.

The M/G/1 queue involving the randomized control problem has been considered by Feinberg and Kim [8]. They considered either \((p, N)\)- or \((N, p)\)-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 [9] considered the system with the \((p, T)\)-policy, exploited its properties and obtained the optimal values of \(T\) and \(p\) for a constrained problem. Most recently, Ke et al. [10] utilized bootstrap methods to investigate the estimation of the expected busy period of an M/G/1 queueing system under \((p, N)\)-policy.

Queueing systems with server vacations are very practical as the server may employ the idle time for additional tasks. There are many researches in the literature which deals with vacation queueing models, such as Chen et al. [11], Doshi [12], Takagi [13] and Wei et al. [14]. For control queueing systems with server vacation and a reliable server, Heyman [3] first introduced the concept of a \(T\)-policy which is defined as follows: when the system is empty, the server deactivates and leaves for a vacation with fixed length of time \(T\) (a vacation). After a vacation period of time with length \(T\), the server returns to the system. It begins to serve if there is at least one customer in the waiting line; otherwise, the server waits for another period of time length \(T\) until at least one customer is present. Tadj [15] derived the probability generating function of queue length and system characteristics in an M/G/1 quorum queueing system under \(T\)-policy. Hur et al. [16] analyzed M/G/1 queue with Min \((N, T)\)-policy, the probability distribution of the number of customers at a steady state condition, and a cost function was constructed to find the optimal operating policy. For queueing systems with an unreliable server, Wang and Ke [17] analyzed an M/G/1 queue with server breakdowns operating under the \(N\)-policy, \(T\)-policy and Min \((N, T)\)-policy. When operating a system, the server often requires a startup time before starting the service. The server startup corresponds to the pre-work of the server before starting the service. Doshi [12] and Takagi [13], respectively, studied GI/G/1 and M/G/1 queueing systems where the server requires a startup time before providing service. Recently, Wang et al. [18] compared the operating cost of the two bicriterion policies, \((T, p)\) and \((p, N)\), for an M/G/1 queue with an unreliable server, \(\text{SOS}\) and startup.
The maximum entropy principle (MEP) has been shown to give a self-consistent method of inference for estimating an unknown but true probability distribution, which is based on information in terms of some given mean value constraints. We note that MEP is not to replace the classical queueing solutions such as embedded Markov chains and matrix-geometric techniques. The basic idea of the MEP is to provide the most probable or the least biased probability distribution complying with the available mean constraints. The maximum entropy principle aims to provide a uniquely correct method of inference for estimating an unknown probability distribution. Several authors have extensively used the MEP to analyze the different queueing models, such as Shore [19], Tadj and Hamdi [20], Lopez-Herrero [21]. Tadj and Hamdi [20] utilized MEP to obtain the approximate solutions to a quorum queuing system. Lopez-Herrero [21] applied the MEP to analyze the busy period of a single server queue with exponentially distributed repeated attempts. Based on the principle, Wang et al. [22] used the same method to study the N policy M/G/1 queue with an unreliable server and derived the approximate steady-state probability distribution of the system size. Later, the N-policy M/G(G)/1 queue, which investigated in Wang et al. [22], was extended to the N-policy M/G(G, G)/1 queue (see Wang et al. [23]). Recently, Ke et al. [24] applied the same approach to approximate the steady-state probability distributions for the NT vacation M/G/1 queueing system with server breakdowns and startup time. To the best of our knowledge, very little work has focused on the steady-state probability distributions of vacation queueing systems with combining a randomized control policy, SOS and general startup time through the MEP. This paper is motivated by the use of the MEP to estimate the queue length distribution for the \((T, p)\)-policy M/G/1 queue with server breakdowns, SOS and startup times.

The entropy serves as a measure of the uncertainty of knowledge about the answer to a well-defined equation. A Lagrangian method is applied to maximize the entropy subject to various known constraints. For the queueing model considered in this paper, MEP provides approximation of the steady-state probability distributions based on different system characteristics. The purpose of this paper is fourfold. First, we present some important system characteristics for the \((T, p)\)-policy M/G/1 queue with server breakdowns, SOS and startup times. Next, we construct the maximum entropy formalism for this queueing system. Thirdly, the maximum entropy approximate solutions are developed through the Lagrange's method. Finally, we perform an accuracy comparison between the exact results and the corresponding approximate results obtained through the MEP.

2. Basic Assumptions for the \((T, p)\)-policy M/G/1 Queue. We consider the \((T, p)\)-policy M/G/1 queue with the following assumptions. It is assumed that customers arrive according to a Poisson process with rate \(\lambda\). 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 is needed to serve all arrived customers for the first required service (here abbreviated as FRS). As soon as FRS of a customer is completed, a customer may leave the system with probability 1\(\theta\) or may opt for SOS with probability \(\theta\). The service times (denoted by \(S_i\) for FRS and \(S_2\) for SOS) are independent and identically distributed (i.i.d.) random variables obeying a general distribution function \(S_i(t) (t \geq 0)\), \(i = 1, 2\), mean service time \(\mu_i\), \(i = 1, 2\), Laplace-Stieltjes transforms (LSTs) \(\tilde{f}_{S_i}(s)\), \(i = 1, 2\), and the k-th moment \(E[S_i^k]\), \(k \geq 1, \ i = 1, 2\), where the sub-index \(i = 1\) (respectively \(i = 2\)) denotes the FRS (respectively the SOS). Further, 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 LST \( \tilde{f}_S(s) = (1 - \theta) \tilde{f}_{S_1}(s) + \theta \tilde{f}_{S_1}(s) \tilde{f}_{S_2}(s) \) with the first two moments of \( S \) are

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

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

When the server is working, it may break down at any time but is immediately repaired. We assume that a server’s breakdown time has an exponential distribution with rate \( \alpha_1 \) in the FRS channel. In the SOS channel, the server fails at an exponential rate \( \alpha_2 \). When the server fails, it is immediately repaired at a repair facility. The repair times of FRS and SOS channels are independent general distributions with distribution functions \( R_1(t) \), \( R_2(t) \), \( t \geq 0 \), mean repair times \( \mu_{R_1} \), \( \mu_{R_2} \) and the \( k \)-th moments \( E[R_1^k] \), \( E[R_2^k] \), \( k \geq 1 \), 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, it immediately returns to serve a customer until the system is empty.

The server operates the \((T, p)\)-policy when the system is empty. The server requires a startup time with random length before starting 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. Let us suggest to the usual independence assumptions between inter-arrival times, service times, breakdown times, vacation times, startup times and repair times. Conveniently, we present this queueing model as the \((T, p)\)-policy M/G(G, G)/1 queue, where 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. Existing Exact Results for the \((T, p)\)-policy M/G(G, G)/1 Queue. We will develop the maximum entropy solutions for steady-state probabilities of the \((T, p)\)-policy M/G(G, G)/1 queue. Some basic known results can be obtained from the literature. These important results facilitate the application of the maximum entropy formalism to study the \((T, p)\)-policy M/G(G, G)/1 queue. Steady-state probabilities \( P_I(n), P_S(n), P_1(n), P_2(n), Q_I(n) \) 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 server is turned off, where \( n = 0, 1, 2, ... \)

\( P_S(n) \equiv \) probability that there are \( n \) customers in the system when the server is startup, where \( n = 1, 2, 3, ... \)

\( P_1(n) \equiv \) probability that there are \( n \) customers in the system when the server is providing FRS, where \( n = 1, 2, 3, ... \)

\( P_2(n) \equiv \) probability that there are \( n \) customers in the system when the server is providing SOS, where \( n = 1, 2, 3, ... \)

\( Q_I(n) \equiv \) probability that there are \( n \) customers in the system when the server is providing FRS but found to be broken down, where \( n = 1, 2, 3, ... \)

\( Q_2(n) \equiv \) probability that there are \( n \) customers in the system when the server is providing SOS but found to be broken down, where \( n = 1, 2, 3, ... \)

According to the results of Yang et al. [25], we obtain the following seven results for the \((T, p)\)-policy M/G(G, G)/1 queue.
The first result is the probability that the server is idle given by
\[
\sum_{n=0}^{\infty} P_I(n) = \frac{T(1 - \rho_H)(2 - p)}{T(2 - p) + \mu U}.
\]
(3)

The second result is the probability that the server is startup given by
\[
\sum_{n=1}^{\infty} P_S(n) = \frac{\mu_U(1 - \rho_H)}{T(2 - p) + \mu U}.
\]
(4)

The third result is the probability that the server is busy of providing FRS given by
\[
\sum_{n=1}^{\infty} P_1(n) = \lambda \mu S_1.
\]
(5)

The fourth result is the probability that the server is busy of providing FRS given by
\[
\sum_{n=1}^{\infty} P_2(n) = \lambda \theta \mu S_2.
\]
(6)

The fifth result is the probability that the server is broken down of providing FRS given by
\[
\sum_{n=1}^{\infty} Q_1(n) = \lambda \alpha_1 \mu S_1 \mu R_1.
\]
(7)

The sixth result is the probability that the server is broken down of providing SOS given by
\[
\sum_{n=1}^{\infty} Q_2(n) = \lambda \theta \alpha_2 \mu S_2 \mu R_2.
\]
(8)

The seventh result is the expected number of customers in the system given by
\[
L_{T,p} = \frac{[\lambda T^2(2 - 3p/2) + T\lambda \mu U(2 - p) + \lambda E(U^2)/2] / T(2 - p) + \mu U + L_H}{T(2 - p) + \mu U + L_H},
\]
(9)

where \( H \) is a random variable representing the (total) completion time of a customer,
\[
L_H = \rho_H + \lambda^2 E[H^2]/2(1 - \rho_H),
\]
(10)

\[
\rho_H = \lambda E[H] = \lambda \mu S_1(1 + \alpha_1 \mu R_1) + \lambda \theta \mu S_2(1 + \alpha_2 \mu R_2),
\]
(11)

\[
E[H^2] = (1 + \alpha_1 \mu R_1)^2 E[S_1^2] + \alpha_1 \mu S_1 E[R_1^2] + 2 \theta \mu S_1 \mu S_2(1 + \alpha_1 \mu R_1)(1 + \alpha_2 \mu R_2) + \theta(1 + \alpha_2 \mu R_2)^2 E[S_2^2] + \theta \alpha_2 \mu S_2 E[R_2^2].
\]
(12)

Note that \( \rho_H \) is a traffic intensity and assumed to be less than unity.

4. Maximum Entropy Approach. Exact probability distributions of the \((T, p)\)-policy M/G(G, G)/1 queue have not been discussed. Consequently, the maximum entropy principle is helpful to estimate probability distributions for a complex queueing system. In this section, we will develop the maximum entropy formalism by using several well-known constraints. These constrains are basic known results shown in the previous section. In order to derive the approximate steady-state probabilities \( P_I(n), P_S(n), P_i(n) (i = 1, 2), Q_i(n) (i = 1, 2) \), it starts to formulate the maximum entropy model in the following.
4.1. The maximum entropy model. Following El-Affendi and Kouvatsos [26], the entropy function \( y \) of an \((T, p)\)-policy M/G(G, G)/1 queue can be formed as

\[
y = - \sum_{n=0}^{\infty} P_{1}(n) \ln P_{1}(n) - \sum_{n=1}^{\infty} P_{S}(n) \ln P_{S}(n) - \sum_{n=1}^{\infty} P_{1}(n) \ln P_{1}(n) - \sum_{n=1}^{\infty} P_{2}(n) \ln P_{2}(n) - \sum_{n=1}^{\infty} Q_{1}(n) \ln Q_{1}(n) - \sum_{n=1}^{\infty} Q_{2}(n) \ln Q_{2}(n).
\] (13)

The maximum entropy solutions for the \((T, p)\)-policy M/G(G, G)/1 queue are obtained by maximizing Equation (13) subject to the following seven constraints:

1. normalizing condition:

\[
\sum_{n=0}^{\infty} P_{1}(n) + \sum_{n=1}^{\infty} P_{S}(n) + \sum_{n=1}^{\infty} P_{1}(n) + \sum_{n=1}^{\infty} P_{2}(n) + \sum_{n=1}^{\infty} Q_{1}(n) + \sum_{n=1}^{\infty} Q_{2}(n) = 1; \tag{14}
\]

2. the probability that the server is startup:

\[
\sum_{n=1}^{\infty} P_{S}(n) = \frac{\mu U (1 - \rho_{H})}{T (2 - p)} + \mu U; \tag{15}
\]

3. the probability that the server is busy of providing FRS:

\[
\sum_{n=1}^{\infty} P_{1}(n) = \lambda \mu_{S_{1}}; \tag{16}
\]

4. the probability that the server is busy of providing SOS:

\[
\sum_{n=1}^{\infty} P_{2}(n) = \lambda \theta \mu_{S_{2}}; \tag{17}
\]

5. the probability that the server is broken down when FRS is provided:

\[
\sum_{n=1}^{\infty} Q_{1}(n) = \lambda \mu_{S_{1}} \alpha_{1} \mu_{R_{1}}; \tag{18}
\]

6. the probability that the server is broken down when SOS is provided:

\[
\sum_{n=1}^{\infty} Q_{2}(n) = \lambda \theta \mu_{S_{2}} \alpha_{2} \mu_{R_{2}}; \tag{19}
\]

7. the expected number of customers in the system:

\[
\sum_{n=0}^{\infty} nP_{1}(n) + \sum_{n=1}^{\infty} nP_{S}(n) + \sum_{n=1}^{\infty} n[P_{1}(n) + P_{2}(n) + Q_{1}(n) + Q_{2}(n)] = L_{T, p}. \tag{20}
\]

In Equations (14)-(20), Equation (14) is multiplied by \( \xi_{1} \), Equation (15) is multiplied by \( \xi_{2} \), Equation (16) is multiplied by \( \xi_{3} \), Equation (17) is multiplied by \( \xi_{4} \), Equation (18) is multiplied by \( \xi_{5} \), Equation (19) is multiplied by \( \xi_{6} \), Equation (20) is multiplied by \( \xi_{7} \).
Thus, the Lagrangian function $\Psi$ is given by
\[
\Psi = -\sum_{n=0}^{\infty} P_I(n) \ln P_I(n) - \sum_{n=1}^{\infty} P_S(n) \ln P_S(n) - \sum_{n=1}^{\infty} P_I(n) \ln P_I(n) - \sum_{n=1}^{\infty} P_2(n) \ln P_2(n)
- \sum_{n=1}^{\infty} Q_1(n) \ln Q_1(n) - \sum_{n=1}^{\infty} Q_2(n) \ln Q_2(n) - \xi_1 \left[ \sum_{n=0}^{\infty} P_I(n) + \sum_{n=1}^{\infty} P_S(n) + \sum_{n=1}^{\infty} P_1(n) \right]
+ \sum_{n=1}^{\infty} P_2(n) + \sum_{n=1}^{\infty} Q_1(n) + \sum_{n=1}^{\infty} Q_2(n) - 1 \right] - \xi_2 \left[ \sum_{n=1}^{\infty} P_S(n) - \mu_U (1 - \rho_H) / T(2 - p) \right]
+ \mu_U \right] - \xi_3 \left[ \sum_{n=1}^{\infty} P_1(n) - \lambda \mu_{S_1} \right] - \xi_4 \left[ \sum_{n=1}^{\infty} P_2(n) - \lambda \mu_{S_2} \right] - \xi_5 \left[ \sum_{n=1}^{\infty} Q_1(n) \right]
- \lambda \mu_{S_1} \alpha_{1 \mu_{R_1}} \right] - \xi_6 \left[ \sum_{n=1}^{\infty} Q_2(n) - \lambda \mu_{S_2} \alpha_{2 \mu_{R_2}} \right] - \xi_7 \left[ \sum_{n=0}^{\infty} n P_1(n) + \sum_{n=1}^{\infty} n P_S(n) \right]
+ \sum_{n=1}^{\infty} n (P_1(n) + P_2(n) + Q_1(n) + Q_2(n)) - L_{T,p} \right],
\]
where $\xi_1 - \xi_7$ are the Lagrangian multipliers corresponding to constrains (14)-(20), respectively.

4.2. The maximum entropy solutions. To find the maximum entropy solutions $P_I(n)$, $P_S(n)$, $P_i(n)$ ($i = 1, 2$) and $Q_i(n)$ ($i = 1, 2$), maximizing in Equation (13) subject to constrains (14)-(20) is equivalent to maximizing (21). The maximum entropy solutions are obtained by taking the partial derivatives of $y$ with respect to $P_I(n)$, $P_S(n)$, $P_i(n)$ ($i = 1, 2$), $Q_i(n)$ ($i = 1, 2$) and setting the results equal to zero, namely,
\[
\frac{\partial \Psi}{\partial P_I(n)} = -1 - \ln P_I(n) - \xi_1 - n \xi_7 = 0, \quad n \geq 0,
\]
\[
\frac{\partial \Psi}{\partial P_S(n)} = -1 - \ln P_S(n) - \xi_1 - \xi_2 - n \xi_7 = 0, \quad n \geq 1,
\]
\[
\frac{\partial \Psi}{\partial P_i(n)} = -1 - \ln P_i(n) - \xi_1 - \xi_3 - n \xi_7 = 0, \quad n \geq 1,
\]
\[
\frac{\partial \Psi}{\partial P_2(n)} = -1 - \ln P_2(n) - \xi_1 - \xi_4 - n \xi_7 = 0, \quad n \geq 1,
\]
\[
\frac{\partial \Psi}{\partial Q_1(n)} = -1 - \ln Q_1(n) - \xi_1 - \xi_5 - n \xi_7 = 0, \quad n \geq 1,
\]
\[
\frac{\partial \Psi}{\partial Q_2(n)} = -1 - \ln Q_2(n) - \xi_1 - \xi_6 - n \xi_7 = 0, \quad n \geq 1.
\]
From Equations (22)-(27), it gives that
\[
P_I(n) = e^{-\left(1 + \xi_1 + n \xi_7 \right)}, \quad n \geq 0,
\]
\[
P_S(n) = e^{-\left(1 + \xi_1 + \xi_2 + n \xi_7 \right)}, \quad n \geq 1,
\]
\[
P_i(n) = e^{-\left(1 + \xi_1 + \xi_3 + n \xi_7 \right)}, \quad n \geq 1,
\]
\[
P_2(n) = e^{-\left(1 + \xi_1 + \xi_4 + n \xi_7 \right)}, \quad n \geq 1,
\]
\[
Q_i(n) = e^{-\left(1 + \xi_1 + \xi_5 + n \xi_7 \right)}, \quad n \geq 1,
\]
\[
Q_2(n) = e^{-\left(1 + \xi_1 + \xi_6 + n \xi_7 \right)}, \quad n \geq 1.
\]
Let $\nu_1 = e^{-(1+\xi_1)}$ and $\nu_j = e^{-\xi_j}$ for $2 \leq j \leq 7$. We transform Equations (28)-(33) in terms of $\nu_j$ ($1 \leq j \leq 7$) given by

\begin{align*}
P_1(n) &= \nu_1 \nu_1^n, \quad n \geq 0, \tag{34} \\
P_S(n) &= \nu_1 \nu_2 \nu_1^n, \quad n \geq 1, \tag{35} \\
P_1(n) &= \nu_1 \nu_3 \nu_1^n, \quad n \geq 1, \tag{36} \\
P_2(n) &= \nu_1 \nu_4 \nu_1^n, \quad n \geq 1, \tag{37} \\
Q_1(n) &= \nu_1 \nu_5 \nu_1^n, \quad n \geq 1, \tag{38} \\
Q_2(n) &= \nu_1 \nu_6 \nu_1^n, \quad n \geq 1. \tag{39}
\end{align*}

Substituting Equations (35)-(39) into Equations (15)-(19), respectively, yields

\begin{align*}
\nu_1 \nu_2 &= \frac{\mu_U (1 - \rho_H)(1 - \nu_2)}{[T(2 - p) + \mu_U] \nu_1}, \tag{40} \\
\nu_1 \nu_3 &= \frac{\lambda \mu_S_1 (1 - \nu_3)}{\nu_1}, \tag{41} \\
\nu_1 \nu_4 &= \frac{\lambda \theta \mu_S_2 (1 - \nu_4)}{\nu_1}, \tag{42} \\
\nu_1 \nu_5 &= \frac{\lambda \mu_S_2 \alpha_1 \mu_{R_1} (1 - \nu_5)}{\nu_1}, \tag{43} \\
\nu_1 \nu_6 &= \frac{\lambda \theta \mu_S_2 \alpha_2 \mu_{R_2} (1 - \nu_6)}{\nu_1}. \tag{44}
\end{align*}

Inserting Equations (15)-(19), Equation (34) in Equation (14) and execute some algebraic manipulations, we may obtain $\nu_1$ given by

\begin{equation}
\nu_1 = \frac{T(2 - p)(1 - \rho_H)(1 - \nu_2)}{T(2 - p) + \mu_U}. \tag{45}
\end{equation}

Substituting (34)-(39) into (20), it yields

\begin{equation}
\nu_1 = \frac{[L_{T, p} - \tau (1 - \rho_H) - \rho_H] \cdot [T(2 - p) + \mu_U]}{T(2 - p)(1 - \rho_H) + L_{T, p}[T(2 - p) + \mu_U]}, \tag{46}
\end{equation}

where $\tau = \mu_U/[T(2 - p) + \mu_U]$.

From Equations (34)-(45), we finally obtain

\begin{align*}
P_1(n) &= \frac{T(2 - p)(1 - \rho_H)(1 - \nu_2) \nu_1^n}{T(2 - p) + \mu_U}, \quad n \geq 0, \tag{47} \\
P_S(n) &= \frac{\mu_U (1 - \rho_H)(1 - \nu_2) \nu_1^n}{[T(2 - p) + \mu_U]} = \tau (1 - \rho_H)(1 - \nu_2) \nu_1^n, \quad n \geq 1, \tag{48} \\
P_1(n) &= \lambda \mu_S_1 (1 - \nu_2) \nu_1^n, \quad n \geq 1, \tag{49} \\
P_2(n) &= \lambda \theta \mu_S_2 (1 - \nu_2) \nu_1^n, \quad n \geq 1, \tag{50} \\
Q_1(n) &= \lambda \mu_S_2 \alpha_1 \mu_{R_1} (1 - \nu_2) \nu_1^n, \quad n \geq 1, \tag{51} \\
Q_2(n) &= \lambda \theta \mu_S_2 \alpha_2 \mu_{R_2} (1 - \nu_2) \nu_1^n, \quad n \geq 1. \tag{52}
\end{align*}

5. **Expected Waiting Time in the Queue.** In this section, we derive the exact and the approximate formulae of the expected waiting time in the queue for the $(T, p)$-policy $M/G (G, G)/1$ queue.
5.1. The exact expected waiting time in the queue. Let $EW_T$, $EW_{2T}$ and $EW_{TP}$ denote the exact expected waiting time in the queue for the $T$, $2T$- and $(T, p)$-policies, respectively. Using the results of Yang et al. [25] and Equation (10) in Little’s formula, we have

$$EW_T = \frac{L_T}{\lambda} - E[H] = \frac{1}{(T + \mu_U)} \left[ \frac{T^2}{2} + \mu_U T + \frac{E(U^2)}{2} \right] + \frac{\lambda E(H^2)}{2(1 - \rho_H)^2},$$  \quad (53)$$

$$EW_{2T} = \frac{L_{2T}}{\lambda} - E[H] = \frac{1}{(2T + \mu_U)} \left[ 2T^2 + 2T \mu_U + \frac{E(U^2)}{2} \right] + \frac{\lambda E(H^2)}{2(1 - \rho_H)^2},$$  \quad (54)$$

$$EW_{TP} = \frac{L_{TP}}{\lambda} - E[H] = \frac{1}{T(2 - p) + \mu_U} \left[ T^2(2 - \frac{3}{2}p) + T \mu_U (2 - p) + \frac{E(U^2)}{2} \right] + \frac{\lambda E(H^2)}{2(1 - \rho_H)^2}. \quad (55)$$

After some algebraic manipulation, we know that $EW_{TP}$ is a convex combination of $EW_T$ and $EW_{2T}$ which represented as follows

$$EW_{TP} = \frac{p(T + \mu_U)}{(2 - p)T + \mu_U} EW_T + \left[ 1 - \frac{p(T + \mu_U)}{(2 - p)T + \mu_U} \right] EW_{2T}. \quad (56)$$

Substituting Equations (53)-(54) into Equation (56), the result is equal to Equation (55). Thus, we demonstrate that the relationships given by Equations (55)-(56) are seen to hold.

5.2. The approximate expected waiting time in the queue. The idle state, the startup state, the busy state, and the repair state are defined as follows:

1. Idle state denoted by $I$: the server is turned off, and the number of customers waiting in the system is greater than or equal to 0.
2. Startup state denoted by $U$: the server begins startup, and the number of customers waiting in the system is greater than or equal to 1.
3. Busy state when $FRS$ is provided denoted by $B_1$: the server is busy and provides $FRS$ to a customer.
4. Busy state when $SOS$ is provided denoted by $B_2$: the server is busy and provides $SOS$ to a customer.
5. Repair state when $FRS$ is provided denoted by $R_1$: the server is broken down when $FRS$ is provided and being repaired.
6. 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 $I$, $U$, $B_1$, $B_2$, $R_1$ 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 $I$, $U$, $B_1$, $B_2$, $R_1$ and $R_2$ are described, respectively, as follows:

1. In idle state $I$: 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. Applying the results of Borthakur et al. [27], the mean remaining residual vacation time is $p(E[T^2]/2E[T]) + (1 - p)(E[(2T)^2]/2E[2T]) = T - pT/2$. Hence, customer $C$ must wait (i) the mean residual vacation time, (ii) the service time of $n$ customers in the queue and (iii) the startup time before providing $FRS$. From the inferences of (i)-(iii), the expected waiting time of customer $C$ at the idle state $I$ is $(T - pT/2 + nE[S] + \mu_U)$.

2. In startup state $U$: the expected waiting time of customer $C$ at the startup state can be derived in the following. Using the same arguments as Borthakur et al. [27] again,
it indicates that the expected waiting time of customer \( C \) at the startup state \( U \) is 
\[
E[S] + E[U^2]/2\mu_U.
\]

(3) 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. Furthermore, the 
server is subject to breakdowns occurring at any time when the server operates. The 
expected waiting time at the busy state \( B_1 \) and \( B_2 \) are discussed in the following.

(i) In busy state \( B_1 \): there are \( n \) customers in front of customer \( C \) will be served for 
FRS. The server is unpredictable breakdowns with rate \( \alpha_1 \). Consequently, the 
expected waiting time of customer \( C \) at the busy state \( B_1 \) is 
\[
E[S] + \alpha_1 \mu_{R_1}.
\]

(ii) In busy state \( B_2 \): this state means that one of \( n \) customers in front of customer 
\( C \) has finished FRS and ready for SOS. The remaining \( (n-1) \) customers are still 
waiting for service. Again, the server also could be breakdown with rate \( \alpha_2 \). It 
implies that the expected waiting time of an arbitrary customer \( C \) at the busy 
state \( B_2 \) is 
\[
E[S] + \mu_{S_2} + \alpha_2 \mu_{R_2}.
\]

(4) 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 
\[
E[S] + E[R_1^2]/2\mu_{R_1} \quad \text{and} \quad (n-1)E[S] + \mu_{S_2} + E[R_2^2]/2\mu_{R_2},
\]
respectively.

Utilizing the listed above results yields the following approximate expected waiting time 
in the queue.

\[
AW_{TP} = \sum_{n=0}^{\infty} (T - pT/2 + nE[S] + \mu_U) P_1(n) + \sum_{n=1}^{\infty} (nE[S] + E[U^2]/2\mu_U) P_S(n)
\]
\[
+ \sum_{n=1}^{\infty} (nE[S] + \alpha_1 \mu_{R_1}) P_1(n) + \sum_{n=1}^{\infty} ((n-1)E[S] + \mu_{S_2} + \alpha_2 \mu_{R_2}) P_2(n)
\]
\[
+ \sum_{n=1}^{\infty} (nE[S] + E[R_1^2]/2\mu_{R_1}) Q_1(n) + \sum_{n=1}^{\infty} ((n-1)E[S] + \mu_{S_2} + E[R_2^2]/2\mu_{R_2}) Q_2(n),
\]

where \( P_1(n), P_S(n), P_1(n), P_2(n), Q_1(n) \) and \( Q_2(n) \) are given in Equations (47)-(52),
respectively.

6. Accuracy Analysis of the Entropy Solutions. This section aims to carry out 
comparisons to examine the accuracy for the maximum entropy results. We perform an 
accuracy comparison between the exact and the maximum entropy solutions, based 
on the mean waiting time in the queue. Numerical comparisons are based on the following 
assumptions:

- service time of FRS channel is a 3-stage Erlang distribution with mean \( \mu_{S_1} = 1/\mu_1 \) 
  and second moment \( E[S_1^2] = 4/(3\mu_1^2) \);
- service time of SOS channel is an exponential distribution with mean \( \mu_{S_2} = 1/\mu_2 \) 
  and second moment \( E[S_2^2] = 2/\mu_2^2 \);
- repair time of FRS channel is a 4-stage Erlang distribution with mean \( \mu_{R_1} = 1/\beta_1 \) 
  and second moment \( E[R_1^2] = 5/(4\beta_1^2) \);
- repair time of SOS channel is a 2-stage Erlang distribution with mean \( \mu_{R_2} = 1/\beta_2 \) 
  and second moment \( E[R_2^2] = 3/(2\beta_2^2) \);
- startup time is a deterministic distribution with mean \( 1/\gamma \) and second moment 
  \( E[U^2] = 1/\gamma^2 \).

First, we fix \( \lambda = 2.0, \mu_1 = 6.0, \mu_2 = 4.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \beta_1 = 8.0, \beta_2 = 5.0, \theta = 0.5 \), and choose various values of \( (T, p) \). Numerical results are summarized 
in Table 1. The relative error percentage, denoted by \( REP \), is used to measure the
accuracy of the approximate values in the following:

\[ REP = \frac{|EW_{TP} - AW_{TP}|}{EW_{TP}} \times 100\% \]

One observes from Table 1 that the approximations are good because the REPs are very small (0-4.6%). Next, choosing \( T = 8 \) and different values \( p = 0.2, 0.5, 0.8 \). The values of \( \lambda, \mu_1, \mu_2, \gamma, \alpha_1, \alpha_2, \beta_1, \beta_2 \) and \( \theta \) are considered in the following six cases:

Case 1. \( \mu_1 = 6.0, \mu_2 = 4.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \beta_1 = 8.0, \beta_2 = 5.0, \theta = 0.5 \), and varying the values of \( \lambda \).

Case 2. \( \lambda = 2.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \beta_1 = 8.0, \beta_2 = 5.0, \theta = 0.5 \), and varying the values of \( (\mu_1, \mu_2) \).

Case 3. \( \lambda = 2.0, \mu_1 = 6.0, \mu_2 = 4.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \theta = 0.5 \), and varying the values of \( (\alpha_1, \alpha_2) \).

Case 4. \( \lambda = 2.0, \mu_1 = 6.0, \mu_2 = 4.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \theta = 0.5 \), and varying the values of \( (\beta_1, \beta_2) \).

Case 5. \( \lambda = 2.0, \mu_1 = 6.0, \mu_2 = 4.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \beta_1 = 8.0, \beta_2 = 5.0, \theta = 0.5 \), and varying the values of \( \gamma \).

Case 6. \( \lambda = 2.0, \mu_1 = 6.0, \mu_2 = 4.0, \gamma = 3.0, \alpha_1 = 0.05, \alpha_2 = 0.1, \beta_1 = 8.0, \beta_2 = 5.0, \theta = 0.5 \), and varying the values of \( \theta \).

<table>
<thead>
<tr>
<th>( T )</th>
<th>0.01</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.771</td>
<td>0.304</td>
<td>0.553</td>
<td>1.048</td>
<td>0.947</td>
<td>0.243</td>
<td>1.409</td>
</tr>
<tr>
<td>2</td>
<td>0.145</td>
<td>0.484</td>
<td>1.680</td>
<td>2.460</td>
<td>2.502</td>
<td>1.047</td>
<td>0.584</td>
</tr>
<tr>
<td>3</td>
<td>0.126</td>
<td>0.828</td>
<td>2.182</td>
<td>3.102</td>
<td>3.232</td>
<td>1.684</td>
<td>0.166</td>
</tr>
<tr>
<td>4</td>
<td>0.277</td>
<td>1.021</td>
<td>2.466</td>
<td>3.468</td>
<td>3.655</td>
<td>2.062</td>
<td>0.086</td>
</tr>
<tr>
<td>5</td>
<td>0.374</td>
<td>1.144</td>
<td>2.648</td>
<td>3.705</td>
<td>3.930</td>
<td>2.313</td>
<td>0.254</td>
</tr>
<tr>
<td>6</td>
<td>0.441</td>
<td>1.230</td>
<td>2.775</td>
<td>3.871</td>
<td>4.124</td>
<td>2.491</td>
<td>0.375</td>
</tr>
<tr>
<td>7</td>
<td>0.490</td>
<td>1.293</td>
<td>2.868</td>
<td>3.993</td>
<td>4.268</td>
<td>2.624</td>
<td>0.465</td>
</tr>
<tr>
<td>8</td>
<td>0.527</td>
<td>1.341</td>
<td>2.940</td>
<td>4.087</td>
<td>4.379</td>
<td>2.727</td>
<td>0.536</td>
</tr>
<tr>
<td>9</td>
<td>0.557</td>
<td>1.379</td>
<td>2.996</td>
<td>4.161</td>
<td>4.467</td>
<td>2.810</td>
<td>0.592</td>
</tr>
<tr>
<td>10</td>
<td>0.582</td>
<td>1.411</td>
<td>3.042</td>
<td>4.222</td>
<td>4.538</td>
<td>2.877</td>
<td>0.638</td>
</tr>
</tbody>
</table>

Numerical results of the \((T, p)\)-policy queue with server breakdowns, \(SOS\) and startup are presented in Table 2 for the above six cases. From Table 2, it appears that (i) \( REP \) decreases as \( \lambda \) increases; (ii) \( REP \) increases as \( \mu_1 \) or \( \mu_2 \) increases; (iii) \( REP \) increases as \( \alpha_1 \) or \( \alpha_2 \) increases; (iv) \( REP \) decreases as \( \beta_1 \) or \( \beta_2 \) increases; (v) \( REP \) increases in \( \gamma \); (vi) \( REP \) decreases as \( \theta \) increases; (vii) \( \lambda \) and \( \mu_1 \) affect \( REP \) sensitively, but \( REP \) is rarely effected by \( \gamma \). A close look at Table 2, we conclude that the accuracy of approximate is quite good because the \( REP \) is remarkably small (below 6.3%). The numerical results also indicate that the entropy solution is sufficient accurate for practical purpose. In this light, the maximum entropy method can be regarded as a helpful method for analyzing complex queueing system.
Table 2. Comparison of the exact $EW_{TP}$ and the approximate $AW_{TP}$ for the $(T, p)$-policy queue with server breakdowns, $SOS$ and startup

<table>
<thead>
<tr>
<th>$p$</th>
<th>$EW_{TP}$</th>
<th>$AW_{TP}$</th>
<th>$REP(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>Case 1: $\mu_1 = 6.0$, $\mu_2 = 4.0$, $\gamma = 3.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\beta_1 = 8.0$, $\beta_2 = 5.0$, $\theta = 0.5$</td>
<td>1.2</td>
<td>7.850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4</td>
<td>7.889</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6</td>
<td>7.936</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.8</td>
<td>7.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>8.071</td>
</tr>
<tr>
<td></td>
<td>Case 2: $\lambda = 2.0$, $\gamma = 3.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\beta_1 = 8.0$, $\beta_2 = 5.0$, $\theta = 0.5$</td>
<td>(0.20, 0.05)</td>
<td>3.0, 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case 3: $\lambda = 2.0$, $\mu_1 = 6.0$, $\mu_2 = 4.0$, $\gamma = 3.0$, $\beta_1 = 8.0$, $\beta_2 = 5.0$, $\theta = 0.5$</td>
<td>(0.05, 0.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case 4: $\lambda = 2.0$, $\mu_1 = 6.0$, $\mu_2 = 4.0$, $\gamma = 3.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\theta = 0.5$</td>
<td>(0.10, 0.10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case 5: $\lambda = 2.0$, $\mu_1 = 6.0$, $\mu_2 = 4.0$, $\gamma = 3.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\beta_1 = 8.0$, $\beta_2 = 5.0$, $\theta = 0.5$</td>
<td>(0.15, 0.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case 6: $\lambda = 2.0$, $\mu_1 = 6.0$, $\mu_2 = 4.0$, $\gamma = 3.0$, $\alpha_1 = 0.05$, $\alpha_2 = 0.1$, $\beta_1 = 8.0$, $\beta_2 = 5.0$, $\theta = 0.5$</td>
<td>(0.20, 0.20)</td>
</tr>
</tbody>
</table>

7. Conclusions. In this paper, we investigated the $(T, p)$-policy $M/G(G, G)/1$ queue for the steady-state probability distribution and the expected waiting time in the queue. Based on the MPE, an analytical approach was applied to compute the steady-state probabilities of the system size. We demonstrate that the maximum entropy method is powerful and is easy to implement. Consequently, the use of the system characteristics is sufficient to obtain accurate estimations. In addition, an extensive numerical computation was performed to compare the exact analytical and the approximate expected waiting time in the queue. The numerical results showed that the relative error percentages are quite small, which leads to the conclusion that the maximum entropy method is accurate enough for practical purpose.
Acknowledgment. The authors would like to thank the National Science Council of Taiwan (Contract No. NSC 99-2218-E-141-001-) for its support.

REFERENCES

