在 *N* 方策和 *T* 方策下具故障及啟動服務者 M/G/1 排隊之研究

Analysis of an M/G/1 Queue with Sever Breakdowns and Startup Times under the *N* Policy and the *T* Policy

研 究 生：王琮胤 Student：Tsung-Yin Wang
指導教授：彭文理 博士 Advisor：Dr. W. L. Pearn

國 立 交 通 大 學
工業工程與管理學系

博士論文

A Dissertation Submitted to
Department of Industrial Engineering and Management
College of Management
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Phillosophy
in
Industrial Engineering and Management

October 2007
Hsinchu, Taiwan, Republic of China

中華民國 九十六 年 十 月

Acknowledgment

誌　謝

離開學生身分已久的我，再重拾學生身分，無論在學習或心境上，都和以往有許多的不同，在體能的負荷上更是吃力。回顧過去四年多來，由於許多貴人的幫助和鼓勵，才能使我有機會進修，也才能使我在家庭、學業及工作三方面的壓力下，順利完成學業。

本論文能夠順利完成，首先要感謝指導老師 彭文理教授熱心的指導，老師除了有卓越的專業知識外，其嚴謹的治學態度與研究精神都是學生學習的標竿；在論文寫作期間，承蒙中興大學應數系 王國雄教授給於學生諸多的意見及研究方向使本論文得以完成，此外還要感謝的本系 鍾淑馨教授、中興大學應數系 王國雄教授、臺灣科技大學工管系 徐世輝教授及臺中技術學院統計系 柯沛程教授，對本論文費心的審查及提供諸多的寶貴意見使本論文更加完備。

在進修期間，筆者服務的學校—臺中技術學院會計系，無論在課程、排課的時間上，都給予筆者最大的方便，在此感謝系上同仁的體諒與幫忙，此外要特別感 劉政淮老師對於筆者在博士修業期間時的關懷、幫忙及鼓勵。再者，要謝謝交大實驗室裏的學長、學弟及同學在課業上的砥礪與啟發，使我獲益良多。

最後要謝謝妻子淑娟這些年來全心的支持與體諒，照顧好家庭與女兒，使我無後顧之憂，能完成夢想。

王琮胤 謹誌於國立交通大學

管理學院工業工程與管理學系

中華民國九十六年十月二十九日

在 *N* 方策和 *T* 方策下具故障及啟動服務者 M/G/1 排隊之研究

學生: 王琮胤　　指導教授：彭文理 博士

國立交通大學管理學院
工業工程與管理學系

摘 要

可控制排隊系統已被廣泛的應用於各領域，諸如製造/存貨系統、通訊及電腦網路系統等等，配合成本分析可幫助決策者在舜息萬變的資訊時代做出最佳決策，追求最大的利潤與效益。本論文主要是探討在 *N* 方策和 *T* 方策下服務者會故障及需一般啟動時間的 M/G/1 排隊系統，所謂 *N* 方策是指服務者一停止服務，要再重新開始提供服務，完全取決於等候線的顧客數是否達到 *N* 人，當顧客數達到 *N* 人時，服務者馬上開始提供服務直到系統中的所有顧客服務完成才停止服務；而所謂 *T* 方策是指服務者一停止服務，在一定時間 *T* 內至少有一位顧客到達，服務者才會開始提供服務直到系統中的所有顧客服務完成才停止服務，若在間 *T* 內無顧客到達，則必需等待直到下一個時間 *T* 內至少有一位有顧客到達。我們分別針對 *N* 方策和 *T* 方策，利用 M/G/1 排隊系統的隨機分解性質求出系統中的期望顧客數及系統期望的閒置、啟動、忙碌、故障期間長度，並推導出系統閒置、啟動、忙碌、故障狀態的機率，利用上述的期望值及機率，我們建構成本函數，分別求得最佳的 *N* 方策和 *T* 方策，並針對最佳的方策做敏感度分析，提供決策者在各種不同的參數下選擇最佳的方策。

此外，鑑於在 *N* 方策和 *T* 方策下服務者會故障及需一般啟動時間的 M/G/1 排隊系統中顧客數的機率分配無法確切求出，我們利用最大熵值法，在有限有用的資訊下諸如排隊系統中的期望顧客數、閒置、啟動、忙碌、故障狀態的機率，在最少偏誤的資訊下，求出估計的系統中顧客數的機率及近似的平均等候時間，並在各種不同分配下，比較最大熵值法解得近似的平均等候時間與真正的平均等候時間兩者之間的誤差。我們驗證得到最大熵值法是一個有用且夠精確的方法，可用來解決複雜的排隊問題。

關鍵字: *N* 方策；*T* 方策；一般修理時間；一般啟動時間；M/G/1 排隊；敏感度分析；最大熵值

Analysis of an M/G/1 Queue with Sever Breakdowns and Startup Times under the *N* Policy and the *T* Policy

Student: Tsung-Yin Wang　　Advisor: Dr. W. L. Pearn

Department of Industrial Engineering and Management
College of Management, National Chiao Tung University

Abstract

The controllable queueing systems have been done by a considerable amount of work in the past and successfully used in various applied problems such as production/inventory systems, communication systems, computer networks and etc. To cooperate with the cost analysis, it can help the decision maker to make the optimal decision to obtain the maximal profit and efficiency for use. In this dissertation, we investigate an M/G/1 queue under the N policy and the T policy with sever breakdowns and general startup times. The N policy means the server returns to provide service only when the number of customers in the system reaches N $(N \geq 1)$ until there are no customers present. The T policy means the server is turned on after a fixed length of time T repeatedly until at least one customer is present in the waiting line. Using the stochastic decomposition property of the M/G/1 queue, we derive various system performance measures, such as the expected number of customers, the expected length of the turned-off, startup, busy, and breakdown periods under the N policy and T policy, respectively. Then we deduce the probabilities of turned-off, startup, busy, and breakdown periods. We also construct the total expected cost function per unit time to determine the optimal threshold N and T, respectively, in order to minimize the cost function for the both policies. Sensitivity investigations on the optimal value of N and T for the both policies, respectively, are studied. Some numerical investigations are presented to demonstrate the analytical results obtained, and show how to make the decision based on minimizing the cost function.

In addition, it is extremely difficult, not impossible, to obtain the explicit formulas such as the steady-state probability mass function of the number of customers and the expected waiting time for the N policy and the T policy M/G/1 queues with repair times and startup times are generally distributed. Under the given available information such as the queue length and the probabilities of idle, startup, busy and breakdown period, we use the maximum entropy principle to derive the approximate formulas for the steady-state probability distributions. We perform a comparative analysis between the approximate waiting time with established waiting time for various distributions, such as exponential (M), k-stage Erlang (E_k), and deterministic (D). We demonstrate that the maximum entropy approach is accurate enough for practical purposes and is a useful method for solving complex queueing systems.

Keywords: *N* policy, *T* policy, general repair time, general startup time, M/G/1 queue, sensitivity analysis, maximum entropy.

List of Contents

List of Table

List of Figures

Chapter 1

Introduction

1.1 Background

Since 1909, A. k. Erlang first brought up the probability knowledge to develop the mathematic model to model the telephone traffic problem. Due to Erlang's concept, many stochastic problems not only can be constructed the mathematic model but also can be done useful engineering computation. Accomplishing the development of telephone system, queueing theory with application has been growing quickly, which it already has been widely used to solve certain practical problems. In addition, the modeling and analysis for the controllable queueing system have been done by a considerable amount of work in the past and successfully used in various applied problems such as production/inventory systems, communication systems, computer networks and etc. (see survey paper by Doshi [11])

A comprehensive and excellent study on the controllable queueing system can be found in Tadj and Choudhury [42] and Takagi [43]. The controllable queueing system tries to find the operating policy, that is, rules for turning the server on and off that result in the minimum long-run cost, in which the optimization is carried out over a class of operating policies rather than over a set of parameters for a single operating policy that is fixed as part of the model. As for such controllable queues, we summary five general categories from Teghem [45] and Gupta [17] as follows:

1. ***Control of the number of servers.*** Servers are removable and may be turned on or off according to the state of the system. The varying number of active servers must be determined.

2. ***Control of the service rate.*** This category generalizes control of the number of servers. We change the service process by varying the service rate rather than modifying the number of servers.

3. ***Control of the arrival rate.*** This category is to control the arrival process. If the arrivals reach a certain level, the arriving customer will be blocked.

4. ***Control of the queue discipline.*** The order of service is determined among different classes of customers, and an allocation of customers to servers is made.

5. ***Control of the admission of customers.*** The arrival rate can be modified, customers may be denied entry, or customers control the decision for entry.

The optimal control problem of a queueing system is to determine when and how to change system parameters to optimize some objective function (minimize the expected system cost function or maximize the expected system profit function). We will find the optimal operating policy to turn the server on and off that result in the lowest long-run cost. The number of published papers concerning the controllable queueing systems is numerous. Among the control policies, the threshold control policies can be divided into the following control policies:

1. **The *N* policy.** An N policy is defined as one where the server is turned on when following the start of the idle period, the total number of customers in the queue reaches the value N and turned off when the system becomes empty. The N policy was first introduced by Yadin and Naor [58] in 1963. Many researchers have worked on this subject since Yadin and Naor [58]; (e.g., Medhi and Templeton [34], Takagi [44], Lee and Park [32], Krishna et al. [31], Hur and Paik [21], Wang et al. [50], Wang and Ke [53], and others.)

2. **The *T* policy.** Following the beginning of the idle period, the server is turned on after a fixed length of time T repeatedly until at least one customer is present in the waiting line. The T policy is developed by Heyman [20], Levy and Yechiali [33], Tijms [46], and Gakis et al. [15] and so on.

3. **The *D* policy.** The D policy takes into account the service times of the waiting customers. After the beginning of the idle period, the server returns to provide service if the sum of service time of the new arriving customers reaches or exceeds a given quantity D ($D \geq 0$) for the first time. The D policy is introduced by Balachandran [6], Tijms [46] and Gakis et al. [15]

4. **The *F* policy.** Consider a queueing system of finite capacity. When the number of customers in the system reaches its capacity K (i.e. the system becomes full), no further arriving customers are allowed to enter the system until there are enough customers in the system have been served so that the number of customers in the system decreases to a threshold value F ($0 \leq F \leq K-1$). The concept of F policy is introduced by Gupta [17]

5. **Combined threshold policies.** Policies are combined by any two of the N, T, and D policies. Doganata [10] first considered the NT-vacation policy M/G/1 queueing system and determined numerically the optimal pair (N,T) that maximizes the length of the vacation period. Gakis et al. [15] proposed six dyadic

policies for an M/G/1 queue. Wang and Ke [54] analyzed an M/G/1 queue with server breakdowns operating under the N policy, the T policy, and the $Min(N,T)$ policy. Hur et al. [22] consider an M/G/1 queue with NT-vacation policy. For a given T, the optimal N-value that minimizes the long-run total expected average cost per unit of time is obtained.

Among the threshold control policies, we will focus on the N policy and T policy. Because the N policy is analytically much easier than other policies, many researchers concentrated on the N policy. To employ the N policy, the server must be continuously monitoring the queue for an arrival when the server is turned off. Adopting the N policy is efficient in utilizing the system facilities and reducing the customer's waiting time. However, in practice, the continue monitor may not be executed in some situations or sometimes the continue monitor results in an expensive cost, the T policy will be proposed. We will analyze various system performance measures and develop the total expected cost function per unit time in which the threshold T or N is a decision variable. We will determine the optimum threshold value and derive analytical results for sensitivity investigations. We also use the maximum entropy principle to obtain the approximate results for the expected waiting time in queue and then perform comparative analysis between the approximate results with established exact results for various distribution functions.

1.2 Theoretical Analysis Techniques

In this section, we will introduce the stochastic decomposition property and the maximum entropy technique to study the N policy and T policy.

1.2.1 Stochastic decomposition property in the M/G/1 *queue*

The stochastic decomposition property in an M/G/l queue was first examined by Gaver [16], and subsequently by Copper [9], Levy and Yechiali [33], Shanthikumar [38], Scholl and Kleinrock [37], Ali and Neuts [2], and many other authors. This type of decomposition property means that the (stationary) number of customers in the system at a random point in time is distributed as the sum of two or more independent random variables, one of which is the (stationary) number of customers in the corresponding standard M/G/l queue (i.e., the server is always available) at a random point in time.

Fuhrmann and Cooper [14] demonstrated the decomposition property holds for a general class of the M/G/l queueing models. They showed that, for any vacation model in an M/G/1 queue, the system size distribution decomposes into two random variables, one of which corresponds to the system size of the ordinary queue without vacations. The interpretation of the other random variable is usually related to the system size given that the server is on vacation. That is,

$$\psi(z) = \chi(z)\pi(z), \tag{1.1}$$

where

$\psi(z) \equiv$ the probability generating function (p.g.f.) for the stationary distribution of the number of customers at a random point.

$\chi(z) \equiv$ the p.g.f. for the stationary distribution of the number of customers at a random point in time when the server on vacation.

$\pi(z) \equiv$ the p.g.f. for the stationary distribution of the number of customers at a random point in the standard M/G/1 queue.

Under the following assumptions:

Assumption 1. The service times of different customers are independent of each other and are independent of the arrival process. In addition, each service time is independent of the sequence of vacation periods that precede that service time.

Assumption 2. All customers arriving to the system are eventually served. Moreover, customers do not balk, defect, or renege from the system.

Assumption 3. Customers are served in an order that is independent of their service times.

Assumption 4. Service is non-preemptive. That is, once selected for service, a customer is served to completion in a continuous manner.

Assumption 5. The rules that govern when the server begins and ends vacations do not anticipate future jumps of the Poisson arrival process.

Assumption 6. The number of customers that arrive during a vacation is independent of the number of customers present in the system when the vacation starts.

The p.g.f. for the stationary distribution of the number of customers at a random point in time when the server on vacation, $\chi(z)$, as below:

$$\chi(z) = \zeta(z)\frac{1-\alpha(z)}{\alpha'(1)(1-z)}, \tag{1.2}$$

where $\alpha(z) \equiv$ the p.g.f. of the number of customers during vacation period.

$\zeta(z) \equiv$ the p.g.f. of the number of customers when the vacation begins.

It should be note that

$$\pi(z) = \frac{(1-\rho)(1-z)\bar{f}_S(\lambda - \lambda z)}{\bar{f}_S(\lambda - \lambda z)}, \tag{1.3}$$

where $\bar{f}_S(\lambda - \lambda z)$ is the Laplace-Stieltjes transform of service time and in vacation system has exhaustive in case $\zeta(z) \equiv 1$.

1.2.2 Maximum entropy technique

Claude E. Shannon (the father of information theory) defined a property of a probability distribution, $H(p) = -\sum p_i \log p_i$, which he called entropy. The principle of maximum entropy is a technique that can be used to estimate input probabilities more generally. The result is a probability distribution that is consistent with known constraints expressed in terms of averages, or expected values, of one or more quantities, but is otherwise as unbiased as possible. The principle of maximum entropy is a method for analyzing the available information in order to determine an unique epistemic probability distribution. It states that, of all the distributions satisfying the constraints supplied by the given information, the minimally prejudiced distribution which should be chosen is the one that maximizes the Shannon entropy $H(p)$. The maximum entropy principle is like other Bayesian methods in that it makes explicit use of prior information. This is an alternative to the methods of inference of classical statistics. The principle of maximum entropy has been applied successfully in a remarkable variety of fields, including statistical mechanics and thermodynamics, reliability estimation, traffic networks, queueing theory, stock market analysis and general probabilistic problem solving.

We consider a system Q that has a finite or countable infinite set B of all possible discrete states $B_0, B_1, B_2, \cdots, B_n, \cdots$. Let $p(B_n)$ represent the probability that the system Q is in state B_n. Following El-Affendi and Kouvatsos [12], we have the entropy function as follows:

$$H = -\sum_{B_n \in B} p(B_n) \ln p(B_n), \tag{1.4}$$

which is maximized subject to the following constraints

$$\sum_{B_n \in B} p(B_n) = 1, \tag{1.5}$$

and

$$\sum_{B_n \in B} f_k(B_n) p(B_n) = F_k, \quad k = 1, 2, \cdots, m \tag{1.6}$$

where $\{F_k\}$ denotes that the expected values defined on a set of suitable functions $\{f_k(B_n), k = 1, 2, \cdots, m\}$. The maximization of (1.4) subject to constraints (1.5) and (1.6) can be achieved using Lagrange's method of undetermined multipliers leading to the solution

$$p(B_n) = \exp\left\{-\beta_0 - \sum_{k=1}^{m} \beta_k f_k(B_n)\right\}, \tag{1.7}$$

where β_0 is a Lagrangian multiplier determined by the normalization constraint (1.5) and $\{\beta_k, k = 1, 2, \cdots, m\}$ are the Lagrangian multipliers determined from the set of constraints (1.6).

1.3 Literature Review

For a reliable server, Yadin and Naor [58] first introduced the concept of controllable queueing system under N policy. The N policy M/G/1 queueing system was first studied by Heyman [19] and was investigated by such authors as Bell [7], Tijms [46], Wang and Ke [53], and others. Exact steady-state solutions of the N policy M/E_k/1 and M/H_k/1 queueing systems were first developed by Wang and Huang [51] and Wang and Yen [56], respectively. Exact steady-state solutions of the N policy M/M/1 queueing system with exponential startup times were first derived by Baker [5]. Borthakur et al. [8] extended Baker's model to general startup times. The N policy M/G/1 queue with startup times was investigated by several authors such as Medhi and Templeton [34], Takagi [44], Lee and Park [32], Krishna et al. [31], Hur and Paik [21], etc. For an unreliable server, exact steady-state solutions of the N policy M/M/1, M/E_k/1, M/H_2/1, and M/H_k/1 queueing systems were developed by Wang [47], Wang [48], Wang et al. [50], and Wang et al. [52], respectively. Wang and Ke [54] studied three control policies in an M/G/1 queueing system and demonstrated that in three control policies, the probability that the server is busy in the steady-state is equal to the traffic intensity. Recently, Ke [24] examined the N policy M/G/1 queueing system with server vacations, startup and breakdowns.

Ke [25] also analyzed the N policy G/M/1/K queueing system with exponential startup times. Furthermore, Ke and Pearn [27] investigated the N policy M/M/1 queueing system with heterogeneous arrivals, in which the server is characterized by breakdowns and vacations. Analytical sensitivity analysis of the N policy M/G/1 queueing system is investigated by Pearn et al. [35]. Exact steady-state solutions of the N policy M/M/1 queueing system with exponential startup times were first developed by Wang [49].

The controllable queueing problem with T policy has been extensively investigated in the literature. A pioneering work in this field is Heyman [20] who first introduced the T policy. The T policy M/G/1 queue with a reliable server was studied by Heyman [20], Levy and Yechiali [33], and Gakis et al. [15]. Gakis et al. [15] developed the distributions and the first two moments of the busy and idle periods in an M/G/1 queue operating under six dyadic policies. They have shown that in all policy cases the steady-state probability that the server is busy is equal to the traffic intensity. Wang and Ke [54] analyzed an M/G/1 queue with server breakdowns operating under the N policy, the T policy, and the $Min(N,T)$ policy. They performed numerical comparisons of three policies which demonstrated that the optimal N policy and the optimal $Min(N,T)$ policy are superior to T policy. Alfa and Li [1] studied the optimal (N,T) policy for an M/G/1 queue with cost structure. Hur et al. [22] investigated an M/G/1 queue with two policies, N - and T - policy simultaneously. They obtained the steady-state distribution of the system size and determined the optimal operating policy. Tadj [41] proposed an M/G/1 quorum queue operating under the T policy. He used probability generating function technique to obtain the steady-state system characteristics. Recently, Ke [26] examined a modified T vacation policy for an M/G/1 queue with an unreliable server and startup. He derived the explicit formulae for various system performance measures such as the expected number of customers in the system, the expected waiting time in the queue, the expected lengths of the idle, busy, breakdown periods, and the busy cycle, etc.

Due to Jaynes [23], the maximally noncommittal distribution with regard to missing information is that, of all the distributions $p_i(n)$, which maximizes the entropy $-\sum p_i(n)\ln p_i(n)$ under the restrictions induced by the given information and which is least biased to the information missing. The maximum entropy principle is applied to analyze the ordinary queueing systems by several researchers such as Ferdinand [13], Shore [39, 40], Arizono et al. [3], Wu and Chan [57], El-Affendi and Kouvatsos [12], Kouvatsos [29, 30], and so on. Ferdinand [13] utilized the maximum entropy method to derive the steady-state solutions for the ordinary M/M/1 queue with finite capacity. Shore [39] derived the steady-state and time-dependent distributions for the ordinary $M/M/\infty/N$ and $M/M/\infty$ queues by means of entropy maximization. For the ordinary M/G/1 and G/G/1 queues, Shore [40] used the maximum entropy approximation to derive the steady-state system performance measures. Arizono et al. [3] proposed an entropy model to derive the probability distributions of the queue length for the ordinary M/M/S queueing system. Applying the method of entropy maximization to the ordinary GI/G/C queue, Wu and Chan [57] derived the approximate formulate for the steady state probability distributions of the number of customers and the expected waiting time in the system. El-Affendi and Kouvatsos [12] provided the maximum entropy formalism to analyze the ordinary M/G/1 and G/M/1 queues in steady-state. Based on the principle of maximum entropy, Kouvatsos [29, 30] developed a closed-form expression for the queue length distribution of the ordinary G/G/1 queue with both finite capacity and infinite capacity, respectively. The maximum entropy principle has been widely applied to the study of more complicated ordinary queueing systems having general interarrival times, or general service times, or general interarrival times and general service times. Wang et al. [55] used the maximum entropy principle to examine the N policy M/G/1 queue with a reliable server. Artalejo and Maria [4] utilized the maximum entropy principle to investigate the probability density function of the busy period in an M/G/1 vacation models operating under N-, T- and D-policies.

1.4 Scope of Dissertation

The objective of this dissertation is fourfold: First, we analyze an unreliable server in the N and T policies M/G/1 queue with general repair times and startup time. We derive the steady state queue length distribution and obtain various system performance measures. Second, optimal threshold N and T to minimize the overall operating cost of the system are found, respectively. Third, we present sensitivity analysis and some numerical computations to verify the analytical results and show how to make the decision based on minimizing the cost function. Finally, we apply the maximum entropy principle to develop the maximum entropy solutions and perform a comparative analysis between the maximum entropy results and exact results.

Chapter 1 is an introduction. In Chapter 1, we review some controllable queues and some techniques relevant to this study such as M/G/l decomposition property and maximum entropy principle.

In Chapter 2, we study the N policy M/G/1 queue with server breakdowns and general startup times. We first develop various system performance measures, such as the expected number of customers, the expected length of the turned-off, complete startup, busy, and breakdown periods. Next, we construct the total expected cost function per unit time to determine the optimal threshold N numerically in order to minimize the cost function. In addition, the analytic results for sensitivity analysis are derived. Furthermore, we investigate some numerical examples.

In Chapter 3, we use maximum entropy principle to study a single removable and unreliable server in the N policy M/G/1 queue with general startup times where arrivals form a Poisson process and service times are generally distributed. The purpose of this chapter is: (i) to provide the maximum entropy formalism for the N policy M/G/1 queue with general repair times and general startup times; (ii) to develop the maximum entropy (approximate) solutions for the N policy M/G/1 queue with general repair times and general startup times by using Lagrange's method; (iii) to obtain approximate results for the expected waiting time in queue; (iv) to perform a comparative analysis between the approximate results with established exact results for various distributions, such as exponential (M), k-stage Erlang (E_k), and deterministic (D). We demonstrate that the maximum entropy approach is accurate enough for practical purposes and is a useful method for solving complex queueing system.

In Chapter 4, we develop the probability generating function and various system performance measures such as the expected number of customers in the system, the expected length of the idle, busy, and breakdown period, and the expected length of the busy cycle, etc. Based on the derived results, we construct the total expected cost function per unit time, including customer holding cost, the system setup cost, turn the server on and off costs, server startup cost, and server breakdown cost. We determine the optimal threshold T numerically to minimize the total expected cost. In addition, numerical results and sensitivity investigations are also presented

In Chapter 5, we use maximum entropy principle to study a single removable and unreliable server in the T policy M/G/1 queue with general repair times and general startup times. First, we develop the maximum entropy (approximate) solutions for the T policy M/G/1 queue with general repair times and general startup times by using Lagrange's method under the constraint "the first moment of the queue length and then to obtain approximate results for the expected waiting time in queue. Next, we develop the maximum entropy solutions under the constraint "the second moment of the queue length" and then to obtain approximate results for the expected waiting time in queue. Finally, we perform a comparative analysis between the approximate results with established exact results for various distributions under the constraints of first moment and second moment of the queue length.

Chapter 6 presents some conclusions based on results of the study, and recommendations for further investigations.

Chapter 2

Optimization of the *N* Policy M/G/1 Queue with Server Breakdowns and Startup Times

In this chapter, we deal with the N policy M/G/1 queue with a single removable and unreliable server whose arrivals form a Poisson process. Service times, repair times, and startup times are assumed to be generally distributed. When the queue length reaches N ($N \geq 1$), the server is immediately turned on but is temporarily unavailable to serve the waiting customers. The server needs a startup time before providing service until there are no customers in the system. Firstly, we develop various system performance measures, such as the expected number of customers, the expected length of the turned-off, complete startup, busy, and breakdown periods. Next, we construct the total expected cost function per unit time to determine the optimal threshold N numerically in order to minimize the cost function In addition, sensitivity analysis and some numerical examples are also investigated

2.1 Assumptions and Notations

It is assumed that arrivals of customers follow a Poisson process with rate λ. The service times for a customer are independent and identically distributed (i.i.d.) random variables obeying an arbitrary distribution function $F_S(t)$ ($t \geq 0$) with a mean service time μ_S and a finite variance σ_S^2. The server is subject to breakdowns at any time with Poisson breakdown rate α when he is working. When the server fails, he is immediately repaired at a repair facility, where the repair times are i.i.d. random variables having a general distribution function $F_R(t)$ ($t \geq 0$) with a mean repair time μ_R and a finite variance σ_R^2. Arriving customers form a single waiting line at a server based on the order of their arrivals. The server can serve only one customer at a time and the service is independent of the arrival process. A customer who arrives and finds the server busy or broken down must wait in the queue until a server is available. Although no service occurs during the repair period of the server, customers continue to arrive following a Poisson process. Furthermore, when the queue length reaches a specific level, denoted by N, the server is immediately turned on (i.e. begin startup) but is temporarily unavailable to serve the waiting customers. He needs a startup time with random length before starting service. Again, the startup times are i.i.d. random variables obeying a general distribution function $F_U(t)$ $(t \geq 0)$, with a mean startup time μ_U and a finite variance σ_U^2. Once the startup is terminated, the server begins serving the waiting customers until the system becomes empty. Service is allowed to be interrupted if the server breaks down, and the server is immediately repaired. Once the server is repaired, he immediately returns to serve customers until there are no customers in the system.

In this chapter, the following notations and probabilities are used.

$N-$ threshold

$S-$ service time random variable

$U-$ startup time random variable

$R-$ repair time random variable

$F_S(\cdot)-$ distribution function of S

$F_U(\cdot)-$ distribution function of U

$F_R(\cdot)-$ distribution function of R

$G(z)-$ p.g.f. of the number of customers in the ordinary M/G/1 queue with unreliable server

$G_N(z)-$ p.g.f. of the number of customers in the N policy M/G/1 queue with server breakdowns and general startup times

$W(z)-$ p.g.f. of the number of customers that arrive during the turned-off plus the startup period

$\bar{f}_U(\cdot)-$ Laplace-Stieltjes transform (LST) of startup time

L_N- expected number of customers in the N policy M/G/1 queue with server breakdowns and general startup times

H_O- complete period (busy period plus breakdown period) of the ordinary M/G/1 queue with server breakdowns

I_N- turned-off period of the N policy M/G/1 queue with server breakdowns and general startup times

U_N- startup period of the N policy M/G/1 queue with server breakdowns and general startup times

B_N – busy period of the N policy M/G/1 queue with server breakdowns and general startup times

D_N – breakdown period of the N policy M/G/1 queue with server breakdowns and general startup times

H_N – complete period which is equal to $(B_N + D_N)$

V_N – complete startup period which is equal to $(U_N + H_N)$

C_N – busy cycle which is equal to $(I_N + V_N)$

$F_{V_N}(\cdot)$ – distribution function of V_N

$F_{H_O}(\cdot)$ – distribution function of H_O

$F_{HO}^{(N+n)}(\cdot)$ – $(N+n)$-fold convolution of $F_{H_O}(\cdot)$

$\bar{f}_{V_N}(\cdot)$ – LST of V_N

P_{I_N} – probability that the server is turned-off in the N policy M/G/1 queue with server breakdowns and general startup times

P_{U_N} – probability that the server is startup in the N policy M/G/1 queue with server breakdowns and general startup times

P_{B_N} – probability that the server is busy in the N policy M/G/1 queue with server breakdowns and general startup times

P_{D_N} – probability that the server is broken down in the N policy M/G/1 queue with server breakdowns and general startup times

C_h – holding cost per unit time for each customer present in the system;

C_s – setup cost for per busy cycle

C_i – cost per unit time for keeping the server off;

C_{sp} – startup cost per unit time for the preparatory work of the server before starting the service

C_b – cost per unit time for keeping the server on and in operation

C_d – breakdown cost per unit time for a failed server

2.2 Justification of Practical Applications

A number of practical problems arise which may be formulated as one in which the server meets unpredictable breakdowns and requires a startup time before providing service. Such models have potentially useful in practical production (manufacturing) systems. For example, in reflow work for Printed Circuit Board (PCB) Surface Mount. Assume that PCB arrives according to a random process. For cost concern, it is desirable that the reflow machine begins operating whenever the number of PCB reaches a critical value N. It takes random time for warming up before the reflow machine starts working. Moreover, the reflow process may be interrupted when machine encounters unpredicted breakdowns. When reflow interruptions occur (breakdowns), it is emergently recovered with a random time. Another possible application is wire bonding in Integrated Circuit (IC) assembly. To save cost, it is desirable that the wire bonder begins operating whenever the number of unbounded IC reaches a critical value N. It requires a random time for setup before the wire bonder starts working. The bonding process may be interrupted when the bonder meets breakdowns. When bonding interruptions occur (breakdowns), it is emergently recovered.

2.3 System Performance Measures

The primary objective of this section is to develop the various system performance measures, such as (i) expected number of customers in the system; (ii) expected length of the turned-off period, the complete startup period, the busy period, and the breakdown period; (iii) expected length of the busy cycle; and (iv) the probability that the server is turned-of, startup, busy and broken down

2.3.1 Expected number of customers in the system

Let H be a random variable representing the completion time of a customer, which includes both the service time of a customer and the repair time of a server. Applying the well-known formula for the p.g.f. of the number of customers in the ordinary M/G/1 queue with reliable server, the p.g.f. of the number of customers in ordinary M/G/1 queue with unreliable server is given by

$$G(z)=\frac{(1-\rho_H)(1-z)\overline{f}_H(\lambda-\lambda z)}{\overline{f}_H(\lambda-\lambda z)-z}, \tag{2.1}$$

where $\rho_H = \lambda E[H]$. In addition, $E[H] = \mu_S(1+\alpha\mu_R)$ and

$$E[H^2] = (1+\alpha\mu_R)^2(\mu_S^2+\sigma_S^2)+\alpha\mu_S(\mu_R^2+\sigma_R^2),$$

(see Wang and Ke [54]). It is to be noted that the traffic intensity ρ_H is assumed to be less than unity. We note that expression (2.1) is obtained only by replacing service times by completion times in the formula of the ordinary M/G/1 queue with reliable server.

For the N policy M/G/1 queue with server breakdowns requiring startup time, we consider that the server is on 'extended vacation' during the turned-off period plus the startup period. Following the result of Medhi and Templeton [34], we obtain

$$G_N(z) = \frac{[1-W(z)]G(z)}{W'(1)(1-z)}, \tag{2.2}$$

$G_N(z) \equiv$ the p.g.f. of number of customers in the N policy M/G/1 queue with server breakdowns and general startup times;

$W(z) \equiv$ the p.g.f. of the number of customers that arrive during the turned-off plus the startup period;

$\equiv$ [the p.g.f. of the number of customers that arrive during the turned-off period]$\times$[the p.g.f. of the number of customers that arrive during the startup period];

$\equiv z^N \bar{f}_U(\lambda-\lambda z)$, where $\bar{f}_U(\cdot)$ is the LST of startup time.

We have $W'(z) = Nz^{N-1}\bar{f}_U(\lambda-\lambda z) - \lambda z^N \bar{f}_U^{(1)}(\lambda-\lambda z)$. It follows that $W'(1) = N +$ $\lambda\mu_U$, where $\mu_U = -\bar{f}_U^{(1)}(0)$ is the mean startup time. Let $\rho_U = \lambda\mu_U$. From (2.1) and (2.2), we obtain

$$G_N(z) = \frac{\left[1-z^N\bar{f}_U(\lambda-\lambda z)\right](1-\rho_H)\,\bar{f}_H(\lambda-\lambda z)}{(N+\rho_U)\left[\bar{f}_H(\lambda-\lambda z)-z\right]}.$$

Let L_N denote the expected number of customers in the N policy M/G/1 queue with server breakdowns and general startup times. Thus we have

$$L_N = G'_N(z)\big|_{z=1}$$

$$= \frac{1}{(N+\rho_U)}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda E[U^2]}{2}\right] + \rho_H + \frac{\lambda^2 E[H^2]}{2(1-\rho_H)}. \qquad (2.3)$$

2.3.2 Expected length of the turned-off, complete startup, busy, and breakdown periods

The turned-off period terminates when the N-th customer arrives in system. Since the complete startup period starts when the turned-off period terminates, the complete startup period is represented by the sum of the startup period and the complete period. The server begins startup when there are at least N waiting customers in the system. This is called the startup period. The startup period terminates when the server starts to serve the waiting customers. Since the complete period begins when the startup period is over and terminates when the system becomes empty, the complete period is represented by the sum of the busy period and the breakdown period. The busy period is initiated when the server completes his startup and begins serving the waiting customers. During the busy period, the server may break down and starts his repair immediately. This is call the breakdown period. After the server is repaired, he returns immediately and provides service until there are no customers in the system.

Let H_O be the complete period of the ordinary M/G/1 queue with server breakdowns. Using the well-known result of Kleinrock [28, p. 213], we obtain the expected length of the complete period for the ordinary M/G/1 queue with server breakdowns as

$$E[H_O] = \frac{\mu_S(1+\alpha\mu_R)}{1-\rho(1+\alpha\mu_R)}, \qquad (\rho = \lambda\mu_S). \qquad (2.4)$$

2.3.2.1 Expected length of the turned-off period

We know that the turned-off period I_N terminates when the N-th customer arrives in system. Since the length of times between two successive arrivals are independently, identically and exponentially distributed with mean $1/\lambda$, thus the expected length of the turned-off period, $E[I_N]$, for the N policy M/G/1 queue with server breakdowns

$$E[I_N]=\frac{N}{\lambda}. \tag{2.5}$$

2.3.2.2 Expected length of the complete startup period

Let V_N represent the complete startup period for the N policy M/G/1 queue with server breakdowns and general startup times. Since the complete startup period is the sum of the complete period and the startup period which implies $V_N = H_N + U_N$, where H_N and U_N denote the complete period and the startup period, respectively. Let $\bar{f}_{V_N}(\cdot)$ be the LST of the distribution of the complete startup period of the N policy M/G/1 queue with server breakdowns.

The following notations are used.

$F_{V_N}(\cdot)-$ distribution function of the complete startup period V_N of the N policy M/G/1 queue with server breakdowns and general startup times.

$\bar{f}_U(\cdot)-$ the LST of startup time

$F_{H_O}(\cdot)-$ distribution function of the complete period H_O of the ordinary M/G/1 queue with server breakdowns.

$F_{H_O}^{(N+n)}(\cdot)-$ $(N+n)-$fold convolution of $F_{H_O}(\cdot)$.

By conditioning V_N on the length of the startup time $=t$ and the number of arrivals during U, we obtain (see [18, p. 277])

$$\begin{aligned}F_{V_N}(x) &= \int_0^x \sum_{n=0}^{\infty} P(\text{given any startup time} = t, \text{completion startup period} \\ &\quad \text{generated by } N \text{ customers arrival plus } n \text{ customers} \\ &\quad \text{arrival in the complete period } H_O \text{ during } t \le x-t)\, dF_U(t) \\ &= \int_0^x \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} F_{H_O}^{(N+n)}(x-t)\, dF_U(t). \end{aligned} \tag{2.6}$$

Taking the LST of both sides of (2.6) yields

$$\bar{f}_{V_N}(s) = \int_0^{\infty}\int_0^x \sum_{n=0}^{\infty} e^{-sx} \frac{e^{-\lambda t}(\lambda t)^n}{n!} F_{H_O}^{(N+n)}(x-t)\, dF_U(t) dx. \tag{2.7}$$

Changing the order of integration of (2.7), it finally gets

$$\bar{f}_{V_N}(s)=\int_0^\infty e^{-\lambda t}\left\{\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\left[\int_t^\infty e^{-sx}F_{H_0}^{(N+n)}(x-t)\,dx\right]\right\}dF_U(t)$$

$$=\int_0^\infty e^{-\lambda t}\left\{\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}e^{-st}\left[\bar{f}_{H_O}(s)\right]^{N+n}\right\}dF_U(t)$$

$$=\left[\bar{f}_{H_O}(s)\right]^N\int_0^\infty e^{-(s+\lambda)t}\sum_{n=0}^{\infty}\frac{\left[\lambda\bar{f}_{H_O}(s)t\right]^n}{n!}dF_U(t)$$

$$=\left[\bar{f}_{H_O}(s)\right]^N\int_0^\infty e^{-\left[s+\lambda-\lambda\bar{f}_{H_O}(s)\right]t}dF_U(t)$$

$$=\left[\bar{f}_{H_O}(s)\right]^N\bar{f}_U\left[\lambda+s-\lambda\bar{f}_{H_O}(s)\right]. \tag{2.8}$$

Differentiating (2.8) with respect to s, we obtain the expected length of the complete startup period as follows:

$$E[V_N]=(N+\lambda\mu_U)E[H_O]+\mu_U=\frac{(N+\lambda\mu_U)\mu_S(1+\alpha\mu_R)}{1-\rho(1+\alpha\mu_R)}+\mu_U. \tag{2.9}$$

2.3.2.3 Expected length of the busy and breakdown periods

The expected length of the complete period and the expected length of the startup period are denoted by $E[H_N]$ and $E[U_N]$, respectively. Recall that $V_N=H_N+U_N$ which implies $E[V_N]=E[H_N]+E[U_N]$. Hence from (2.9) and (2.4), we obtain

$$E[H_N]=(N+\lambda\mu_U)E[H_O]=\frac{(N+\lambda\mu_U)\mu_S(1+\alpha\mu_R)}{1-\rho(1+\alpha\mu_R)}, \tag{2.10}$$

$$E[U_N]=\mu_U. \tag{2.11}$$

Let $E[B_N]$ and $E[D_N]$ be the expected length of the busy period and the expected length of the breakdown period, respectively. Recall that the complete period is the sum of the busy period and the breakdown period which implies $E[H_N]=E[B_N]+E[D_N]$. Hence from (2.10) we have

$$E[B_N]=\frac{(N+\lambda\mu_U)\mu_S}{1-\rho(1+\alpha\mu_R)}, \tag{2.12}$$

$$E[D_N]=\frac{(N+\lambda\mu_U)\alpha\mu_S\mu_R}{1-\rho(1+\alpha\mu_R)}. \tag{2.13}$$

2.3.3 Expected length of the busy cycle

The busy cycle for the N policy M/G/1 queue with server breakdowns and general startup times, denoted by C_N, is the length of time from the beginning of the last turned-off period to the beginning of the next turned-off period. Since the busy cycle is the sum of the turned-off period (I_N), the startup period (U_N), the busy period (B_N), and the breakdown period (D_N), we get

$$E[C_N] = E[I_N] + E[U_N] + E[B_N] + E[D_N] = E[I_N] + E[V_N]. \tag{2.14}$$

From (2.5) and (2.9), we obtain

$$E[C_N] = \frac{N + \lambda\mu_U}{\lambda\left[1 - \rho(1 + \alpha\mu_R)\right]}. \tag{2.15}$$

2.3.4 Probability that the server is turned-off, startup, busy and broken down

In steady-state, let

$P_{I_N} \equiv$ probability that the server is turned-off.

$P_{U_N} \equiv$ probability that the server is startup.

$P_{B_N} \equiv$ probability that the server is busy.

$P_{D_N} \equiv$ probability that the server is broken down.

We obtain

$$P_{I_N} = \frac{E[I_N]}{E[C_N]}, \tag{2.16}$$

$$P_{U_N} = \frac{E[U_N]}{E[C_N]}, \tag{2.17}$$

$$P_{B_N} = \frac{E[B_N]}{E[C_N]}, \tag{2.18}$$

$$P_{D_N} = \frac{E[D_N]}{E[C_N]}. \tag{2.19}$$

Substituting $E[I_N]$ in (2.5), $E[U_N]$ in (2.11), $E[B_N]$ in (2.12), $E[D_N]$ in (2.13), and $E[C_N]$ in (2.15) into relations (2.16)-(2.19) yields the probability that the server is turned-off, startup, busy and broken down in the following:

$$P_{I_N} = \frac{N(1 - \rho_H)}{N + \rho_U}, \tag{2.20}$$

$$P_{U_N} = \frac{\rho_U(1-\rho_H)}{N+\rho_U}, \qquad (2.21)$$

$$P_{B_N} = \rho, \qquad (2.22)$$

$$P_{D_N} = \alpha\rho\mu_R. \qquad (2.23)$$

We prove from (2.22) that the probability that the server is busy in the steady-state is equal to ρ.

2.4 The Optimal *N* Policy

We develop an expected cost function per unit time for the N policy M/G/1 queue with server startup and breakdowns in which N is a decision variable. Our objective is to determine the optimum value of the control parameter N, say N^*, so as to minimize this function. We define the following cost elements:

$C_h \equiv$ holding cost per unit time for each customer present in the system;

$C_s \equiv$ setup cost for per busy cycle;

$C_i \equiv$ cost per unit time for keeping the server off;

$C_{sp} \equiv$ startup cost per unit time for the preparatory work of the server before starting the service;

$C_b \equiv$ cost per unit time for keeping the server on and in operation;

$C_d \equiv$ breakdown cost per unit time for a failed server.

Utilizing the definition of each cost element listed above, the expected cost function per unit time per customer is given by

$$F_O(N) = C_h L_N + C_s \frac{1}{E[C_N]} + C_i \frac{E[I_N]}{E[C_N]} + C_{sp} \frac{E[U_N]}{E[C_N]} + C_b \frac{E[B_N]}{E[C_N]} + C_d \frac{E[D_N]}{E[C_N]}, \qquad (2.24)$$

where L_N is given in (2.3). We note that $\rho_H + \left(\lambda^2 E[H^2] \big/ 2(1-\rho_H)\right)$, $E[B_N] \big/ E[C_N]$ and $E[D_N] \big/ E[C_N]$ do not involve the decision variable N. Omitting these cost terms are not functions of the decision variable N. The optimization problem in (2.24) is equivalent to minimize the following equation:

$$F(N)=\frac{1}{N+\rho_U}\left\{\frac{C_h N^2}{2}-\left[C_h\left(\frac{1}{2}-\rho_U\right)-C_i\left(1-\rho_H\right)\right]N\right.$$

$$\left.+\frac{C_h\lambda^2 E[U^2]}{2}+\left(C_s\lambda+C_{sp}\rho_U\right)\left(1-\rho_H\right)\right\}. \qquad (2.25)$$

Differentiating $F(N)$ with respect to N, we get

$$\frac{dF(N)}{dN}=\frac{C_h}{2}-\frac{C_h}{2\left(N+\rho_U\right)^2}\left\{\rho_U+\lambda^2\sigma_U^2+\frac{2}{C_h}\left[C_s\lambda+\left(C_{sp}-C_i\right)\rho_U\right](1-\rho_H)\right\}.$$

Setting $dF(N)/dN=0$ yields

$$N^*=-\rho_U+\sqrt{\rho_U+\lambda^2\sigma_U^2+\frac{2}{C_h}\left[C_s\lambda+\left(C_{sp}-C_i\right)\rho_U\right](1-\rho_H)}, \qquad (2.26)$$

where

$$\rho_U=\lambda\mu_U=\frac{\lambda}{\gamma} \quad (\mu_U=\frac{1}{\gamma}),$$

and

$$\rho_H=\lambda E[H]=\lambda\mu_S(1+\alpha\mu_R)=\frac{\lambda}{\mu}\left(1+\frac{\alpha}{\beta}\right) \quad (\mu_S=\frac{1}{\mu},\ \mu_R=\frac{1}{\beta}).$$

Differentiating $F(N)$ with respect to N twice and using (2.26), we obtain

$$\frac{d^2F(N)}{dN^2}=C_h\left\{\rho_U+\lambda^2\sigma_U^2+\frac{2}{C_h}\left[C_s\lambda+\left(C_{sp}-C_i\right)\rho_U\right](1-\rho_H)\right\}^{\frac{-1}{2}}>0, \quad (\rho_H<1).$$

Thus N^* is the unique minimizer of $F(N)$. If N^* is not an integer, the best positive integer value of N is one of the integers surrounding N^*.

2.5 Sensitivity Analysis

A system analyst often concerns with how the system performance measures can be affected by the changes of the input parameters in the investigated queueing service model. Sensitivity investigation on the queueing model with critical input parameters may provide some answers to this question. In the following, we conduct some sensitivity investigations on the optimal value N^* based on changes in the values of the system parameters $\lambda, \mu, \alpha, \beta, \gamma$ and cost parameters C_h, C_s, C_i, C_{sp}.

From (2.26), we perform some algebraic manipulations with respect to system parameters $\lambda, \mu, \alpha, \beta, \gamma$. Differentiating N^* with respect to λ, we obtain

$$\frac{\partial N^*}{\partial \lambda} = -\mu_U + \frac{(\mu_U + \theta_1)^{\frac{1}{2}}}{2\sqrt{\lambda}}, \quad if \quad \sigma_U^2 - \theta_1 E[H] = 0, \tag{2.27}$$

$$\frac{\partial N^*}{\partial \lambda} = -\mu_U + \frac{\mu_U + 2\lambda\sigma_U^2 + \theta_1(1-2\rho_H)}{2\sqrt{\rho_U + \lambda^2\sigma_U^2 + \lambda\theta_1(1-\rho_H)}}, \quad if \quad \sigma_U^2 - \theta_1 E[H] \neq 0, \tag{2.28}$$

where

$$\theta_1 = \frac{2\left[C_s + (C_{sp} - C_i)\mu_U\right]}{C_h}.$$

Setting (2.27) and (2.28) to zero, and then solving for λ, we find

$$\lambda = \frac{\mu_U + \theta_1}{4\mu_U^2}, \quad if \quad \sigma_U^2 - \theta_1 E[H] = 0, \tag{2.29}$$

$$\lambda = \frac{\mu_U + \theta_1}{2(\sigma_U^2 - \theta_1 E[H])}\left[-1 + \frac{\mu_U}{\sqrt{\mu_U^2 - \sigma_U^2 + \theta_1 E[H]}}\right], \quad if \quad \sigma_U^2 - \theta_1 E[H] \neq 0. \tag{2.30}$$

Note that in (2.30) the conditions of $\mu_U^2 - \sigma_U^2 + \theta_1 E[H] > 0$ are required. Differentiating (2.27) and (2.28) with respect to λ again and substituting (2.29) and (2.30) into the resulting differentiation from (2.27) and (2.28), respectively, we have

$$\frac{\partial^2 N^*}{\partial \lambda^2} = -\frac{2\mu_U^3}{\mu_U + \theta_1} < 0, \quad if \quad \sigma_U^2 - \theta_1 E[H] = 0, \tag{2.31}$$

$$\frac{\partial^2 N^*}{\partial \lambda^2} = -\frac{2\left(\mu_U^2 - \sigma_U^2 + \theta_1 E[H]\right)^{\frac{3}{2}}}{\mu_U + \theta_1} < 0, \quad if \quad \sigma_U^2 - \theta_1 E[H] \neq 0. \tag{2.32}$$

The above results show that the graph of N^* is concave downward with respect to λ, which attains its maximum value under two different parameter settings satisfying (2.29) and (2.30), respectively. Differentiating N^* with respect to μ yields

$$\frac{\partial N^*}{\partial \mu} = \frac{\rho\theta_1\rho_H}{2\sqrt{\rho_U + \lambda^2\sigma_U^2 + \lambda\theta_1(1-\rho_H)}} > 0, \tag{2.33}$$

where $\rho = \lambda\mu_S = \lambda/\mu$. Thus N^* is increasing in μ. Similarly, differentiating N^* with respect to α and β, respectively, we obtain

$$\frac{\partial N^*}{\partial \alpha} = \frac{-\lambda\rho\theta_1}{2\beta\sqrt{\rho_U + \lambda^2\sigma_U^2 + \lambda\theta_1(1-\rho_H)}} < 0, \tag{2.34}$$

$$\frac{\partial N^*}{\partial \beta} = \frac{\lambda\rho\alpha\theta_1}{2\beta^2\sqrt{\rho_U + \lambda^2\sigma_U^2 + \lambda\theta_1(1-\rho_H)}} > 0. \tag{2.35}$$

From (2.34) and (2.35), we see that N^* is decreasing in α and N^* is increasing in β. If the startup time distribution is given, we can easily see how γ affects N^* because σ_U^2 is a function of the parameter γ. For example, suppose the startup time distribution obeys the Erlang-*k* (*k* > 1) stage distribution with mean μ_U ($=1/\gamma$). Substituting $\sigma_U^2 = \mu_U^2/k$ into (2.26) and then differentiating N^* with respect to μ_U , we get

$$\frac{\partial N^*}{\partial \mu_U} = -\lambda + \frac{\left(\frac{2}{k}\lambda^2\mu_U + \theta_2\right)}{2\sqrt{\frac{1}{k}\lambda^2\mu_U^2 + \mu_U\theta_2 + \theta_3}}, \tag{2.36}$$

where

$$\theta_2 = 1 + \frac{2(C_{sp} - C_i)(1-\rho_H)}{C_h},$$

and

$$\theta_3 = \frac{2C_s(1-\rho_H)}{C_h}.$$

Two situations are considered while investigating the behavior of $\partial N^*/\partial \mu_U$:

Case (i): If $\theta_2^2 - 4\lambda\theta_3 > 0$ and setting $\partial N^*/\partial \mu_U = 0$, then we obtain

$$\mu_U = \frac{k}{2\lambda}\left(-\theta_2 + \sqrt{\frac{k\theta_2^2 - 4\lambda\theta_3}{k-1}}\right), \tag{2.37}$$

which is a unique solution. Differentiating $\partial N^*/\partial \mu_U$ with respect to μ_U again and using (2.37), we finally get

$$\frac{\partial^2 N^*}{\partial \mu_U^2} = \frac{\frac{\sqrt{\lambda}}{k}\left(k\theta_2^2 - 4\lambda\theta_3\right)}{4\left(\frac{\lambda}{k}\mu_U^2 + \mu_U\theta_2 + \theta_3\right)^{\frac{3}{2}}} < 0 . \tag{2.38}$$

Hence N^* is a concave downward function of μ_U. Since $\mu_U = 1/\gamma$, then it implies that N^* is also a concave downward function of γ. Therefore, from (2.37), we may obtain

$$\gamma = \frac{2(k-1)\lambda}{k\left(\theta_2^2 - 4\lambda\theta_3\right)}\left(\theta_2 + \sqrt{\frac{k\theta_2^2 - 4\lambda\theta_3}{k-1}}\right). \tag{2.39}$$

Case (ii): If $\theta_2^2 - 4\lambda\theta_3 \le 0$, we can see from (2.36) that N^* is a decreasing function of μ_U. It also implies that N^* is an increasing function of γ.

In Case (i) and Case (ii) we see how the value N^* is affected by the input parameter γ. On the other hand, it can easily see from (2.26) that N^* is increasing in C_s, C_{sp} and decreasing in C_h, C_i.

2.6 Numerical Computations

We present some numerical computations to demonstrate the analytical results obtained, and show how to make the decision based on minimizing the cost function (see (2.25)). Since the cost function is only related to system parameters λ, μ, α, β, γ in which $\mu_S = 1/\mu$, $\mu_R = 1/\beta$, $\mu_U = 1/\gamma$ and σ_U^2 is a function of γ, then (2.25) is independent of service time distribution and repair time distribution except for startup time distribution. The sensitivity investigation focuses on the Erlang-2 startup time distribution. First, we fix the following cost parameters C_s=1000, C_h=5, C_{sp}= 100, C_i=60 and consider the following five cases.

Case 1: We select μ=0.5, 1, 1.5, 2, α=0.05, β=3, γ=3, and vary the values of λ.

Case 2: We select λ=0.2, 0.4, 0.6, 0.8, α = 0.05, β=3, γ=3, and vary the values of μ.

We observe from Figure 2.1 that (i) the local maximum value of N^* is moving from left to right as μ increases; and (ii) as λ is fixed, N^* is getting larger as μ increases. From Figure 2.2, we see that (i) N^* increases in μ; (ii) if μ is small enough, N^* increases quickly; (iii) if μ is large and $\rho = \lambda/\mu$ **i**s small enough,

N^* is insensitive; and (iv) if μ is fixed and large enough, N^* increases in λ. Numerical results of Case 1 and Case 2 are provided in Table 2.1.

Case 3: We select $\lambda = 0.5$, $\mu = 1$, $\beta = 1, 2, 3, 4$, $\gamma = 3$ and vary the values of α.

Case 4: We select $\lambda = 0.5$, $\mu = 1$, $\alpha = 0.4, 0.8, 1.2, 1.6$, $\gamma = 3$ and vary the values of β.

We observe from Figure 2.3 that (i) N^* decreases in α. As α is fixed, the larger β has larger N^*; (ii) N^* has an upper bound as α closes to zero, and (iii) N^* is not insensitive to α. It can easily observe from Figure 4 that (i) N^* increases in β but N^* is insensitive to β as β is large; and (ii) as β is fixed, the larger α has smaller N^*. Numerical results of Case 3 and Case 4 are listed in Table 2.2.

Case 5: We select $\lambda = 0.3$, $\mu = 1$, $\alpha = 0.5$, $\beta = 3$ and vary the values of γ.

Figure 2.5 indicates that N^* increases in γ if $\theta_2^2 - 4\lambda\theta_3 \le 0$ 2. However, for another set of cost parameters C_s =500, C_h =5, C_{sp} =100, C_i =40. Parameters satisfying $\theta_2^2 - 4\lambda\theta_3 > 0$, N^* has a unique maximum value at

$$\gamma = \frac{2(k-1)\lambda}{k\left(\theta_2^2 - 4\lambda\theta_3\right)}\left(\theta_2 + \sqrt{\frac{k\theta_2^2 - 4\lambda\theta_3}{k-1}}\right).$$

(see Figure 2.6). Figure 2.5 and Figure 2.6 show that N^* may be too insensitive to changes in γ as γ is greater than 0.4. The numerical results are presented in Table 2.3.

To see how N^* changes when cost parameter changes, we select λ =0.3, μ =1, α =0.05, β =3, γ =3, choose C_{sp} =100, C_i =60, and vary the specified values of (C_s , C_h). Table 2.4 shows that N^* increases in C_s and decreases in C_h. On the other hand, we select C_s =1000, C_h =5 and change the specified values of (C_{sp} , C_i). Table 2.5 reveals that N^* is insensitive to (C_{sp} , C_i).

Finally, we make comparisons between our model and existing literature (see Pearn et al. [35]). According to the parameters setting by [35], we perform a numerical experiment based on λ =0.4, μ_S =1, σ_S =1, α =0.05, μ_R =0.2, σ_R =1, C_h =5, C_b =50, C_i =10, C_d =100 and C_s =200. In addition, we fix startup cost C_{sp} = 90 and vary the parameter value (γ) of exponential startup distribution from 0.1 to 1 and N from 1 to 25. Figure 2.7 shows that our model approaches to that by [35] as γ tends to large enough (μ_U tends to small enough).

Figure 2.1 Plots of (λ, N^*) with μ =0.5, 1.0, 1.5, 2.0, α =0.05, β =3, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

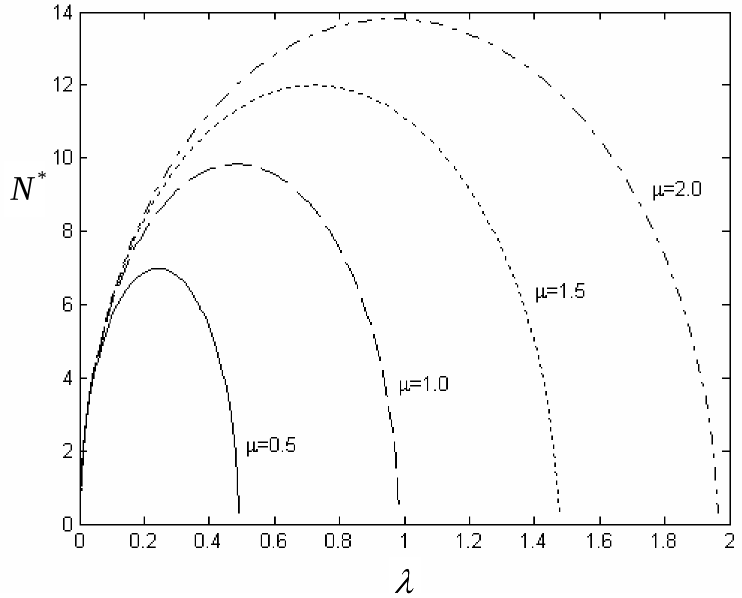


Figure 2.2 Plots of (μ, N^*) with λ =0.2, 0.4, 0.6, 0.8, α =0.05, β =3, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

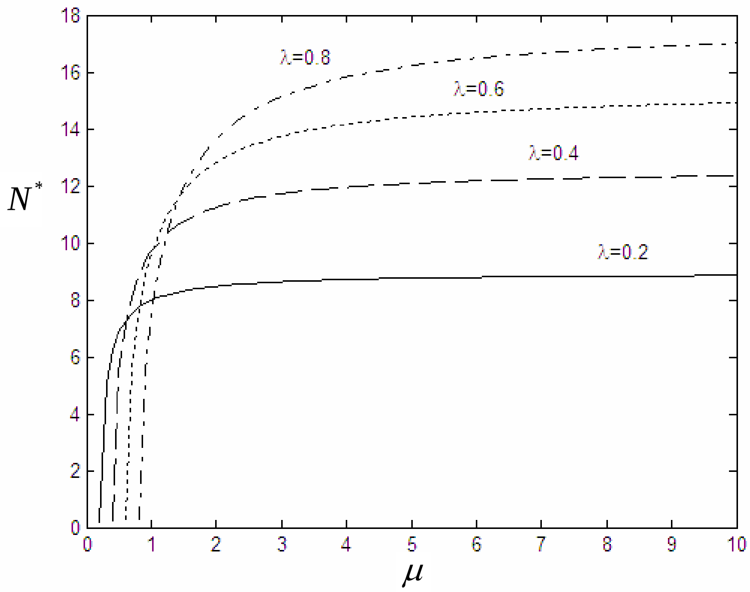


Table 2.1 The optimal N^* and minimum expected cost $F(N^*)$ with various values of (λ, μ).

$\alpha = 0.05$, $\beta = 3$, $\gamma = 3$, $C_s = 1000$, $C_h = 5$, $C_{sp} = 100$, $C_i = 60$								
(λ, μ)	(0.3, 0.5)	(0.3, 1.0)	(0.3, 1.5)	(0.3, 2.0)	(0.2, 1.0)	(0.4, 1.0)	(0.6, 1.0)	(0.8, 1.0)
N^*	7	9	10	10	8	10	10	8
$F(N^*)$	55.3856	85.1964	94.5549	99.1349	85.5034	82.2018	69.7010	47.7634

Figure 2.3 Plots of (α, N^*) with λ =0.5, μ =1, β =1, 2, 3, 4, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

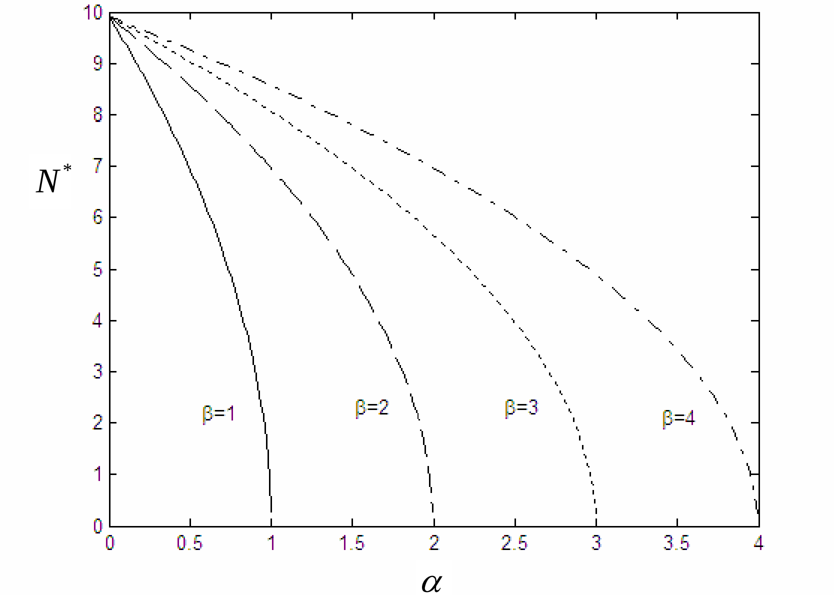


Figure 2.4 Plots of (β, N^*) with λ =0.5, μ =1, α =0.4, 0.8, 1.2, 1.6, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

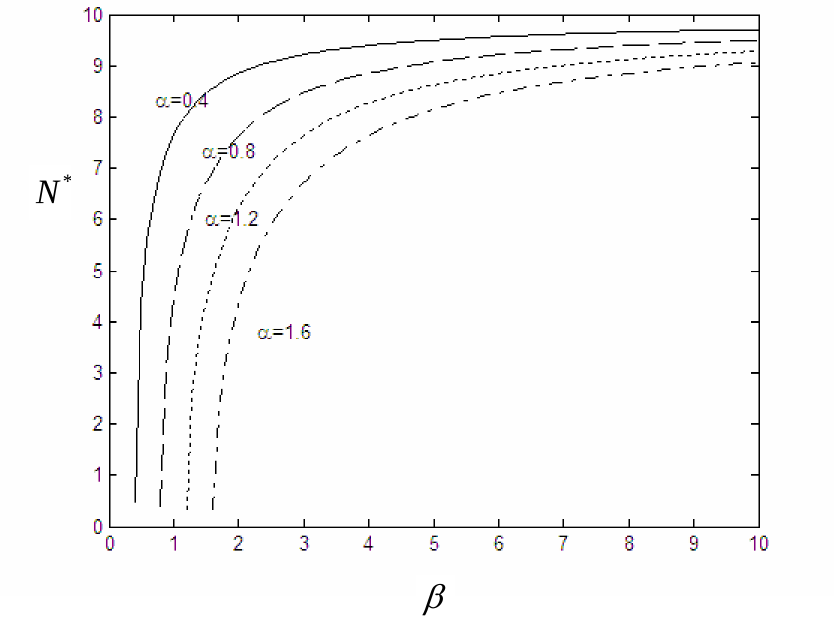


Table 2.2 The optimal N^* and minimum expected cost $F(N^*)$ with various values of (α, β).

$\lambda = 0.5$, $\mu = 1$, $\gamma = 3$, $C_s = 1000$, $C_h = 5$, $C_{sp} = 100$, $C_i = 60$								
(α, β)	(0.5, 1.0)	(0.5, 2.0)	(0.5, 3.0)	(0.5, 4.0)	(0.4, 2.0)	(0.8, 2.0)	(1.2, 2.0)	(1.6, 2.0)
N^*	7	9	9	9	9	8	6	4
$F(N^*)$	48.6806	63.8626	68.4962	70.8162	66.7410	54.7399	41.5367	26.2598

Figure 2.5 Plots of (γ, N^*) with λ=0.3, μ=1, α=0.05, β=3, C_s=1000, C_h=5, C_{sp} = 100, C_i=60

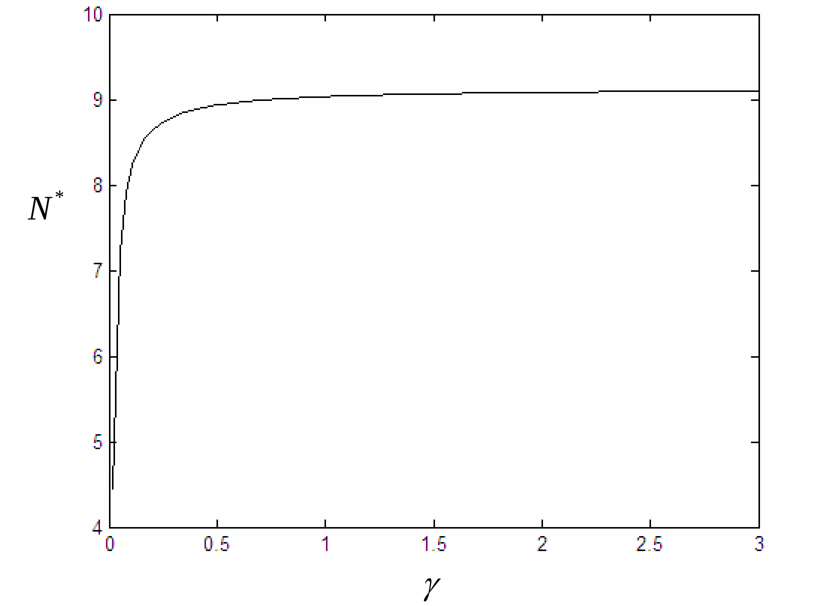


Figure 2.6 Plots of (γ, N^*) with λ=0.3, μ=1, α=0.05, β=3, C_s=500, C_h=5, C_{sp} = 100, C_i=40

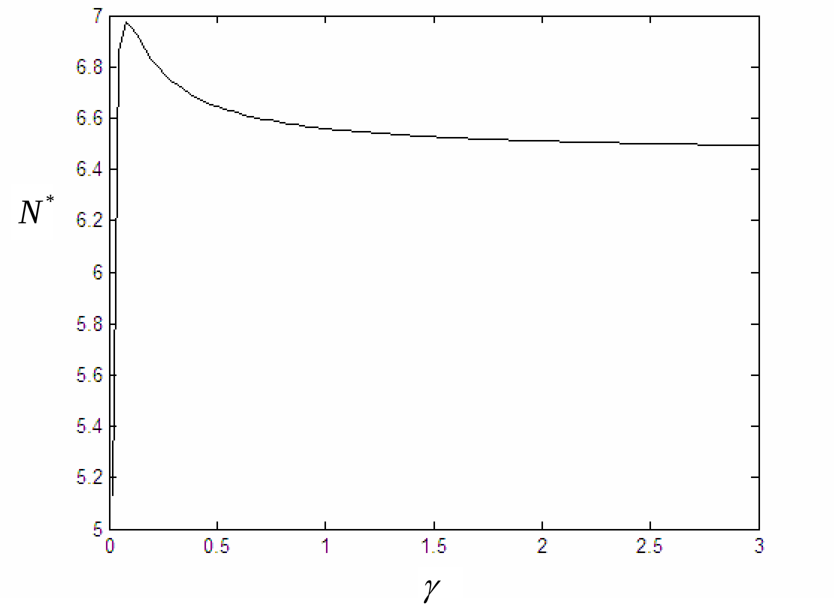


Table 2.3 The optimal N^* and minimum expected cost $F(N^*)$ with various values of γ.

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $C_s=1000$, $C_h=5$, $C_{sp}=100$, $C_i=60$								
γ	0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2
N^*	9	9	9	9	9	9	9	9
$F(N^*)$	85.4241	85.2503	85.2181	85.2068	85.2016	85.1987	85.1970	85.1959

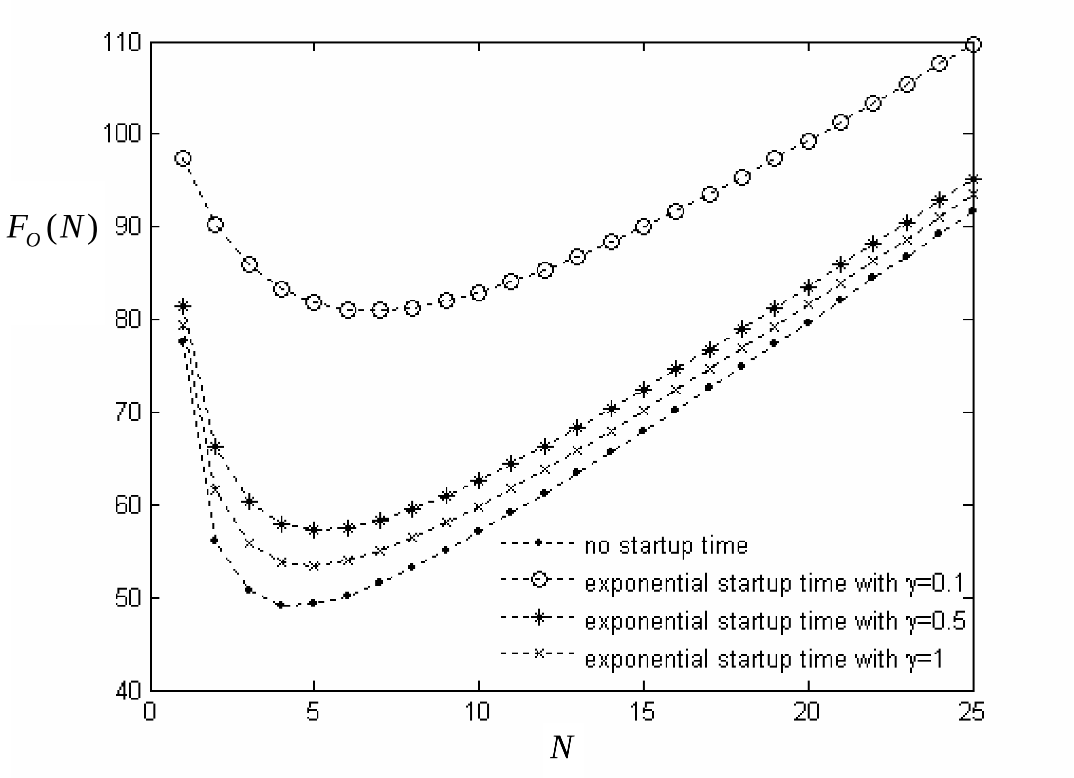
Table 2.4 The optimal N^* and minimum expected cost $F(N^*)$ with various values of (C_s, C_h).

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $\gamma=3$, $C_{sp}=100$, $C_i=60$								
(C_s, C_h)	(1000, 5)	(1000, 10)	(1000, 15)	(1000, 20)	(400, 10)	(600,10)	(800, 10)	(900,10)
N^*	9	6	5	5	4	5	6	6
$F(N^*)$	85.1964	101.9221	114.0319	124.3333	78.3476	87.3775	95.0861	98.5041

Table 2.5 The optimal N^* and minimum expected cost $F(N^*)$ with various values of (C_{sp}, C_i).

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $\gamma=3$, $C_s=1000$, $C_h=5$								
(C_{sp}, C_i)	(80, 20)	(80, 30)	(80, 40)	(80, 50)	(35, 25)	(45, 25)	(55 , 25)	(65, 25)
N^*	9	9	9	9	9	9	9	9
$F(N^*)$	57.5492	64.4228	71.2964	78.1701	60.6423	60.7187	60.7951	60.8714

Figure 2.7 The total expected cost $F_O(N)$ for different values of γ and N.



Chapter 3

Maximum Entropy Analysis to the *N* Policy M/G/1 Queue with Server Breakdowns and Startup Times

We study a single removable and unreliable server in the N policy M/G/1 queue with general startup times where arrivals form a Poisson process and service times are generally distributed. When N customers are accumulated in the system, the server is immediately turned on but is temporarily unavailable to the waiting customers. He needs a startup time before providing service until the system becomes empty. The server is subject to breakdowns according to a Poisson process and his repair time obeys an arbitrary distribution. We use the maximum entropy principle to derive the approximate formulas for the steady-state probability distributions of the queue length. We perform a comparative analysis between the approximate results with established exact results for various distributions, such as exponential (M), k-stage Erlang (E_k), and deterministic (D). We demonstrate that the maximum entropy approach is accurate enough for practical purposes and is a useful method for solving complex queueing systems.

3.1 Assumptions and Notations

It is assumed that customers arrive according to a Poisson process with parameter λ and service times are independent and identically distributed (i.i.d.) random variables having a general distribution function, $F_S(t)$ $(t \geq 0)$ with a mean service time μ_S and a finite variance σ_S^2. The server is subject to breakdowns at any time with Poisson breakdown rate α when he is turned on and working. When the server fails, he is immediately repaired at a repair facility, where the repair times are independent and identically distributed random variables obeying a general distribution function $F_R(t)$ $(t \geq 0)$ with a mean repair time μ_R and a finite variance σ_R^2. Arriving-customers form a single waiting line based on the first-come, first-served (FCFS) discipline. The server can serve only one customer at a time and the service is independent of the arrival of the customers. A customer who arrives and finds the server busy or broken down must wait in the queue until a server is available. Although no service occurs during the repair period of the server, customers continue to arrive following a Poisson process. Furthermore, when the number of customers in the system reaches a specific level, denoted by N, the server is immediately turned on (i.e. begin startup) but is temporarily unavailable to the waiting customers. He requires a startup time with random length before starting service. The startup times are independent and identically distributed random variables obeying a general distribution function $F_U(t)$ ($t \geq 0$) with a mean startup time μ_U and a finite variance σ_U^2. Once the startup is over, the server begins serving the waiting customers until there are no customers in the system. Service is allowed to be interrupted if the server breaks down, and the server is immediately repaired. Once the server is repaired, he immediately returns to serve customers until the system becomes empty.

The following notations and probabilities are used throughout this chapter.

N – threshold

S – service time random variable

U – startup time random variable

R – repair time random variable.

$F_S(\cdot)$ – distribution function of S

$F_U(\cdot)$ – distribution function of U

$F_R(\cdot)$ – distribution function of R

$G_O(z)$ – p.g.f. of the number of customers in the ordinary M/G/1 queue with reliable server

$G(z)$ – p.g.f. of the number of customers in the ordinary M/G/1 queue with unreliable server

$G_N(z)$ – p.g.f. of the number of customers in the N policy M/G/1 queue with server breakdowns and general startup times

$W(z)$ – p.g.f. of the number of customers that arrive during the turned-off plus the startup period

$\overline{f}_U(\cdot)$ – Laplace-Stieltjes transform (LST) of startup time

L_N – expected number of customers in the N policy M/G/1 queue with server breakdowns and general startup times

$P_{0,I}(n)$ – probability that there are n customers in the system when the server is turned off, where $n = 0, 1, 2, \cdots, N-1$

$P_{0,U}(n)$ – probability that there are n customers in the system when the server is startup, where $n = N, N+1, N+2, \cdots$

$P_1(n)$ – probability that there are n customers in the system when the server is turned on and working, where $n = 1, 2, 3, \cdots$

$P_2(n)$ – probability that there are n customers in the system when the server is in operation but found to be broken down, where $n = 1, 2, 3, \cdots$

W_q – exact expected waiting time in the queue

I_S – idle state

U_S – startup state

B_S – busy state

R_S – repair state

U_r – remaining startup time for the server begin startup

W_q^* – approximate waiting time in the queue

3.2 The Expected Number of Customers in the System

Let $G_O(z)$ denote the probability generating function (p.g.f.) of the number of customers in the ordinary M/G/1 queue with reliable server. From Kleinrock [28, p. 194], we have

$$G_O(z) = \frac{(1-\rho)(1-z)\,\bar{f}_S(\lambda - \lambda z)}{\bar{f}_S(\lambda - \lambda z) - z}, \qquad (3.1)$$

where $\rho = \lambda \mu_S$ and $\bar{f}_S(\cdot)$ is the Laplace-Stieltjes transform (abbreviated LST) of service time.

Let H be a random variable representing the completion time of a customer, which includes both the service time of a customer and the repair time of a server. Applying the well-known formula for the p.g.f. of the number of customers in the ordinary M/G/1 queue with reliable server, the p.g.f. of the number of customers in the ordinary M/G/1 queue with unreliable server is given by

$$G(z)=\frac{(1-\rho_H)(1-z)\overline{f}_H(\lambda-\lambda z)}{\overline{f}_H(\lambda-\lambda z)-z}, \tag{3.2}$$

where $\rho_H=\lambda E[H]$ ($E[H]$ is the mean completion time) and $\overline{f}_H(\cdot)$ is the LST of completion time. Note that ρ_H is traffic intensity and it should be assumed to be less than unity. It should be noted that expression (3.2) is obtained only by replacing service times by completion times in the formula of the ordinary M/G/1 queue with a reliable server.

We consider that the server is on "extended vacation" during the turned-off period I_N and startup period U_N, the lengths of which equal (I_N+U_N). Following the result of Medhi and Templeton [34], we obtain

$$G_N(z)=\frac{\left[1-W(z)\right]G(z)}{W'(1)(1-z)}, \tag{3.3}$$

where

$G_N(z)\equiv$ the p.g.f. of number of customers in the N policy M/G/1 queue system with server breakdowns and general startup times;

$W(z)\equiv$ the p.g.f. of the number of customers that arrive during a vacation of length I_N+U_N;

$\equiv$ [the p.g.f. of the number of customers that arrive during I_N]$\times$[the p.g.f. of the number of customers that arrive during U_N];

$\equiv$ $z^N\overline{f}_U(\lambda-\lambda z)$, where $\overline{f}_U(\cdot)$ is the LST of startup time.

We have

$$W'(z)=Nz^{N-1}\overline{f}_U(\lambda-\lambda z)+z^{N-1}\overline{f}_U^{(1)}(\lambda-\lambda z)(-\lambda).$$

It follows $W'(1)=N+\lambda\mu_U=N+\rho_U$, where $\rho_U=\lambda\mu_U$. From (3.2) and (3.3), we obtain

$$G_N(z)=\frac{\left[1-z^N\overline{f}_U(\lambda-\lambda z)\right](1-\rho_H)\overline{f}_H(\lambda-\lambda z)}{(N+\rho_U)\left[\overline{f}_H(\lambda-\lambda z)-z\right]}.$$

Let L_N denote the expected number of customers in the N policy M/G/1 queue with server breakdowns and general startup times. Thus we have

$$L_N=G_N'(z)\big|_{z=1}$$

$$= \frac{1}{(N+\rho_U)}\left[\frac{N(N-1)}{2} + N\rho_U + \frac{\lambda E[U^2]}{2}\right] + \rho_H + \frac{\lambda^2 E[H^2]}{2(1-\rho_H)}, \qquad (3.4)$$

where $E[H] = \mu_S(1+\alpha\mu_R)$ and $E[H^2] = (1+\alpha\mu_R)^2(\mu_S^2+\sigma_S^2) + \alpha\mu_S(\mu_R^2+\sigma_R^2)$.

3.3 The Maximum Entropy Results

In this section, we will develop the maximum entropy solutions for the steady-state probabilities of the N policy M/G/1 queue with server breakdowns and general startup times. Let us define

$P_{0,I}(n) \equiv$ probability that there are n customers in the system when the server is turned off, where $n = 0, 1, 2, \cdots, N-1$.

$P_{0,U}(n) \equiv$ probability that there are n customers in the system when the server is startup, where $n = N, N+1$, N+2, $\cdots$.

$P_1(n) \equiv$ probability that there are n customers in the system when the server is turned on and working, where $n = 1, 2, 3, \cdots$.

$P_2(n) \equiv$ probability that there are n customers in the system when the server is in operation but found to be broken down, where $n = 1, 2, 3, \cdots$.

In order to derive the steady-state probabilities $P_{0,I}(n)$, $P_{0,U}(n)$ and $P_i(n)$ (i = 1, 2) by using the maximum entropy principle, we formulate the maximum entropy model in the following. Following El-Affendi and Kouvatsos [12], the entropy function Y can be illustrated mathematically as

$$Y = -\sum_{n=0}^{N-1} P_{0,I}(n)\ln P_{0,I}(n) - \sum_{n=N}^{\infty} P_{0,U}(n)\ln P_{0,U}(n) - \sum_{n=1}^{\infty} P_1(n)\ln P_1(n) - \sum_{n=1}^{\infty} P_2(n)\ln P_2(n),$$

or equivalently

$$Y = -N\ln P_{0,I}(0) - \sum_{n=N}^{\infty} P_{0,U}(n)\ln P_{0,U}(n) - \sum_{n=1}^{\infty} P_1(n)\ln P_1(n) - \sum_{n=1}^{\infty} P_2(n)\ln P_2(n). \qquad (3.5)$$

There are five basic known results from the literature (see [8] and [54]) that facilitate the application of the maximum entropy formalism to study the N policy

M/G/1 queue with server breakdowns and general startup times. The maximum entropy solutions are obtained by maximizing (3.5) subject to the following five constraints, written as,

(i) normalizing condition

$$NP_{0,I}(0)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_{1}(n)+\sum_{n=1}^{\infty}P_{2}(n)=1, \tag{3.6}$$

(ii) the probability that the server is startup

$$\sum_{n=N}^{\infty}P_{0,U}(n)=\frac{\rho_U}{N+\rho_U}\left[1-\rho(1+\alpha\mu_R)\right]=\rho_U\Theta_1(1-\rho_H), \tag{3.7}$$

where $\Theta_1=\dfrac{1}{N+\rho_U}$ and $\rho_H=\rho(1+\alpha\mu_R)$,

(iii) the probability that the server is busy

$$\sum_{n=1}^{\infty}P_{1}(n)=\rho, \tag{3.8}$$

(iv) the probability that the server is broken down

$$\sum_{n=1}^{\infty}P_{2}(n)=\rho\alpha\mu_R, \tag{3.9}$$

(v) the expected number of customers in the system

$$\sum_{n=0}^{N-1}P_{0,I}(n)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_{1}(n)+\sum_{n=1}^{\infty}P_{2}(n)=L_N, \tag{3.10}$$

where L_N is given by (3.4).

It yields from (3.6) to (3.9)

$$P_{0,I}(0)=P_{0,I}(n)=\Theta_1(1-\rho_H),\qquad n=1,2,\cdots,N-1. \tag{3.11}$$

In (3.6)-(3.10), (3.6) is multiplied by τ_1, (3.7) is multiplied by τ_2, (3.8) is multiplied by τ_3, (3.9) is multiplied by τ_4 and (3.10) is multiplied by τ_5. Thus the Lagrangian function y is given by

$$y=-N\ln P_{0,I}(0)-\sum_{n=N}^{\infty}P_{0,U}(n)\ln P_{0,U}(n)-\sum_{n=1}^{\infty}P_{1}(n)\ln P_{1}(n)-\sum_{n=1}^{\infty}P_{2}(n)\ln P_{2}(n)$$
$$-\tau_1\left[NP_{0,I}(0)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_{1}(n)+\sum_{n=1}^{\infty}P_{2}(n)-1\right]$$

$$-\tau_2\left[\sum_{n=N}^{\infty}P_{0,U}(n)-\rho_U\Theta_1(1-\rho_H)\right]-\tau_3\left[\sum_{n=1}^{\infty}P_1(n)-\rho\right]-\tau_4\left[\sum_{n=1}^{\infty}P_2(n)-\rho\alpha\mu_R\right]$$

$$-\tau_5\left[\frac{N(N-1)}{2}P_{0,I}(0)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_1(n)+\sum_{n=1}^{\infty}P_2(n)-L_N\right], \qquad (3.12)$$

where τ_1 to τ_5 are the Lagrangian multipliers corresponding to constraints (3.6)-(3.10), respectively.

3.3.1 The maximum entropy solutions

To get the maximum entropy solutions, $P_{0,U}(n)$, $P_1(n)$, $P_2(n)$, maximizing in (3.5) subject to constraints (3.6)-(3.10) is equivalent to maximizing (3.12).

The maximum entropy solutions are obtained by taking the partial derivatives of y with respect to $P_{0,I}(n)$, $P_{0,U}(n)$ and $P_i(n)$ (i = 1, 2), and setting the results equal to zero, namely,

$$\frac{\partial y}{\partial P_{0,I}(0)}=-N\ln P_{0,I}(0)-N-N\tau_1-\frac{N(N-1)}{2}\tau_5=0, \qquad (3.13)$$

$$\frac{\partial y}{\partial P_{0,U}(n)}=-N\ln P_{0,U}(n)-1-\tau_1-\tau_2-n\tau_5=0, \qquad (3.14)$$

$$\frac{\partial y}{\partial P_1(n)}=-\ln P_1(n)-1-\tau_1-\tau_3-n\tau_5=0, \qquad (3.15)$$

$$\frac{\partial y}{\partial P_2(n)}=-\ln P_2(n)-1-\tau_1-\tau_4-n\tau_5=0. \qquad (3.16)$$

It implies from (3.13)-(3.16) that we obtain

$$P_{0,I}(0)=P_{0,I}(n)=e^{-(1+\tau_1)}e^{-\frac{(N-1)}{2}\tau_5},\quad n=1,2,\cdots,N-1, \qquad (3.17)$$

$$P_{0,U}(n)=e^{-(1+\tau_1+\tau_2)}e^{-n\tau_5},\quad n=N,N+1,\cdots \qquad (3.18)$$

$$P_1(n)=e^{-(1+\tau_1+\tau_3)}e^{-n\tau_5},\quad n=1,2,\ \cdots \qquad (3.19)$$

$$P_2(n)=e^{-(1+\tau_1+\tau_4)}e^{-n\tau_5},\quad n=1,2,\ \cdots. \qquad (3.20)$$

Let $\phi_1=e^{-(1+\tau_1)}$, $\phi_2=e^{-\tau_2}$, $\phi_3=e^{-\tau_3}$, $\phi_4=e^{-\tau_4}$ and $\phi_5=e^{-\tau_5}$. We transform

(3.17)-(3.20) in terms ϕ_1, ϕ_2, ϕ_3, ϕ_4 and ϕ_5 given by

$$P_{0,I}(n) = \Theta_1(1-\rho_{\mathrm{H}}), \quad n = 0, 1, 2, \cdots, N-1 \tag{3.21}$$

$$P_{0,U}(n) = \phi_1\phi_2\phi_5^n, \quad n = N, N+1, \cdots \tag{3.22}$$

$$P_1(n) = \phi_1\phi_3\phi_5^n, \quad n = 1, 2, \cdots \tag{3.23}$$

$$P_2(n) = \phi_1\phi_4\phi_5^n, \quad n = 1, 2, \cdots. \tag{3.24}$$

Substituting (3.22)-(3.24) into (3.7)-(3.9), respectively, yields

$$\phi_1\phi_2 = \frac{\rho_U\Theta_1(1-\rho_H)(1-\phi_5)}{\phi_5^N}, \tag{3.25}$$

$$\phi_1\phi_3 = \frac{\rho(1-\phi_5)}{\phi_5}, \tag{3.26}$$

$$\phi_1\phi_4 = \frac{\rho\alpha\mu_R(1-\phi_5)}{\phi_5}. \tag{3.27}$$

Substituting (3.11) and (3.22)-(3.24) into (3.10) and doing the algebraic manipulations, we obtain

$$\phi_5 = 1 - \frac{\rho_H + \rho_U\Theta_1(1-\rho_H)}{L_N - \Theta_1(N-1)(1-\rho_H)\left(\frac{N}{2}+\rho_U\right)}. \tag{3.28}$$

Finally, we get

$$P_{0,I}(n) = \Theta_1(1-\rho_{\mathrm{H}}), \quad n = 0, 1, 2, \cdots, N-1, \tag{3.29}$$

$$P_{0,U}(n) = \rho_U\Theta_1(1-\rho_H)(1-\phi_5)\phi_5^{n-N}, \quad n = N, N+1, \cdots \tag{3.30}$$

$$P_1(n) = \rho(1-\phi_5)\phi_5^{n-1}, \quad n = 1, 2, \cdots \tag{3.31}$$

$$P_2(n) = \rho\alpha\mu_R)(1-\phi_5)\phi_5^{n-1}, \quad n = 1, 2, \cdots. \tag{3.32}$$

3.4 The Exact and Approximate Expected Waiting Time in the Queue

In this section, we develop the exact and the approximate formulae for the expected waiting time in the N policy M/G/1 queue with server breakdowns and general startup times as follows.

3.4.1 The exact expected waiting time in the queue

Let W_q denote the exact expected waiting time in the queue. Using (3.4) and Little's formula, we obtain

$$W_q = \frac{L_N}{\lambda} - E[H] = \Theta_1\left[\frac{N(N-1)}{2\lambda} + N\mu_U + \frac{\lambda E[U^2]}{2}\right] + \frac{\lambda E[H^2]}{2(1-\rho_H)}. \qquad (3.33)$$

3.4.2 The approximate expected waiting time in the queue

We define the idle state, the startup state, the busy state, and the repair state as follows:

(i) Idle state denoted by I_S: the server is turned off and the number of customers waiting in the system is less than or equal to $N-1$.

(ii) Startup state denoted by U_S: the server begins startup and the number of customers waiting in the system is greater than or equal to N.

(iii) Busy state denoted by B_S: the server is busy and provides service to a customer.

(iv) Repair state denoted by R_S: the server is broken down and being repaired.

Following Borthakur et al. [8], we find the expected waiting time of customer C at the states I_S, U_S, B_S and R_S as follows. Suppose that a 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_S, U_S, B_S and R_S are described, respectively, as follows:

(i) In idle state I_S: The server will begin startup after $(N-n-1)$ customers arrive in the system. Thus customer C will be served until $(N-n-1)$ customers arrive and n customers in front of him waiting for service. The expected waiting time of customer C at the idle state is

$$\frac{(N-n-1)}{\lambda} + \mu_U + n\mu_S .$$

(ii) In startup state U_S: We derive the expected waiting time of customer C at the startup state in the following. Let us define

$U_r \equiv$ remaining startup time for the server begin startup.

Following Borthakur et al. [8], the cumulative distribution function (c.d.f.) of $F_{U_r}(t)$ is given by

$$F_{U_r}(t) = P\{U_r \le t\} = \frac{1}{\mu_U}\int_0^t [1 - F_U(x)]\, dx,$$

where $F_U(x)$ is the c.d.f of startup time. Thus, we get the probability density function of remaining startup time for the server startup as

$$f_{U_r}(t) = F'_{U_r}(t) = \frac{1}{\mu_U}[1 - F_U(t)].$$

Let $E[U_r]$ be the mean remaining startup time.

$$E[U_r] = \int_0^\infty t\ f_{U_r}(t)dt = \int_0^\infty t\frac{1}{\mu_U}[1 - F_U(t)]dt = \frac{1}{\mu_U}\int_0^\infty t[1 - F_U(t)]dt$$
$$= \frac{1}{\mu_U}\int_0^\infty t[1 - F_U(t)]dt = \frac{1}{\mu_U}\left\{\frac{t^2}{2}[1 - F_U(t)]\Big|_{t=0}^{\infty} + \int_0^\infty \frac{t^2}{2} f_U(t)\, dt\right\}$$
$$= \frac{E[U^2]}{2\mu_U}.$$

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 $n\mu_S + E[U^2]/2\mu_U$.

(iii) In busy state B_S: Since the server is turned on and working, customer C only waits n customers in front of him to be served. The expected waiting time of customer C at the busy state is $n\mu_S$.

(iv) In repair state and R_S: Using the same argument as (ii), we have the expected waiting time of customer C at the repair state is $n\mu_S + E[R^2]/2\mu_R$.

Finally, using the listed above results, we obtain the approximate expected waiting time in the queue given by

$$W_q^* = \sum_{n=0}^{N-1}\left(\frac{N-n-1}{\lambda} + \mu_U + n\mu_S\right)P_{0,I}(0) + \sum_{n=N}^{\infty}\left(n\mu_S + \frac{E[U^2]}{2\mu_U}\right)P_{0,U}(n)$$
$$+\sum_{n=1}^{\infty} n\mu_S P_1(n) + \sum_{n=1}^{\infty}\left(n\mu_S + \frac{E[R^2]}{2\mu_R}\right)P_2(n). \qquad (3.34)$$

where $P_{0,I}(0)$, $P_{0,U}(n)$, $P_1(n)$, and $P_2(n)$ are given in (3.29)-(3.32), respectively.

3.5 Comparative Analysis

The primary objective of this section is to examine the accuracy of the maximum entropy results. We present specific numerical comparisons between the exact results and the maximum entropy (approximate) results for the N policy M/G/1 queue with general service times, general repair times and general startup times. Conveniently, we represent this queueing system as the N policy M/G(G,G)/l queue where the second, third, fourth symbols denote the general distribution of service time, repair time, and startup time, respectively.

This section includes the following three subsections:

(i) Comparative analysis for the N policy M/M(M,M)/1 and M/D(D,D)/1 queues.

(ii) Comparative analysis for the N policy M/E_3(E_4,E_3)/1 and M/M(E_3,E_2)/1 queues.

(iii) Comparative analysis for the N policy M/E_3(E_4,D)/1 and M/E_3(E_4,M)/1 queues.

3.5.1 Comparative analysis for the N policy **M/M(M,M)/1** *and* **M/D(D,D)/1** *queues*

Here we perform a comparative analysis between the exact W_q and the approximate (maximum entropy) W_q^* for the N policy M/M(M,M)/1 and M/D(D, D)/1 queues. For the N policy M/M(M,M)/1 queue, we obtain $\mu_S = 1/\mu$, $E[S^2] = 2/\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 2/\beta^2$, $\mu_U = 1/\gamma$ and $E[U^2] = 2/\gamma^2$. For the N policy M/D(D,D)/1 queue, we have $\mu_S = 1/\mu$, $E[S^2] = 1/\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 1/\beta^2$, $\mu_U = 1/\gamma$ and $E[U^2] = 1/\gamma^2$.

We set N =5 and N =10,and choose the various values of λ, μ, α, β, and γ. The numerical results are obtained by considering the following parameters:

Case 1: We fix μ=1.0, α=0.05, β=3.0, γ=3.0, and vary the values of λ from 0.2 to 0.8.

Case 2: We fix λ=0.3, α=0.05, β=3.0, γ=3.0, and vary the values of μ from 0.5 to 2.0.

Case 3: We fix λ=0.3, μ=1.0, β=3.0, γ=3.0, and vary the values of α from 0.05 to 0.2.

Case 4: We fix λ=0.3, μ=1.0, α=0.05, γ=3.0, and vary the values of β from 2.0 to 6.0.

Case 5: We fix λ =0.3, μ =1.0, α =0.05, β =3.0, and vary the values of γ from 2.0 to 5.0.

Numerical results for the N policy M/M(M,M)/1 and M/D(D,D)/1 queues are shown in Table 3.1 for the above five cases. The relative error percentages are very small (0-6.8%).

3.5.2 Comparative analysis for the N policy **$M/E_3(E_4,E_3)/1$** *and* **$M/M(E_3,E_2)/1$** *queues*

Here we perform a comparative analysis between the exact W_q and the approximate (maximum entropy) W_q^* for the N policy $M/E_3(E_4,E_3)/1$ and $M/M(E_3, E_2)/1$ queues. For the N policy $M/E_3(E_4,E_3)/1$ queue, we have $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 5/4\beta^2$, $\mu_U = 1/\gamma$, and $E[U^2] = 4/3\gamma^2$. For the N policy $M/M(E_3,E_2)/1$ queue, we get $\mu_S = 1/\mu$, $E[S^2] = 2/\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 4/3\beta^2$, $\mu_U = 1/\gamma$ and $E[U^2] = 3/2\gamma^2$.

Numerical results for the the N policy $M/E_3(E_4,E_3)/1$ and $M/M(E_3,E_2)/1$ queues are shown in Table 3.2 for the above five cases. The relative error percentages are also very small (0-3.5%).

3.5.3 Comparative analysis for the N policy **$M/E_3(E_4,D)/1$** *and* **$M/E_3(E_4,M)/1$** *queues*

Here we perform a comparative analysis between the exact W_q and the approximate (maximum entropy) W_q^* for the N policy $M/E_3(E_4,D)/1$ and $M/E_3(E_4, M)/1$ queues. For the N policy $M/E_3(E_4,D)/1$ queue, we get $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 5/4\beta^2$, $\mu_U = 1/\gamma$, and $E[U^2] = 1/\gamma^2$. For the N policy $M/E_3(E_4,M)/1$ queue, we obtain $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $\mu_R = 1/\beta$, $E[R^2] = 5/4\beta^2$, $\mu_U = 1/\gamma$, and $E[U^2] = 2/\gamma^2$.

Numerical results for the N policy $M/E_3(E_4,D)/1$ and $M/E_3(E_4,M)/1$ queues are shown in Table 3.3 for the above five cases. Again, the relative error percentages are very small (0-3.5%).

Table 3.1. Comparison of exact W_q and approximate W_q^* for the N policy M/M(M,M)/1 and M/D(D,D)/1 queues.

	M/M(M,M)/1						M/D(D,D)/1					
	$N=5$			$N=10$			$N=5$			$N=10$		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	10.4626	10.4244	0.3657	22.9452	22.8653	0.3481	10.3300	10.3955	0.6345	22.8137	22.8376	0.1049
0.4	5.9040	5.8579	0.7815	12.1359	12.0482	0.7225	5.5494	5.7124	2.9363	11.7834	11.9048	1.0302
0.6	5.1372	5.0756	1.1979	9.2850	9.1820	1.1095	4.3314	4.5880	5.9242	8.4824	8.6975	2.5358
0.8	7.1603	7.0513	1.5226	10.2658	10.1154	1.4654	4.9251	5.2593	6.7861	8.0347	8.3274	3.6437
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	10.0582	9.9373	1.2022	18.3737	18.1696	1.1107	8.4605	8.9757	6.0898	16.7776	17.2097	2.5752
1.0	7.3178	7.2762	0.5695	15.6334	15.5501	0.5325	7.0903	7.2048	1.6150	15.4074	15.4804	0.4733
1.5	7.0437	7.0179	0.3654	15.3592	15.3057	0.3480	6.9532	6.9967	0.6252	15.2704	15.2861	0.1031
2.0	6.9617	6.9431	0.2683	15.2773	15.2378	0.2583	6.9122	6.9324	0.2924	15.2294	15.2288	0.0038
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	7.3178	7.2762	0.5695	15.6334	15.5501	0.5325	7.0903	7.2048	1.6150	15.4074	15.4804	0.4733
0.10	7.3384	7.2546	1.1408	15.6539	15.4870	1.0660	7.1006	7.1794	1.1102	15.4177	15.4134	0.0282
0.15	7.3594	7.2333	1.7140	15.6750	15.4241	1.6005	7.1111	7.1540	0.6041	15.4282	15.3464	0.5302
0.20	7.3810	7.2121	2.2890	15.6966	15.3613	2.1359	7.1219	7.1288	0.0968	15.4391	15.2796	1.0328
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	7.3298	7.2672	0.8549	15.6454	15.5204	0.7991	7.0963	7.1930	1.3625	15.4134	15.4478	0.2226
3.0	7.3178	7.2762	0.5695	15.6334	15.5501	0.5325	7.0903	7.2048	1.6150	15.4074	15.4804	0.4733
4.0	7.3123	7.2811	0.4269	15.6279	15.5655	0.3993	7.0875	7.2109	1.7411	15.4047	15.4969	0.5986
6.0	7.3072	7.2864	0.2845	15.6227	15.5811	0.2661	7.0850	7.2172	1.8670	15.4021	15.5136	0.7238
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	7.4211	7.3789	0.5685	15.7269	15.6432	0.5323	7.1895	7.3035	1.5858	15.4989	15.5714	0.4675
3.0	7.3178	7.2762	0.5695	15.6334	15.5501	0.5325	7.0903	7.2048	1.6150	15.4074	15.4804	0.4733
4.0	7.2667	7.2253	0.5700	15.5869	15.5039	0.5326	7.0406	7.1553	1.6299	15.3617	15.4348	0.4762
5.0	7.2362	7.1949	0.5702	15.5591	15.4762	0.5327	7.0107	7.1256	1.6390	15.3342	15.4075	0.4779

Table 3.2. Comparison of exact W_q and approximate W_q^* for the N policy M/E_3(E_4,E_3)/1 and M/M(E_3,E_2)/1 queues.

	M/E3(E4,E3)/1						M/M(E3,E2)/1					
	$N=5$			$N=10$			$N=5$			$N=10$		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0$, $\alpha=0.05$, $\beta=3.0$, $\gamma=3.0$)											
0.2	10.3742	10.4051	0.2982	22.8575	22.8468	0.0467	10.4611	10.4228	0.3657	22.9442	22.8643	0.3481
0.4	5.6675	5.7607	1.6454	11.9007	11.9524	0.4344	5.9006	5.8545	0.7815	12.1335	12.0459	0.7225
0.6	4.5996	4.7502	3.2733	8.7496	8.8586	1.2465	5.1311	5.0696	1.1981	9.2805	9.1775	1.1095
0.8	5.6692	5.8557	3.2895	8.7774	8.9224	1.6524	7.1482	7.0393	1.5230	10.2557	10.1054	1.4655
μ	Case 2. ($\lambda=0.3$, $\alpha=0.05$, $\beta=3.0$, $\gamma=3.0$)											
0.5	8.9927	9.2959	3.3715	17.3093	17.5293	1.2711	10.0537	9.9328	1.2022	18.3700	18.1660	1.1107
1.0	7.1660	7.2285	0.8714	15.4827	15.5035	0.1348	7.3154	7.2737	0.5695	15.6318	15.5485	0.5325
1.5	6.9833	7.0037	0.2921	15.2999	15.2926	0.0478	7.0416	7.0158	0.3654	15.3579	15.3045	0.3480
2.0	6.9287	6.9359	0.1046	15.2453	15.2318	0.0888	6.9598	6.9411	0.2683	15.2761	15.2367	0.2583
α	Case 3. ($\lambda=0.3$, $\mu=1.0$, $\beta=3.0$, $\gamma=3.0$)											
0.05	7.1660	7.2285	0.8714	15.4827	15.5035	0.1348	7.3154	7.2737	0.5695	15.6318	15.5485	0.5325
0.10	7.1796	7.2043	0.3433	15.4962	15.4377	0.3776	7.3351	7.2514	1.1409	15.6515	15.4846	1.0660
0.15	7.1936	7.1802	0.1863	15.5102	15.3720	0.8907	7.3554	7.2293	1.7141	15.6717	15.4209	1.6005
0.20	7.2079	7.1562	0.7175	15.5245	15.3064	1.4045	7.3761	7.2073	2.2892	15.6925	15.3573	2.1359
β	Case 4. ($\lambda=0.3$, $\mu=1.0$, $\alpha=0.05$, $\gamma=3.0$)											
2.0	7.1739	7.2175	0.6073	15.4905	15.4717	0.1214	7.3264	7.2638	0.8549	15.6427	15.5177	0.7991
3.0	7.1660	7.2285	0.8714	15.4827	15.5035	0.1348	7.3154	7.2737	0.5695	15.6318	15.5485	0.5325
4.0	7.1624	7.2343	1.0033	15.4790	15.5197	0.2628	7.3103	7.2791	0.4269	15.6266	15.5642	0.3993
6.0	7.1590	7.2403	1.1350	15.4756	15.5361	0.3907	7.3054	7.2846	0.2845	15.6217	15.5801	0.2661
γ	Case 5. ($\lambda=0.3$, $\mu=1.0$, $\alpha=0.05$, $\beta=3.0$)											
2.0	7.2666	7.3285	0.8525	15.5748	15.5952	0.1310	7.4166	7.3744	0.5685	15.7242	15.6405	0.5323
3.0	7.1660	7.2285	0.8714	15.4827	15.5035	0.1348	7.3154	7.2737	0.5695	15.6318	15.5485	0.5325
4.0	7.1159	7.1786	0.8811	15.4367	15.4578	0.1367	7.2650	7.2236	0.5700	15.5856	15.5026	0.5326
5.0	7.0858	7.1486	0.8870	15.4091	15.4303	0.1378	7.2348	7.1935	0.5703	15.5580	15.4751	0.5327

Table 3.3. Comparison of exact W_q and approximate W_q^* for the N policy M/E_3(E_4,D)/1 and M/E_3(E_4,M)/1 queues.

	M/E_3(E_4,D)/1						M/E_3(E_4,M)/1					
	$N=5$			$N=10$			$N=5$			$N=10$		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	10.3734	10.4044	0.2983	22.8571	22.8464	0.0467	10.3756	10.4065	0.2981	22.8582	22.8475	0.0467
0.4	5.6660	5.7593	1.6460	11.9000	11.9517	0.4345	5.6703	5.7636	1.6443	11.9022	11.9539	0.4343
0.6	4.5975	4.7481	3.2753	8.7485	8.8575	1.2467	4.6039	4.7544	3.2693	8.7517	8.8608	1.2459
0.8	5.6664	5.8529	3.2917	8.7760	8.9210	1.6529	5.6748	5.8612	3.2849	8.7803	8.9253	1.6515
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	8.9916	9.2948	3.3720	17.3087	17.5288	1.2712	8.9949	9.2980	3.3704	17.3104	17.5304	1.2710
1.0	7.1650	7.2274	0.8716	15.4821	15.5030	0.1348	7.1682	7.2307	0.8710	15.4838	15.5046	0.1347
1.5	6.9822	7.0026	0.2922	15.2994	15.2920	0.0478	6.9855	7.0059	0.2919	15.3010	15.2937	0.0479
2.0	6.9276	6.9349	0.1047	15.2448	15.2312	0.0888	6.9309	6.9381	0.1045	15.2464	15.2329	0.0889
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	7.1650	7.2274	0.8716	15.4821	15.5030	0.1348	7.1682	7.2307	0.8710	15.4838	15.5046	0.1347
0.10	7.1785	7.2032	0.3435	15.4957	15.4372	0.3776	7.1818	7.2064	0.3429	15.4973	15.4388	0.3777
0.15	7.1925	7.1791	0.1861	15.5096	15.3715	0.8907	7.1957	7.1823	0.1867	15.5113	15.3731	0.8908
0.20	7.2068	7.1551	0.7173	15.5239	15.3059	1.4045	7.2100	7.1583	0.7178	15.5256	15.3075	1.4046
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	7.1728	7.2164	0.6076	15.4900	15.4712	0.1214	7.1761	7.2196	0.6069	15.4916	15.4728	0.1214
3.0	7.1650	7.2274	0.8716	15.4821	15.5030	0.1348	7.1682	7.2307	0.8710	15.4838	15.5046	0.1347
4.0	7.1613	7.2332	1.0035	15.4785	15.5192	0.2628	7.1646	7.2365	1.0029	15.4801	15.5208	0.2627
6.0	7.1579	7.2392	1.1352	15.4751	15.5355	0.3907	7.1612	7.2424	1.1346	15.4767	15.5372	0.3907
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	7.2642	7.3261	0.8529	15.5736	15.5940	0.1311	7.2714	7.3334	0.8516	15.5773	15.5977	0.1309
3.0	7.1650	7.2274	0.8716	15.4821	15.5030	0.1348	7.1682	7.2307	0.8710	15.4838	15.5046	0.1347
4.0	7.1152	7.1779	0.8812	15.4363	15.4574	0.1367	7.1171	7.1798	0.8809	15.4373	15.4584	0.1366
5.0	7.0854	7.1482	0.8871	15.4089	15.4301	0.1378	7.0866	7.1494	0.8868	15.4095	15.4307	0.1378

Chapter 4

Optimization of the *T* Policy M/G/1 Queue with Server Breakdowns and Startup Times

We consider a T policy M/G/1 queue in which the server is typically subject to unpredictable breakdowns. It is assumed that arriving customers follow a Poisson process and the breakdown times of the server follow the negative exponential distribution. We also assume that the service times, the repair times, and the startup times obey a general distribution. After a period of length T, the server is immediately turned on but is temporarily unavailable to serve waiting customers if there is at least one customer in the waiting line; otherwise, the server waits another period of length T and so on until at least one customer is present. When the server turns on, he requires for the preparatory work (i.e. begin startup) before starting service. Once the startup is terminated, the server immediately starts serving the waiting customers. We develop the probability generating function and various system performance measures such as the expected number of customers in the system, the expected length of the idle, busy, and breakdown period, and the expected length of the busy cycle, etc. Based on the derived results, we construct the total expected cost function per unit time, including customer holding cost, the system setup cost, server on and off costs, server startup cost, and server breakdown cost. We determine the optimal threshold T numerically to minimize the total expected cost. In addition, numerical results and sensitivity investigations are also presented

4.1 Assumptions and Notations

It is assumed that customers arrive according to a Poisson process with parameter λ. The service times of the customers are independent and identically distributed (i.i.d.) random variables obeying an arbitrary distribution function $F_S(t)$ ($t \geq 0$) with a finite mean μ_S and a finite variance σ_S^2. The server is subject to breakdowns at any time with Poisson breakdown rate α when he is working. When the server fails, he is immediately repaired at a repair facility, where the repair times are i.i.d. random variables having a general distribution function $F_R(t)$ ($t \geq 0$) with a finite mean μ_R and a finite variance σ_R^2. Arriving customers form a single waiting line at a server based on the order of their arrivals. The server can serve only one customer at a time and the service is independent of the arrival process. A customer who arrives and finds the server busy or broken down must wait in the queue until a server is available. Although no service occurs during the repair period of the server, customers continue to arrive following a Poisson process. Furthermore, if there is at least one customer in waiting line after a period of length T, the server is immediately turned on (i.e. begin startup) but is temporarily unavailable to serve the waiting customers. He needs a startup time with random length before starting service. Again, the startup times are i.i.d. random variables obeying a general distribution function $F_U(t)$ ($t \geq 0$) with a finite mean μ_U and a finite variance σ_U^2. Once the startup is terminated, the server begins serving the waiting customers until the system becomes empty. Service is allowed to be interrupted if the server breaks down, and the server is immediately repaired. Once the server is repaired, he immediately returns to serve customers until there are no customers in the system.

The following notations and probabilities are used throughout this chapter.

$T-$ threshold

$S-$ service time random variable

$U-$ startup time random variable

$R-$ repair time random variable

$F_S(\cdot)-$ distribution function of S

$F_U(\cdot)-$ distribution function of U

$F_R(\cdot)-$ distribution function of R

$A(t)-$ number of customers arriving into the system during $[0,t]$

A_m- arrival time of the *m*-th customer

$F_{A_m}(\cdot)-$ distribution function of A_m

$G_I(z)-$ p.g.f. of the number of customers waiting in the queue during an idle period

$G_U(z)$ p.g.f. of the number of customers arriving during a startup period

$\bar{f}_U(\cdot)-$ Laplace-Stieltjes transform (LST) of startup time

$W(z)-$ p.g.f. of the number of customers that arrive during the turned-off plus the startup period

$G(z)-$ p.g.f. of the number of customers in the ordinary M/G/1 queue with unreliable server

$G_T(z)-$ p.g.f. of the number of customers in the T policy M/G/1 queue with server breakdowns and general startup times

L_T- expected number of customers in the T policy M/G/1 queue with server breakdowns and general startup times

H_O- complete period of the ordinary M/G/1 queue with server breakdowns

I_T- turned-off period of the T policy M/G/1 queue with server breakdowns and general startup times

U_T- startup period of the T policy M/G/1 queue with server breakdowns and general startup times

B_T- busy period of the T policy M/G/1 queue with server breakdowns and general startup times

D_T- breakdown period of the T policy M/G/1 queue with server breakdowns and general startup times

H_T- complete period which is equal to (B_T+D_T)

V_T- complete startup period which is equal to (U_T+H_T)

C_T- busy cycle which is equal to (I_T+V_T)

$F_{V_T}(\cdot)-$ distribution function of V_T

$F_{H_O}(\cdot)-$ distribution function of H_O

$F_{H_O}^{(m+n)}(\cdot)-$ $(m+n)$-fold convolution of $F_{H_O}(\cdot)$

$\overline{f}_{V_T}(\cdot)-$ LST of V_T

$P_{I_T}-$ probability that the server is idle in the T policy M/G/1 queue with server breakdowns and general startup times

$P_{U_T}-$ probability that the server is startup in the T policy M/G/1 queue with server breakdowns and general startup times

$P_{B_T}-$ probability that the server is busy in the T policy M/G/1 queue with server breakdowns and general startup times

P_{D_T} – probability that the server is broken down in the T policy M/G/1 queue with server breakdowns and general startup times

C_h – holding cost per unit time for each customer present in the system;

C_s – setup cost for per busy cycle;

C_i – cost per unit time for keeping the server off;

C_{sp} – startup cost per unit time for the preparatory work of the server before starting the service;

C_b – cost per unit time for keeping the server on and in operation;

C_d – breakdown cost per unit time for a failed server

4.2 System Performance Measures

In this section, we focus mainly on developing some important system performance measures, such as (i) the expected number of customers in the system; (ii) the expected length of the idle period, the complete startup period, the busy period, and the breakdown period; (iii) the expected length of the busy cycle; and (iv) the probability that the server is idle, startup, busy and broken down.

4.2.1 Expected number of customers in the system

Let H be a random variable representing the completion time of a customer, which includes both the service time of a customer and the repair time of a server. Applying the well-known results of Medhi and Templeton [34], the probability generating function (p.g.f.) of the number of customers in the ordinary M/G/1 queue with server breakdowns is given by

$$G(z) = \frac{(1-\rho_H)(1-z)\overline{f}_H(\lambda - \lambda z)}{\overline{f}_H(\lambda - \lambda z) - z}, \tag{4.1}$$

where $\rho_H = \lambda E[H]$. In addition, $E\left[H\right] = \mu_S(1+\alpha\mu_R)$ and $E[H^2] = (1+\alpha\mu_R)^2 (\mu_S^2+\sigma_S^2) + \alpha\mu_S(\mu_R^2+\sigma_R^2)$. The traffic intensity ρ_H is assumed to be less than 1. We consider a Poisson arrival process. Let ξ_i denote the elapsed time between the (i-1)st and the i-th arriving customer. Following Ross [36], the time intervals ξ_i are i.i.d. exponential random variables with mean $1/\lambda$. Let $A(t)$ denote the number of customers arriving into the system during $[0,t]$. Let A_m be the arrival time of the

m-th customer and let $A_m = \sum_{i=1}^{m} \xi_i$. The distribution of A_m occurring by time t is given by

$$F_{A_m}(t) = P\{A_m \le t\} = \int_0^t \frac{\lambda(\lambda x)^{m-1}}{(m-1)!} e^{-\lambda x} dx = 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = P\{A(t) \ge m\}.$$

It is obvious that

$$P\{A(t) = m\} = F_{A_m}(t) - F_{A_{m+1}}(t),$$

where $\sum_a^b \cdot = 0$ for $b < a$.

Hence, in a period of length T and when there are at least $m\ (m \ge 1)$ customers in the system, the distribution of A_m is given

$$F_{A_m}(T) = P\{A_m \le T\} = 1 - \sum_{k=0}^{m-1} \frac{(\lambda T)^k}{k!} e^{-\lambda T}.$$

Let $G_I(z)$ be the p.g.f. of the number of customers waiting in the queue during idle period I_T. Thus we have

$$G_I(z) = \sum_{m=0}^{\infty} z^m \left[F_{A_m}(T) - F_{A_{m+1}}(T) \right] = e^{-(1-z)\lambda T}.$$

Let $G_U(z)$ be the p.g.f. of the number of customers arriving during startup period U_T. Then we get $G_U(z) = e^{-(1-z)\lambda t}$. The Laplace-Stieltjes transform (abbreviated LST) of $G_U(z)$ is given by

$$\int_0^{\infty} G_U(z)\, dF_U(t) = \overline{f}_U\left[\lambda(1-z)\right].$$

Because the Poisson process from any point on is independent of all that has previously occurred, we have $W(z) = G_I(z)\, \overline{f}_U[\lambda(1-z)] = e^{-(1-z)\lambda T}\, \overline{f}_U[\lambda(1-z)]$. For the T policy M/G/1 queue with server breakdowns and general startup times, we get the idle period plus the startup period as $W(z)$. We extending the well-known decomposition property concerning M/G/1 vacation queue studied by Fuhrmann and Cooper [14], we obtain the p.g.f. of number of customers in the T policy M/G/1 queue with server breakdowns and general startup times as follows:

$$G_T(z) = G(z)\left[\frac{1 - W(z)}{W'(1)(1-z)} \right], \tag{4.2}$$

where $G(z)$ is given in (4.1). Let L_T denote the expected number of customers in the T policy M/G/1 queue with server breakdowns and general startup times. Thus we obtain from (4.2) that

$$L_T = G_T'(z)\Big|_{z=1} = \frac{1}{(\lambda T + \rho_U)}\left[\frac{\lambda^2 T^2}{2} + \lambda\rho_U T + \frac{\lambda^2\sigma_U^2 + \rho_U^2}{2}\right] + \rho_H + \frac{\lambda^2 E[H^2]}{2(1-\rho_H)}, \qquad (4.3)$$

where $\rho_U = \lambda\mu_U$.

4.2.2 Expected length of the idle, complete startup, busy and breakdown periods

The idle period terminates when at least one customer arrives in system during the period T. Since the startup period starts when the idle period terminates, the complete startup period is represented by the sum of the startup period and the complete period. The server begins startup when there is at least one waiting customer at the end of the fixed period T in the system. This is called the startup period. The startup period terminates when the server starts to serve the waiting customers. Since the complete period begins when the startup period is over and terminates when the system becomes empty, the complete period is represented by the sum of the busy period and the breakdown period. The busy period is initiated when the server completes his startup and begins serving the waiting customers. During the busy period, the server may break down and starts his repair immediately. This is called the breakdown period. After the server is repaired, he returns immediately and provides service until the system is empty. Let H_O be the complete period of the ordinary M/G/1 queue with server breakdowns. Using the well-known result of Kleinrock [28, p. 213], we obtain the expected length of the complete period for the ordinary M/G/1 queue with server breakdowns as

$$E[H_O] = \frac{E[H]}{1-\lambda E[H]} = \frac{\mu_S(1+\alpha\mu_R)}{1-\rho_H}. \qquad (4.4)$$

4.2.2.1 Expected length of the idle period

The idle period I_T terminates when at least one customer arrives during the period T. It is obvious that

$$E[I_T] = T. \qquad (4.5)$$

4.2.2.2 Expected length of the complete startup period

Let V_T represent the complete startup period for the T policy M/G/1 queue with server breakdowns and general startup times. Thus we have $V_T = H_T + U_T$, where H_T and U_T denote the complete period and the startup period, respectively. Let $\overline{f}_{V_T}(\cdot)$ be the LST of the distribution of the complete startup period of the ordinary M/G/1 queue with server breakdowns.

The following notations are used.

$F_{V_T}(\cdot)$ − distribution function of the complete startup period V_T of the T policy M/G/1 queue with server breakdowns and general startup times;

$\overline{f}_U(\cdot)$ − the LST of startup times;

$F_{H_O}(\cdot)$ − distribution function of the complete period H_O of the ordinary M/G/1 queue with server breakdowns;

$F_{H_O}^{(m+n)}(\cdot)$ − $(m+n)$-fold convolution of $F_{H_O}(\cdot)$.

By conditioning V_T on the length of the first startup time $= t$ and the number of arrivals during U, we obtain from Gross and Harris [18 , p. 277] that

$$\begin{aligned} F_{V_T/A=m}(x) &= \int_0^x \sum_{n=0}^{\infty} P(given\ any\ startup\ time = t, complete\ startup\ period \\ &\quad generated\ by\ m\ customers\ arrival\ plus\ n\ customers \\ &\quad arrival\ in\ the\ complete\ period\ H_O\ during\ t \le x-t)\, dF_U(t) \\ &= \int_0^x \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} F_{H_O}^{(m+n)}(x-t)\, dF_U(t)\,. \end{aligned} \tag{4.6}$$

Taking the LST on both sides of (4.6) yields

$$\overline{f}_{V_T/A=m}(s) = \left[\overline{f}_{H_O}(s)\right]^m \overline{f}_U\left[\lambda + s - \lambda \overline{f}_{H_O}(s)\right]. \tag{4.7}$$

Differentiating (4.7) with respect to s and then setting $s = 0$, we obtain the expected length of the complete startup period given by $A = m$ customers arriving during the period T as follows:

$$E[V_T \mid A = m] = \left(m + \lambda\mu_U\right) E\left[H\right] + \mu_U\,.$$

Using the fact that

$$E\left[V_T\right] = E\left[E\left[V_T \mid A\right]\right],$$

and

$$E[A] = \sum_{m=0}^{\infty} m \frac{(\lambda T)^m e^{-\lambda T}}{m!} = \lambda T \,,$$

we get

$$E[V_T] = E\left[E(V_T \mid A)\right] = E\left\{\left(A + \lambda \mu_U\right) E\left[H_O\right] + \mu_U\right\}.$$

It follows that

$$E[V_T] = E[A]E\left[H_O\right] + \lambda \mu_U E\left[H_O\right] + \mu_U = \frac{(T + \mu_U)\,\rho_H}{1 - \rho_H} + \mu_U \,. \tag{4.8}$$

4.2.2.3 Expected Length of the Busy and Breakdown Periods

Recall that $V_T = H_T + U_T$ which implies $E[V_T] = E[H_T] + E[U_T]$. Hence from (4.8), we obtain

$$E[H_T] = \frac{(T + \mu_U)\rho_H}{1 - \rho_H} = \frac{\mu_S (1 + \alpha \mu_R)(\lambda T + \rho_U)}{1 - \rho_H}, \tag{4.9}$$

and

$$E[U_T] = \mu_U. \tag{4.10}$$

Because the complete period is the sum of the busy period and the breakdown period, $H_T = B_T + D_T$, which implies $E[H_T] = E[B_T] + [D_T]$. From (4.9), we find that

$$E[B_T] = \frac{(\lambda T + \rho_U)\mu_S}{1 - \rho_H}, \tag{4.11}$$

and

$$E[D_T] = \frac{(\lambda T + \rho_U)\alpha \mu_S \mu_R}{1 - \rho_H}. \tag{4.12}$$

4.2.3 Expected length of the busy cycle

The busy cycle for the T policy M/G/1 queue with server breakdowns and general startup times, denoted by C_T, is the length of time from the beginning of the last idle period to the beginning of the next idle period. Since the busy cycle is the sum of the idle period (I_T), the startup period (U_T), the busy period (B_T), and the breakdown period (D_T), we get

$$E[C_T]=E[I_T]+E[U_T]+E[B_T]+E[D_T]=E[I_T]+E[V_T]. \quad (4.13)$$

From (4.5) and (4.8), we obtain

$$E[C_T]=\frac{T+\mu_U}{1-\rho_H}. \quad (4.14)$$

4.2.4 Probability that the server is turned-off, startup, busy and broken down

In steady-state, let

P_{I_T} ≡ probability that the server is idle;

P_{U_T} ≡ probability that the server is startup;

P_{B_T} ≡ probability that the server is busy;

P_{D_T} ≡ probability that the server is broken down.

From (4.5), (4.10)-(4.12) and (4.14), we get

$$P_{I_T}=\frac{E[I_T]}{E[C_T]}=\frac{T(1-\rho_H)}{T+\mu_U}, \quad (4.15)$$

$$P_{U_T}=\frac{E[U_T]}{E[C_T]}=\frac{\mu_U(1-\rho_H)}{T+\mu_U}, \quad (4.16)$$

$$P_{B_T}=\frac{E[B_T]}{E[C_T]}=\rho, \quad (4.17)$$

$$P_{D_T}=\frac{E[D_T]}{E[C_T]}=\alpha\rho\mu_R, \quad (4.18)$$

where $\rho=\lambda\mu_S$. We demonstrate from (4.17) that the probability that the server is busy in the steady-state is equal to ρ.

4.3 The Optimal *T* Policy

We develop the total expected cost function per unit time for the T policy M/G/1 queue with server breakdowns and general startup times in which T is a decision variable. We determine the optimum value of the control parameter T for our constructing expected cost function. Let us define the cost elements as follows:

C_h ≡ holding cost per unit time for each customer present in the system;

$C_s \equiv$ setup cost per busy cycle;

$C_i \equiv$ cost per unit time for keeping the server off;

$C_{sp} \equiv$ startup cost per unit time for the preparatory work of the server before starting the service;

$C_b \equiv$ cost per unit time for keeping the server on and in operation;

$C_d \equiv$ breakdown cost per unit time for a failed server.

Utilizing the definition of each cost element listed above, the expected cost function per customer per unit time is given by

$$F_O(T) = C_h L_T + C_s \frac{1}{E[C_T]} + C_i P_{I_T} + C_{sp} P_{U_T} + C_b P_{B_T} + C_d P_{D_T} . \tag{4.19}$$

Omitting these cost terms are not functions of the decision variable T, the optimization problem in (4.19) is equivalent to minimize the following equation:

$$F(T) = \frac{1}{(T+\mu_U)} \left\{ C_h \left[\frac{\lambda}{2} T^2 + \rho_U T + \frac{\lambda(\sigma_U^2 + \mu_U^2)}{2} \right] + \left(C_s + C_i T + C_{sp}\mu_U \right)(1-\rho_H) \right\} . \tag{4.20}$$

Differentiating $F(T)$ with respect to T, we get

$$\frac{dF(T)}{dT} = \frac{C_h \left[\frac{\lambda}{2} T^2 + \lambda \mu_U T - \frac{\lambda(\sigma_U^2 - \mu_U^2)}{2} \right] - \left[C_s + (C_{sp} - C_i)\mu_U \right](1-\rho_H)}{(T+\mu_U)^2} . \tag{4.21}$$

Setting $dF(T)/dT = 0$ yields

$$T^* = -\mu_U + \sqrt{\sigma_U^2 + \frac{2\left[C_s + (C_{sp} - C_i)\mu_U \right](1-\rho_H)}{C_h \lambda}} , \tag{4.22}$$

and since

$$\left. \frac{d^2 F(T)}{dT^2} \right|_{T=T^*} = \frac{C_h \lambda \sigma_U^2 + 2\left[C_s + (C_{sp} - C_i)\mu_U \right](1-\rho_H)}{(T^* + \mu_U)^3} > 0 . \tag{4.23}$$

Thus T^* is the unique minimizer of $F(T)$.

4.4 Sensitivity Analysis

A system analyst is concerned with varying the system parameters over a reasonable range and observing the relative change in the system performance measures. A sensitivity investigation of different system parameters (λ, μ, α, β, γ) and cost parameters (C_h, C_s, C_i, C_{sp},) levels is particularly valuable when evaluating future conditions, where $\mu_S = 1/\mu$, $\mu_R = 1/\beta$ and $\mu_U = 1/\gamma$. We can assess how robust the results are to system input parameters. In the following, we conduct some sensitivity investigations on the optimal value T^* based on changes in the values of system parameters and cost parameters. From (4.22), differentiating T^* with respect to λ, we obtain

$$\frac{\partial T^*}{\partial \lambda} = \frac{-\theta_1}{\lambda^2 \sqrt{\sigma_U^2 + \frac{2\theta_1(1-\rho_H)}{\lambda}}} < 0, \tag{4.24}$$

where $\theta_1 = \frac{\left[C_s + (C_{sp} - C_i)\mu_U\right]}{C_h}$.

It follows from (4.24) that T^* decreases in λ. Differentiating T^* with respect to μ yields

$$\frac{\partial T^*}{\partial \mu} = \frac{\theta_1(1+\alpha\mu_R)}{\mu^2 \sqrt{\sigma_U^2 + \frac{2\theta_1(1-\rho_H)}{\lambda}}} > 0. \tag{4.25}$$

It follows from (4.25) that T^* increases in μ. Similarly, differentiating T^* with respect to α and β respectively, we obtain

$$\frac{\partial T^*}{\partial \alpha} = \frac{-\theta_1\mu_S\mu_R}{\sqrt{\sigma_U^2 + \frac{2\theta_1(1-\rho_H)}{\lambda}}} < 0, \tag{4.26}$$

$$\frac{\partial T^*}{\partial \mu_R} = \frac{-\theta_1\alpha\mu_S}{\sqrt{\sigma_U^2 + \frac{2\theta_1(1-\rho_H)}{\lambda}}} < 0. \tag{4.27}$$

The above results imply that T^* decreases in α and μ_R, respectively. Recalling that $\mu_R = 1/\beta$, we conclude that T^* increases in β. Since σ_U^2 is a function of γ, we can see how γ affects T^* while startup time distribution is given. For special case, suppose that the startup time distribution obeys an exponential distribution with mean $\mu_U = 1/\gamma$. Substituting $\sigma_U^2 = \mu_U^2$ into (4.22) and then differentiating T^* with respect to μ_U, we get

$$\frac{\partial T^*}{\partial \mu_U} = -1 + \frac{\mu_U + \theta_2}{\sqrt{\mu_U^2 + 2\theta_2 \mu_U + \dfrac{2C_s(1-\rho_H)}{C_h \lambda}}}, \tag{4.28}$$

where $\theta_2 = \dfrac{(C_{sp} - C_i)(1-\rho_H)}{C_h \lambda}$.

If $\theta_2(C_{sp} - C_i) < 2C_s$, then we have $\partial T^* / \partial \mu_U < 0$. It follows that T^* increases in γ. If $\theta_2(C_{sp} - C_i) = 2C_s$, we get $\partial T^* / \partial \mu_U = 0$. Thus T^* is independent of γ. Furthermore, if $\theta_2(C_{sp} - C_i) > 2C_s$, we have $\partial T^* / \partial \mu_U > 0$. The result implies that T^* decreases in γ. On the other hand, it can easily see from (4.22) that (i) T^* increases in C_s and C_{sp}; and (ii) T^* decreases in C_i and C_h.

4.5 Numerical Computations

We present some numerical computations to verify the analytical results, and show how to make the decision based on minimizing the cost function $F(T)$. The sensitivity investigation concentrates mainly on the exponential startup time distribution. The cost parameters $C_s = 1000$, $C_h = 5$, $C_{sp} = 100$, $C_i = 60$ are fixed. We consider the following four cases.

Case 1: Choose $\mu = 0.5, 1.0, 1.5, 2.0$, $\alpha = 0.05$, $\beta = 3$, $\gamma = 3$, and vary the values of λ.

Case 2: Choose $\lambda = 0.2, 0.4, 0.6, 0.8$, $\alpha = 0.05$, $\beta = 3$, $\gamma = 3$, and vary the values of μ.

Case 3: Choose $\lambda = 0.5$, $\mu = 1$, $\beta = 1,2,3,4$, $\gamma = 3$ and vary the values of α.

Case 4: Choose $\lambda = 0.5$, $\mu = 1$, $\alpha = 0.2, 0.4, 0.6, 0.8$, $\gamma = 3$ and vary the values of β.

Figure 4.1 reveals that (i) T^* is decreasing in λ; (ii) as λ is fixed, T^* increases as μ increases; and (iii) if λ is small enough, T^* increases more quickly and the values of T^* rarely change for different values of μ. From Figure 4.2, we see that (i) T^* increases in μ; (ii) if μ is small enough, T^* increases quickly; (iii) if μ is large and $\rho = \lambda/\mu$ is small enough, T^* is insensitive; and (iv) if μ is fixed and large enough, T^* decreases as λ increases. Furthermore, the optimal value, T^*, and the corresponding minimum expected cost $F(T^*)$ are displayed in Table 4.1 for parameters $\alpha = 0.05$, $\beta = 3$, and $\gamma = 3$.

It appears from Figure 4.3 that (i) T^* decreases in α; (ii) as α is fixed, the larger β has larger T^*; and (iii) T^* has an upper bound as α closes to zero. Figure 4.4 reveals that (i) T^* increases in β but T^* is insensitive to β as β is large; and (ii) as β is fixed, the larger α has the smaller T^*. Furthermore, the optimal value, T^*, and the corresponding minimum expected cost $F(T^*)$ are shown in Table 4.2 for parameters $\lambda = 0.5$, $\mu = 1$, and $\gamma = 3$.

Figure 4.5 indicates that (i) T^* increases in γ; and (ii) as γ is smaller than 0.4, T^* increases quickly but T^* is insensitive to γ as γ is larger than 0.4. The optimal value, T^*, and the corresponding minimum expected cost $F(T^*)$ are displayed in Table 4.3 for parameters $\lambda = 0.3$, $\mu = 1$, $\alpha = 0.05$, and $\gamma = 3$.

To see how T^* changes when the cost parameter changes, we set $\lambda = 0.3$, $\mu = 1$, $\alpha = 0.5$, $\beta = 3$, $\gamma = 3$, choose $C_{sp} = 100$, $C_i = 60$, and vary the specified values of (C_s, C_h). We observe from Table 4.4 that T^* increases in C_s and decreases in C_h. On the other hand, we select $C_s = 1000$, $C_h = 5$, and change the specified values of (C_{sp}, C_i). Table 4.5 reveals that T^* increases in C_{sp} and decreases in C_i, but T^* is insensitive to (C_{sp}, C_i).

Figure 4.1 Plots of (λ, T^*) with μ =0.5, 1.0, 1.5, 2.0, α =0.05, β =3, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

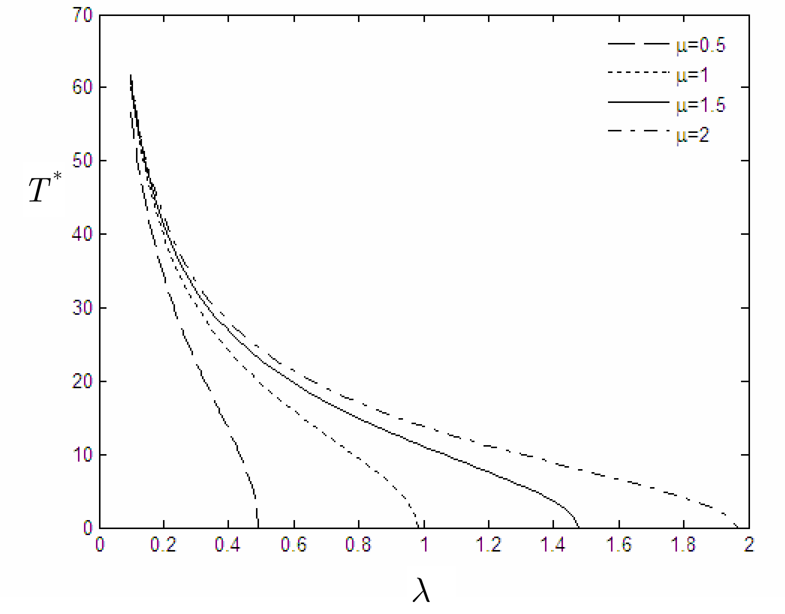


Figure 4.2 Plots of (μ, T^*) with λ =0.2, 0.4, 0.6, 0.8, α =0.05, β =3, γ =3, C_s =1000, C_h =5, C_{sp} = 100, C_i =60

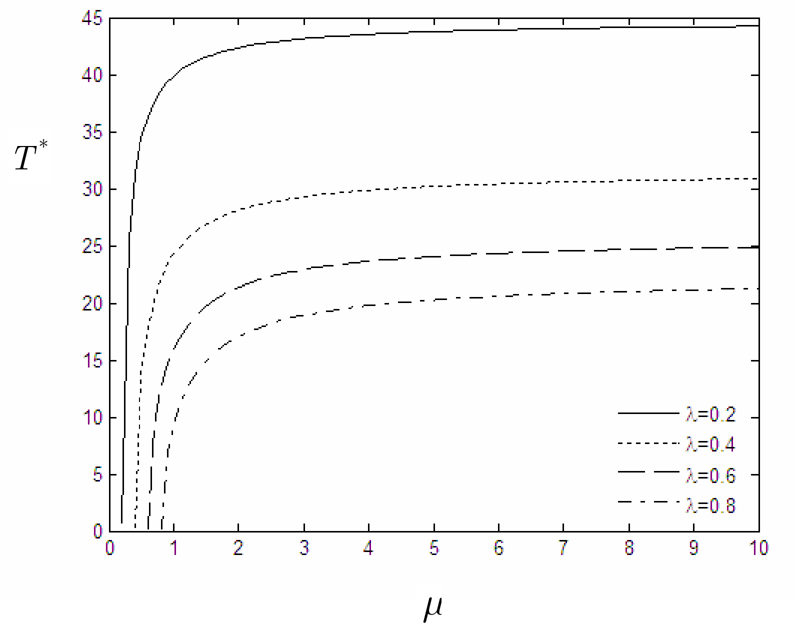


Table 4.1. The optimal T^* and minimum expected cost $F(T^*)$ with various values of (λ, μ).

$\alpha = 0.05$, $\beta = 3$, $\gamma = 3$, $C_s = 1000$, $C_h = 5$, $C_{sp} = 100$, $C_i = 60$								
$(\lambda,\ \mu)$	(0.3, 0.5)	(0.3, 1.0)	(0.3, 1.5)	(0.3, 2.0)	(0.2, 1.0)	(0.4, 1.0)	(0.6, 1.0)	(0.8, 1.0)
T^*	22.6241	30.3119	32.4767	33.5071	39.8499	24.1892	15.9017	9.3975
$F(T^*)$	57.8362	87.6679	97.0150	101.6107	87.9832	84.6451	72.1052	50.1233

Figure 4.3 Plots of (α, T^*) with λ=0.5, μ=1, β=1, 2, 3, 4, γ=3, C_s=1000, C_h=5, C_{sp}= 100, C_i=60

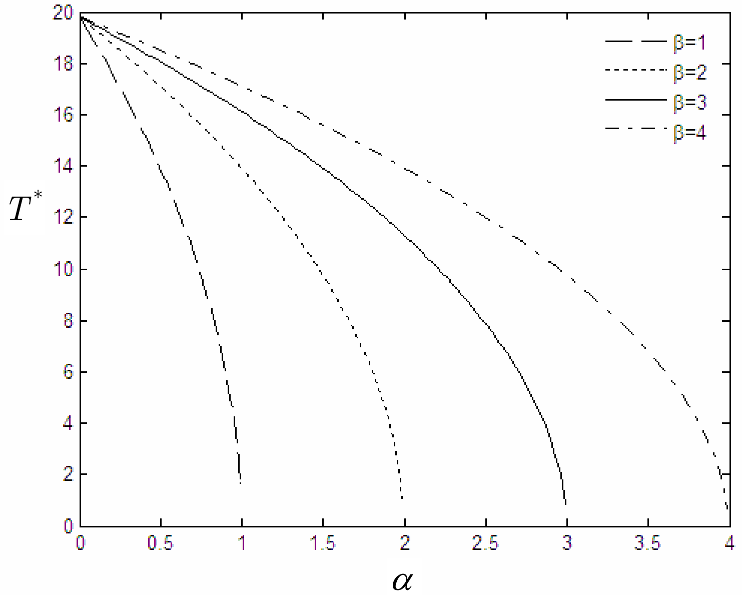


Figure 4.4 Plots of (β, T^*) with λ=0.5, μ=1, α=0.4, 0.8, 1.2, 1.6, γ=3, C_s=1000, C_h=5, C_{sp}= 100, C_i=60

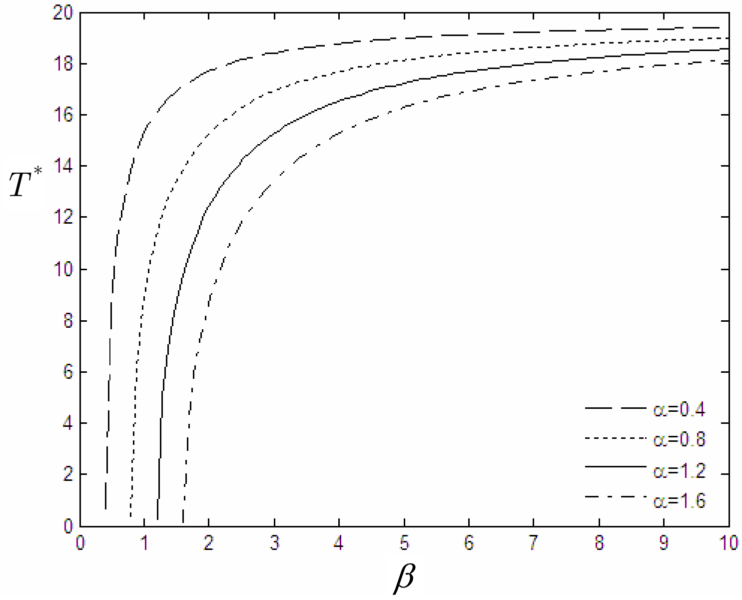


Table 4.2. The optimal T^* and minimum expected cost $F(T^*)$ with various values of (α, β).

$\lambda=0.5$, $\mu=1$, $\gamma=3$, $C_s=1000$, $C_h=5$, $C_{sp}=100$, $C_i=60$								
$(\alpha,\ \beta)$	(0.5, 1.0)	(0.5, 2.0)	(0.5, 3.0)	(0.5, 4.0)	(0.4, 2.0)	(0.8, 2.0)	(1.2, 2.0)	(1.6, 2.0)
T^*	6.7925	8.3908	8.8621	9.0889	8.6765	7.4712	6.0420	4.1808
$F(T^*)$	59.4781	76.9783	82.4260	85.0945	80.2654	66.7251	51.7803	34.1213

Figure 4.5 Plots of (γ, T^*) with λ=0.3, μ=1, α=0.05, β=3, C_s=1000, C_h=5, C_{sp}= 100, C_i=60

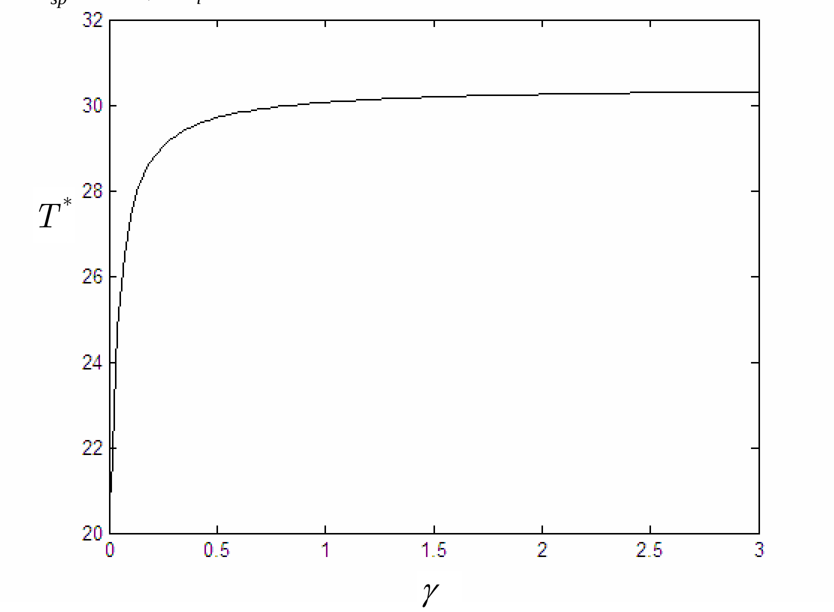


Table 4 3. The optimal T^* and minimum expected cost $F(T^*)$ with various values of γ

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $C_s=1000$, $C_h=5$, $C_{sp}=100$, $C_i=60$								
γ	0.4	0.8	1.2	1.6	2	2.4	2.8	3.2
T^*	7.3990	8.1910	8.4873	8.6419	8.7368	8.8009	8.8471	8.8821
$F(T^*)$	126.8279	126.2007	125.7975	125.5548	125.3955	125.2834	125.2006	125.1368

Table 4.4. The optimal T^* and minimum expected cost $F(T^*)$ with various values of (C_s, C_h).

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $\gamma=3$, $C_{sp}=100$, $C_i=60$								
(C_s, C_h)	(1000, 5)	(1000, 10)	(1000, 15)	(1000, 20)	(400, 10)	(600,10)	(800, 10)	(900,10)
T^*	8.8657	6.1757	4.9847	4.2753	3.8317	4.7349	5.4999	5.8470
$F(T^*)$	125.1667	159.6880	186.1419	208.4142	116.9593	133.4406	147.3823	153.7043

Table 4.5. The optimal T^* and minimum expected cost $F(T^*)$ with various values of (C_{sp}, C_i).

$\lambda=0.3$, $\mu=1$, $\alpha=0.05$, $\beta=3$, $\gamma=3$, $C_s=1000$, $C_h=5$								
(C_{sp}, C_i)	(80, 20)	(80, 30)	(80, 40)	(80, 50)	(35, 25)	(45, 25)	(55 , 25)	(65, 25)
T^*	8.8959	8.8808	8.8657	8.8506	8.8203	8.8355	8.8506	8.8657
$F(T^*)$	97.6411	104.4540	111.2667	118.0792	100.4285	100.5665	100.7042	100.8417

Chapter 5

Maximum Entropy Analysis to the *T* Policy M/G/1 Queue with Server Breakdowns and Startup Times

In this chapter, we use the maximum entropy approach to solve the steady-state probabilities of the T policy M/G/1 queue with server breakdowns and general startup time. Besides the constraints of normalizing condition and the probability of the various server statuses, the maximum entropy solutions are used to derive the queue length distributions using the first moment and second moment of the number of customers in the system, respectively. We derive the approximate formulas for the steady-state probability distributions of the queue length and perform a comparative analysis between the approximate results with established exact results for various distributions, such as exponential (M), k-stage Erlang (E_k), and deterministic (D). The experiment demonstrates that the maximum entropy approach is accurate enough for practical purposes and is a useful method for solving complex queueing systems by using the first moment of the number of customers in the system, which the use is better than the second moment of the number of customers in the system.

5.1 Assumptions and Notations

We consider a T policy M/G/1 queue in which the server performs a startup before providing service and is typically subject to unpredictable breakdowns. As soon as the system is empty, the server is immediately turned off. After the time is elapsed length T, the server is immediately turned on but is temporarily unavailable to serve the waiting customers if there is at least one customer in the waiting line; otherwise, the server waits another period of length T and so on until at least one customer is present. When the server turns on, he requires for the preparatory work (i.e. begin startup) before starting service. Once the startup is terminated, the server immediately starts serving the waiting customers.

It is assumed that customers arrive according to a Poisson process with parameter λ. The service times of the customers are independent and identically distributed (i.i.d.) random variables obeying an arbitrary distribution function $F_S(t)\,(t \geq 0)$ with a finite mean μ_S and a finite variance σ_S^2. The server is subject to breakdowns at any time with Poisson breakdown rate α when he is working. When the server fails, he is immediately repaired at a repair facility, where the repair times are i.i.d. random variables having a general distribution function $F_R(t)\,(t \geq 0)$ with a finite mean μ_R and a finite variance σ_R^2. Arriving customers form a single waiting line at a server based on the order of their arrivals. The server can serve only one customer at a time and the service is independent of the arrival process. A customer who arrives and finds the server busy or broken down must wait in the queue until a server is available. Although no service occurs during the repair period of the server, customers continue to arrive following a Poisson process. The server is turned off when the system is empty. If some customers are accumulated in the queue after T unit time is elapsed since system empty, the server is immediately turned on (i.e. begin startup) but is temporarily unavailable for the waiting customers. He needs a startup time with random length before starting service. Again, the startup times are i.i.d. random variables obeying a general distribution function $F_U(t)\,(t \geq 0)$ with a finite mean μ_U and a finite variance σ_U^2. Once the startup is terminated, the server begins serving the waiting customers until the system becomes empty. Service is allowed to be interrupted if the server breaks down, and the server is immediately repaired. Once the server is repaired, he immediately returns to serve customers until there are no customers in the system.

The exact steady-state solutions to the T policy M/G/1 queue with service times, repair times or startup times distribution of the general type have not been found. It is extremely difficult, if not impossible, to obtain the explicit formulas such as the steady-state probability mass function of the number of customers and the expected waiting time for the T policy M/G/1 queue in which the repair times and startup times are generally distributed. However, one can utilize the maximum entropy principle to approximate the T policy M/G/1 queue with general repair times and general startup times. This becomes particularly helpful when some system performance measures (for instance, the expected number of customers in the system, the probability that the server is busy, broken down, etc) are known. In this paper, we utilize the maximum entropy principle associated with five basic known results from the literature to study the T policy M/G/1 queue with general repair times and general startup times. Next, we replace the first moment of customers in the system by the second moment of the queue length to study the T policy M/G/1 queue with general repair times and general startup times

The purpose of this chapter is:

(ii) to present some important system performance measures for the T policy M/G/1 queue with general repair times and general startup times;

(iii) to develop the maximum entropy (approximate) solutions for the T policy M/G/1 queue with general repair times and general startup times by using Lagrange's method;

(iv) to obtain approximate results for the expected waiting time in the queue;

(v) to perform a comparative analysis between exact results and two approximate results obtained through maximum entropy principle by using (i) the first moment of the queue length and (ii) the second moment of the queue length.

In this chapter, the following notations and probabilities are used.

T – threshold

S – service time random variable

U – startup time random variable

R – repair time random variable

$F_S(\cdot)$ – distribution function of S

$F_U(\cdot)$ – distribution function of U

$F_R(\cdot)$ – distribution function of R

$G(z)$ – p.g.f. of the number of customers in the ordinary M/G/1 queue with unreliable server

$A(t)$ – number of customers arriving in the system during $[0, t]$

A_m – arrival time of the m-th customer

$F_{A_m}(\cdot)$ – distribution function of A_m

$G_I(z)$ – p.g.f. of the number of customers waiting in the queue during an idle period

$G_U(z)$ p.g.f. of the number of customers arriving during startup period

$\bar{f}_U(\cdot)$ – Laplace-Stieltjes transform (LST) of startup time

$W(z)$ – p.g.f. of the number of customers that arrive during turned-off period and startup period

$G_T(z)$ – p.g.f. of the number of customers in the T policy M/G/1 queue with server breakdowns and general startup times

L_{1T} – expected number of customers in the T policy M/G/1 queue with server breakdowns and general startup times

L_{2T} – second moment of the queue length in the T policy M/G/1 queue with server breakdowns and general startup times

$P_{0,I}(\cdot)$ – probability that there are n customers in the system when the server is turned off

$P_{0,U}(\cdot)$ – probability that there are n customers in the system when the server is startup

$P_1(\cdot)$ – probability that there are n customers in the system when the server is turned on and working

$P_2(\cdot)$ – probability that there are n customers in the system when the server is in operation but found to be broken down

W_q – exact expected waiting time in the queue

I_S – idle state

U_S – startup state

B_S – busy state

R_S – repair state

W_q^* – approximate expected waiting time in the queue

5.2 The Expected Number of Customers in the System

Let H be a random variable representing the completion time of a customer, which includes both the service time of a customer and the repair time of a server. Applying the well-known results of Medhi and Templeton [34], the probability generating function (p.g.f.) of the number of customers in the ordinary M/G/1 queue with server breakdowns is given by

$$G(z) = \frac{(1-\rho_H)(1-z)\overline{f}_H(\lambda-\lambda z)}{\overline{f}_H(\lambda-\lambda z)-z}, \tag{5.1}$$

where $\rho_H = \lambda E[H]$. In addition, $E\left[H\right] = \mu_S(1+\alpha\mu_R)$ and $E[H^2] = (1+\alpha\mu_R)^2$

$(\mu_S^2+\sigma_S^2)+\alpha\mu_S(\mu_R^2+\sigma_R^2)$. The traffic intensity ρ_H is assumed to be less than 1. We consider a Poisson arriving process. Let ξ_i denote the elapsed time between the (i-1)st and the i-th customer arriving. Following Ross [36], the event ξ_i are i.i.d. exponential random variables with mean $1/\lambda$. Let $A(t)$ denote the number of customer arriving at the system during $[0,t]$. Let A_m be the arrival time of the m-th customer and $A_m=\sum_{i=1}^{m}\xi_i$. The distribution of A_m occurring by time t is given by

$$F_{A_m}(t)=P\{A_m\le t\}=\int_0^t\frac{\lambda(\lambda x)^{m-1}}{(m-1)!}e^{-\lambda x}dx=1-\sum_{k=0}^{m-1}\frac{(\lambda t)^k}{k!}e^{-\lambda t}=P\{A(t)\ge m\}.$$

It is obvious

$$P\{A(T)=m\}=F_{A_m}(T)-F_{A_{m+1}}(T),$$

where $\sum_{a}^{b}\cdot=0$ for $b<a$.

Hence in a period of length T, at least $m\ge1$ customers in the system is given

$$F_{A_m}(T)=P\{A_m\le T\}=1-\sum_{k=0}^{m-1}\frac{(\lambda T)^k}{k!}e^{-\lambda T}.$$

Let G_I be the p.g.f. of the number of customers waiting in the queue during idle period I_T. Thus we have

$$G_I(z)=\sum_{m=0}^{\infty}z^m\left[F_{A_m}(T)-F_{A_{m+1}}(T)\right]=e^{-(1-z)\lambda T}.$$

Let G_U be the p.g.f. of the number of customers arriving during startup period U_T. Then we get $G_U(z)=e^{-(1-z)\lambda t}$ and the Laplace-Stieltjes transform (abbreviated LST) of $G_U(z)$ given by

$$\int_0^\infty G_U(z)\,dF_U(t)=\overline{f}_U\left[\lambda(1-z)\right].$$

Because the Poisson process from any point on arrival is independent of all that has previously occurred, $W(z)=G_I(z)\overline{f}_U[(1-z)\lambda]=e^{-(1-z)\lambda T}\overline{f}_U[(1-z)\lambda]$. For the T policy M/G/1 queue with server breakdowns and startup time, we get the idle period plus the startup period as $W(z)$. Using the well-known decomposition property concerning M/G/1 vacation queue of Fuhrmann and Cooper [14], we obtain the p.g.f. of number of customers in the T policy M/G/1 queue with server breakdowns and general startup times as follows:

$$G_T(z)=G(z)\left[\frac{1-W(z)}{W'(1)(1-z)}\right], \tag{5.2}$$

where $G(z)$ is given in (5.1). Let L_{1T} and L_{2T} denote the first moment and the second moment of the number of customers in the T policy M/G/1 queue with server breakdowns and general startup times, respectively. Thus we obtain

$$L_{1T}=G_T'(z)\big|_{z=1}=\frac{1}{(\lambda T+\rho_U)}\left[\frac{\lambda^2T^2}{2}+\lambda T\rho_U+\frac{\lambda^2\sigma_U^2+\rho_U^2}{2}\right]+\rho_H+\frac{\lambda^2E[H^2]}{2(1-\rho_H)}, \quad (5.3)$$

and

$$L_{2T}=2\Pi\left(\frac{\lambda T}{3}+\rho_H\right)+\frac{\lambda^3\left(T^2\mu_U+2TE[U^2]+E[U^3]\right)}{3(\lambda T+\rho_U)}$$
$$+\lambda^2E[H^2]\left(1+\frac{L_{1T}}{(1-\rho_H)}\right)+\frac{\lambda^3E[H^3]}{3(1-\rho_H)}+L_{1T}, \quad (5.4)$$

where $\rho_U=\lambda\mu_U$, $\Pi=\dfrac{\lambda^2T^2+2\lambda T\rho_U+\lambda^2E[U^2]}{2(\lambda T+\rho_U)}$ and

$$E[H^3]=(1+\alpha\mu_R)^3E[S^3]+3\alpha(1+\alpha\mu_R)E[S^2]E[R^2]+\alpha\mu_SE[R^3].$$

5.3 The Maximum Entropy Results

In this section, we will develop the maximum entropy solutions for the steady state probabilities of the T policy M/G/1 queue with server breakdowns and general startup times. Let us define

$P_{0,I}(n)\equiv$ probability that there are n customers in the system when the server is turned off, where $n=0,1,2,\cdots$.

$P_{0,U}(n)\equiv$ probability that there are n customers in the system when the server is startup, where $n=1,2,3,\cdots$.

$P_1(n)\equiv$ probability that there are n customers in the system when the server is turned on and working, where $n=1,2,3,\cdots$.

$P_2(n)\equiv$ probability that there are n customers in the system when the server is in operation but found to be broken down, where $n=1,2,3,\cdots$.

In order to derive the steady-state probabilities $P_{0,I}(n)$, $P_{0,U}(n)$ and $P_i(n)$ (i = 1, 2) by using the maximum entropy principle, we formulate the maximum entropy model in the following. Following El-Affendi and Kouvatsos [12], the entropy function Y can be illustrated mathematically as

$$Y=-\sum_{n=0}^{\infty}P_{0,I}(n)\ln P_{0,I}(n)-\sum_{n=1}^{\infty}P_{0,U}(n)\ln P_{0,U}(n)-\sum_{n=1}^{\infty}P_1(n)\ln P_1(n)$$

$$-\sum_{n=1}^{\infty} P_2(n)\ln P_2(n)\,. \tag{5.5}$$

There are five basic known results from the literature (see [8] and [54]) that facilitate the application of the maximum entropy formalism to study the T policy M/G/1 queue with server breakdowns and general startup times. The maximum entropy solutions are obtained by maximizing (5.5) subject to the following five constraints, written as,

(i) normalizing condition

$$\sum_{n=0}^{\infty} P_{0,I}(n)+\sum_{n=1}^{\infty} P_{0,U}(n)+\sum_{n=1}^{\infty} P_1(n)+\sum_{n=1}^{\infty} P_2(n)=1, \tag{5.6}$$

(ii) the probability that the server is startup

$$\sum_{n=1}^{\infty} P_{0,U}(n)=\frac{\mu_U(1-\rho_H)}{T+\mu_U}, \tag{5.7}$$

(iii) the probability that the server is busy

$$\sum_{n=1}^{\infty} P_1(n)=\rho, \tag{5.8}$$

(iv) the probability that the server is broken down

$$\sum_{n=1}^{\infty} P_2(n)=\rho\alpha\mu_R, \tag{5.9}$$

(v) the expected number of customers in the system

$$\sum_{n=0}^{\infty} nP_{0,I}(n)+\sum_{n=1}^{\infty} nP_{0,U}(n)+\sum_{n=1}^{\infty} nP_1(n)+\sum_{n=1}^{\infty} nP_2(n)=L_{1T}, \tag{5.10}$$

where L_{1T} is given in (5.3).

In (5.6)-(5.10), (5.6) is multiplied by τ_1, (5.7) is multiplied by τ_2, (5.8) is multiplied by τ_3, (5.9) is multiplied by τ_4 and (5.10) is multiplied by τ_5. Thus the Lagrangian function y_1 is given by

$$y_1=-\sum_{n=0}^{\infty} P_{0,I}(n)\ln P_{0,I}(n)-\sum_{n=1}^{\infty} P_{0,U}(n)\ln P_{0,U}(n)$$

$$-\sum_{n=1}^{\infty} P_{0,I}(n)\ln P_1(n)-\sum_{n=1}^{\infty} P_2(n)\ln P_2(n)$$

$$-\tau_1\left[\sum_{n=0}^{\infty}P_{0,I}(n)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_1(n)+\sum_{n=1}^{\infty}P_2(n)-1\right]$$

$$-\tau_2\left[\sum_{n=1}^{\infty}P_{0,U}(n)-\frac{\mu_U(1-\rho_H)}{T+\mu_U}\right]-\tau_3\left[\sum_{n=1}^{\infty}P_1(n)-\rho\right]-\tau_4\left[\sum_{n=1}^{\infty}P_2(n)-\rho\alpha\mu_R\right]$$

$$-\tau_5\left[\sum_{n=0}^{\infty}nP_{0,I}(n)+\sum_{n=1}^{\infty}nP_{0,U}(n)+\sum_{n=1}^{\infty}nP_1(n)+\sum_{n=1}^{\infty}nP_2(n)-L_{1T}\right]. \tag{5.11}$$

where τ_1 - τ_5 are the Lagrangian multipliers corresponding to constraints (5.6)-(5.10), respectively.

5.3.1 The maximum entropy solutions with the first moment of the queue length

To get the maximum entropy solutions, $P_{0,U}(n)$, $P_1(n)$, $P_2(n)$, maximizing in (5.5) subject to constraints (5.6)-(5.10) is equivalent to maximizing (5.11).

The maximum entropy solutions are obtained by taking the partial derivatives of y_1 with respect to $P_{0,I}(n)$, $P_{0,U}(n)$ and $P_i(n)$ (i = 1, 2), and setting the results equal to zero, namely,

$$\frac{\partial y_1}{\partial P_{0,I}(0)}=-1-\ln P_{0,I}(0)-\tau_1=0, \tag{5.12}$$

$$\frac{\partial y_1}{\partial P_{0,I}(n)}=-1-\ln P_{0,I}(n)-\tau_1-n\tau_5=0,\quad n=1,2,\cdots \tag{5.13}$$

$$\frac{\partial y_1}{\partial P_{0,U}(n)}=-1-\ln P_{0,U}(n)-\tau_1-\tau_2-n\tau_5=0,\quad n=1,2,\cdots \tag{5.14}$$

$$\frac{\partial y_1}{\partial P_1(n)}=-1-\ln P_1(n)-\tau_1-\tau_3-n\tau_5=0,\quad n=1,2,\cdots \tag{5.15}$$

$$\frac{\partial y_1}{\partial P_2(n)}=-1-\ln P_2(n)-\tau_1-\tau_4-n\tau_5=0,\quad n=1,2,\cdots. \tag{5.16}$$

It implies from (5.12)-(5.16) that we obtain

$$P_{0,I}(0)=e^{-(1+\tau_1)}, \tag{5.17}$$

$$P_{0,I}(n)=e^{-(1+\tau_1+n\tau_5)},\quad n=1,2,\cdots \tag{5.18}$$

$$P_{0,U}(n)=e^{-(1+\tau_1+\tau_2+n\tau_5)},\quad n=1,2,\cdots \tag{5.19}$$

$$P_1(n)=e^{-(1+\tau_1+\tau_3+n\tau_5)},\quad n=1,2,\cdots \tag{5.20}$$

$$P_2(n) = e^{-(1+\tau_1+\tau_4+n\tau_5)}, \quad n = 1, 2, \cdots. \tag{5.21}$$

Let $\phi_1 = e^{-(1+\tau_1)}$, $\phi_2 = e^{-\tau_2}$, $\phi_3 = e^{-\tau_3}$, $\phi_4 = e^{-\tau_4}$ and $\phi_5 = e^{-\tau_5}$. We transform (5.17)-(5.21) in terms ϕ_1, ϕ_2, ϕ_3, ϕ_4 and ϕ_5 given by

$$P_{0,I}(0) = \phi_1 \tag{5.22}$$

$$P_{0,I}(n) = \phi_1\phi_5^n, \quad n = 1, 2, \cdots \tag{5.23}$$

$$P_{0,U}(n) = \phi_1\phi_2\phi_5^n, \quad n = 1, 2, \cdots \tag{5.24}$$

$$P_1(n) = \phi_1\phi_3\phi_5^n, \quad n = 1, 2, \cdots \tag{5.25}$$

$$P_2(n) = \phi_1\phi_4\phi_5^n, \quad n = 1, 2, \cdots. \tag{5.26}$$

Substituting (5.23)-(5.25) into (5.7)-(5.9), respectively, yields

$$\phi_1\phi_2 = \frac{\mu_U(1-\rho_H)(1-\phi_5)}{\phi_5(T+\mu_U)}, \tag{5.27}$$

$$\phi_1\phi_3 = \frac{\rho(1-\phi_5)}{\phi_5}, \tag{5.28}$$

$$\phi_1\phi_4 = \frac{\rho\alpha\mu_R(1-\phi_5)}{\phi_5}. \tag{5.29}$$

Substituting (5.22)-(5.26) into (5.6) and using (5.27)-(5.29), we obtain

$$\phi_1 = \frac{T(1-\rho_H)(1-\phi_5)}{T+\mu_U}. \tag{5.30}$$

Substituting (5.22)-(5.26) and (5.30) into (5.10) and taking the algebraic manipulations, we obtain

$$\phi_5 = \frac{[L_{1T} - \Theta_2(1-\rho_H) - \rho_H](T+\mu_U)}{T(1-\rho_H) + L_{1T}(T+\mu_U)}, \tag{5.31}$$

where $\Theta_2 = \dfrac{\mu_U}{T+\mu_U}$.

Finally, we get

$$P_{0,I}(0) = \frac{T(1-\rho_H)(1-\phi_5)}{T+\mu_U}, \tag{5.32}$$

$$P_{0,I}(n)=\frac{T\left(1-\rho_H\right)(1-\phi_5)\phi_5^n}{T+\mu_U},\quad n=1,2,\cdots \tag{5.33}$$

$$P_{0,U}(n)=\Theta_2(1-\rho_H)(1-\phi_5)\phi_5^{n-1},\quad n=1,2,\cdots \tag{5.34}$$

$$P_1(n)=\rho(1-\phi_5)\phi_5^{n-1},\quad n=1,2,\cdots \tag{5.35}$$

$$P_2(n)=\rho\alpha\mu_R(1-\phi_5)\phi_5^{n-1},\quad n=1,2,\cdots. \tag{5.36}$$

5.3.2 The maximum entropy solutions with the second moment of the queue length

In this section, we develop another maximum entropy solutions based on the second moment of the queue length. The maximum entropy is obtained by maximizing (5.4) by replacing the (5.10) with the second moment of the number of customers in the system. That is, maximize the Lagrangian function y_2 given by

$$\begin{aligned}
y_2 =& -\sum_{n=0}^{\infty}P_{0,I}(n)\ln P_{0,I}(n)-\sum_{n=1}^{\infty}P_{0,U}(n)\ln P_{0,U}(n)\\
&-\sum_{n=1}^{\infty}P_{0,I}(n)\ln P_1(n)-\sum_{n=1}^{\infty}P_2(n)\ln P_2(n)\\
&-\omega_1\left[\sum_{n=0}^{\infty}P_{0,I}(n)+\sum_{n=N}^{\infty}P_{0,U}(n)+\sum_{n=1}^{\infty}P_1(n)+\sum_{n=1}^{\infty}P_2(n)-1\right]\\
&-\omega_2\left[\sum_{n=1}^{\infty}P_{0,U}(n)-\frac{\mu_U(1-\rho_H)}{T+\mu_U}\right]-\omega_3\left[\sum_{n=1}^{\infty}P_1(n)-\rho\right]-\omega_4\left[\sum_{n=1}^{\infty}P_2(n)-\rho\alpha\mu_R\right]\\
&-\omega_5\left[\sum_{n=0}^{\infty}n^2P_{0,I}(n)+\sum_{n=1}^{\infty}n^2P_{0,U}(n)+\sum_{n=1}^{\infty}n^2P_1(n)+\sum_{n=1}^{\infty}n^2P_2(n)-L_{2T}\right],
\end{aligned} \tag{5.37}$$

which is subject to (5.6), (5.7), (5.8), (5.9) and

$$\sum_{n=0}^{\infty}n^2P_{0,I}(n)+\sum_{n=1}^{\infty}n^2P_{0,U}(n)+\sum_{n=1}^{\infty}n^2P_1(n)+\sum_{n=1}^{\infty}n^2P_2(n)=L_{2T}. \tag{5.38}$$

The maximum entropy solutions are obtained by taking the partial derivatives of y_2 with respect to $P_{0,I}(n)$, $P_{0,U}(n)$ and $P_i(n)$ (i = 1, 2), and setting the results equal to zero, namely,

$$\frac{\partial y_2}{\partial P_{0,I}(0)}=-1-\ln P_{0,I}(0)-\omega_1=0, \tag{5.39}$$

$$\frac{\partial y_2}{\partial P_{0,I}(n)}=-1-\ln P_{0,I}(n)-\omega_1-n^2\omega_5=0,\quad n=1,2,\cdots \tag{5.40}$$

$$\frac{\partial y_2}{\partial P_{0,U}(n)} = -1 - \ln P_{0,U}(n) - \omega_1 - \omega_2 - n^2\omega_5 = 0, \quad n = 1, 2, \cdots \tag{5.41}$$

$$\frac{\partial y_2}{\partial P_1(n)} = -1 - \ln P_1(n) - \omega_1 - \omega_3 - n^2\omega_5 = 0, \quad n = 1, 2, \cdots \tag{5.42}$$

$$\frac{\partial y_2}{\partial P_2(n)} = -1 - \ln P_2(n) - \omega_1 - \omega_4 - n^2\omega_5 = 0, \quad n = 1, 2, \cdots. \tag{5.43}$$

Let $\varphi_1 = e^{-(1+\omega_1)}$, $\varphi_2 = e^{-\omega_2}$, $\varphi_3 = e^{-\omega_3}$, $\varphi_4 = e^{-\omega_4}$ and $\varphi_5 = e^{-\omega_5}$. We transform (5.39)-(5.43) in terms φ_1, φ_2, φ_3, φ_4 and φ_5 given by

$$P_{0,I}(0) = \varphi_1, \tag{5.44}$$

$$P_{0,I}(n) = \varphi_1\varphi_5^{n^2}, \quad n = 1, 2, \cdots \tag{5.45}$$

$$P_{0,U}(n) = \varphi_1\varphi_2\varphi_5^{n^2}, \quad n = 1, 2, \cdots \tag{5.46}$$

$$P_1(n) = \varphi_1\varphi_3\varphi_5^{n^2}, \quad n = 1, 2, \cdots \tag{5.47}$$

$$P_2(n) = \varphi_1\varphi_4\varphi_5^{n^2}, \quad n = 1, 2, \cdots. \tag{5.48}$$

Substituting (5.46)-(5.48) into (5.7)-(5.9), respectively, yields

$$\varphi_1\varphi_2 = \frac{\mu_U(1-\rho_H)}{(T+\mu_U)\sum_{n=1}^{\infty}\varphi_5^{n^2}}, \tag{5.49}$$

$$\varphi_1\varphi_3 = \frac{\rho}{\sum_{n=1}^{\infty}\varphi_5^{n^2}}, \tag{5.50}$$

$$\varphi_1\varphi_4 = \frac{\rho\alpha\mu_R}{\sum_{n=1}^{\infty}\varphi_5^{n^2}}. \tag{5.51}$$

Substituting (5.44)-(5.48) into (5.6) and then replacing by (5.49)-(5.51), we get

$$\varphi_1 = \frac{T(1-\rho_H)}{(T+\mu_U)\sum_{n=0}^{\infty}\varphi_5^{n^2}}. \tag{5.52}$$

Substituting (5.44)-(5.48) and (5.52) into (5.37) and taking the algebraic manipulations, we obtain

$$\left[\frac{T(1-\rho_H)}{\sum_{n=0}^{\infty}\varphi_5^{n^2}}+\frac{(\mu_U+T\rho_H)}{\sum_{n=1}^{\infty}\varphi_5^{n^2}}\right]\frac{\sum_{n=0}^{\infty}n^2\varphi_5^{n^2}}{\left(T+\mu_U\right)}=L_{2T}\,. \tag{5.53}$$

The explicit solution for φ_5 in (5.53) does not exist. It should be to note that the approximate solution of φ_5 is obtained using Newton method. In iteration process, one requires the successive approximate solution of φ_5 is less than 0.00001.

5.4 The Exact and Approximate Expected Waiting Time in the Queue

In this section, we develop the exact and the approximate formulae for the expected waiting time in the T policy M/G/1 queue with server breakdowns and general startup times as follows.

5.4.1 The exact expected waiting time in the queue

Let W_q denote the exact expected waiting time in the queue. Using (5.3) and Little's formula, we obtain

$$W_q=\frac{L_{1T}}{\lambda}-E[H]=\frac{1}{(T+\mu_U)}\left[\frac{T^2}{2}+T\mu_U+\frac{\sigma_U^2+\mu_U^2}{2}\right]+\frac{\lambda E[H^2]}{2(1-\rho_H)}\,. \tag{5.54}$$

5.4.2 The approximate expected waiting time in the queue

We define the idle state, the startup state, the busy state, and the repair state as follows:

(i) Idle state denoted by I_S: the server is turned off and the number of customers waiting in the system is greater than or equal to 0.

(ii) Startup state denoted by U_S: the server begins startup and the number of customers waiting in the system is greater than or equal to 1.

(iii) Busy state denoted by B_S: the server is busy and provides service to a customer.

(iv) Repair state denoted by R_S: the server is broken down and being repaired.

Following Borthakur et al. [8], we find the expected waiting time of customer C at the states I_S, U_S, B_S and R_S as follows. Suppose that a 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_S, U_S, B_S and R_S are described, respectively, as follows: we have

(i) In idle state I_S: The server will begin startup after customer C arrive and n customers in front of him waiting for service. Following the well known theorem of renewal theory (mean residual life), the mean remaining idle time is $T/2$. The expected waiting time of customer C at the idle state is

$$\left(\frac{T}{2}+\mu_U+n\mu_S\right).$$

(ii) In startup state U_S: Using the same argument as (i), we get the mean remaining startup time $E[U^2]/2\mu_U$. Thus we obtain the expected waiting time of customer C at the startup state is $n\mu_S+E[U^2]/2\mu_U$.

(iii) In busy state B_S: Since the server is turned on and working, customer C only waits n customers in front of him to be served. The expected waiting time of customer C at the busy state is $n\mu_S$.

(iv) In repair state R_S: Using the same argument as (ii), we have the expected waiting time of customer C at the repair state is $n\mu_S+E[R^2]/2\mu_R$.

Finally, using the listed above results we obtain the approximate expected waiting time in the queue given by

$$\begin{aligned} W_q^* &= \sum_{n=0}^{\infty}\left(\frac{T}{2}+\mu_U+n\mu_S\right)P_{0,I}(n)+\sum_{n=1}^{\infty}\left(n\mu_S+\frac{E[U^2]}{2\mu_U}\right)P_{0,S}(n) \\ &+\sum_{n=1}^{\infty}n\mu_S P_1(n)+\sum_{n=1}^{\infty}\left(n\mu_S+\frac{E[R^2]}{2\mu_R}\right)P_2(n). \end{aligned} \tag{5.55}$$

where $P_{0,I}(n)$, $P_{0,S}(n)$, $P_1(n)$, and $P_2(n)$ are given in (5.32)-(5.36) (or (5.44)-(5.48)), respectively.

5.5 Comparative Analysis

The primary objective of this section is to examine the accuracy of the two maximum entropy results. We present specific numerical comparisons between the exact results and the two maximum entropy (approximate) results for the T policy M/G/1 queue with general service times, general repair times and general startup times. Conveniently, we represent this queue as the T policy M/G(G,G)/l queue where the second, third, fourth symbols denote the general distribution of service time, repair time, and startup time, respectively.

This section includes the following three subsections:

(i) comparative analysis for the T policy M/M(M,M)/1 and M/D(D,D)/1 queues.

(ii) comparative analysis for the T policy M/E_3(E_4,E_3)/1 and M/M(E_3,E_2)/1 queues.

(iii) comparative analysis for the T policy M/E_3(E_4,D)/1 and M/E_3(E_4,M)/1 queues.

5.5.1 Comparative analysis for the T policy **M/M(M,M)/1** *and* **M/D(D,D)/1 q*ueues***

Here we perform a comparative analysis between the exact W_q and the two approximate (maximum entropy) W_q^* for the T policy M/M(M,M)/1 and M/D(D,D)/1 queue. For the T policy M/M(M,M)/1 queues, we obtain $\mu_S = 1/\mu$, $E[S^2] = 2/\mu^2$, $E[S^3] = 6/\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 2/\beta^2$, $E[R^3] = 6/\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 2/\gamma^2$, and $E[U^3] = 6/\gamma^3$. For the T policy M/D(D,D)/1 queue, we have $\mu_S = 1/\mu$, $E[S^2] = 1/\mu^2$, $E[S^3] = 1/\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 1/\beta^2$, $E[R^3] = 1/\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 1/\gamma^2$ and $E[U^3] = 1/\gamma^3$.

We set T =5 and T =10, and choose the various values of λ, μ, α, β, and γ. The numerical results are obtained by considering the following parameters:

Case 1: We fix μ =1.0, α =0.05, β =3.0, γ =3.0, and vary the values of λ from 0.2 to 0.8.

Case 2: We fix λ =0.3, α =0.05, β =3.0, γ =3.0, and vary the values of μ from 0.5 to 2.0.

Case 3: We fix λ =0.3, μ =1.0, β =3.0, γ =3.0, and vary the values of α from 0.05 to 0.2.

Case 4: We fix λ =0.3, μ =1.0, α =0.05, γ =3.0, and vary the values of β from 2.0 to 5.0.

Case 5: We fix λ =0.3, μ =1.0, α =0.05, β =3.0, and vary the values of γ from 2.0 to 5.0.

Numerical results of W_q and W_q^* for the T policy M/M(M,M)/1 and M/D(D,D)/1 queues are shown in Table 5.1 for the above five cases using the constraint of the first moment of the queue length. The most relative error percentages are small (0.2-6.0%). Table 5.2 presents numerical results using the constraint of the second moment of the queue length. The range of relative error percentages is wider (1.3-27.9%).

5.5.2 Comparative analysis for the T policy **M/E$_3$(E$_4$,E$_3$)/1** *and* **M/M(E$_3$,E$_2$)/1** *queues*

Here we perform a comparative analysis between the exact W_q and the two approximate (maximum entropy) W_q^* for the T policy M/E$_3$(E$_4$,E$_3$)/1 and M/M(E$_3$, E$_2$)/1 queues. For the T policy M/E$_3$(E$_4$,E$_3$)/1 queue, we have $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $E[S^3] = 20/9\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 4/3\beta^2$, $E[R^3] = 15/8\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 4/3\gamma^2$, and $E[U^3] = 20/9\gamma^3$. For the T policy M/M(E$_3$,E$_2$)/1 queue, we get $\mu_S = 1/\mu$, $E[S^2] = 2/\mu^2$, $E[S^3] = 6/\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 4/3\beta^2$, $E[R^3] = 20/9\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 3/2\gamma^2$ and $E[U^3] = 3/\gamma^3$.

Numerical results of W_q and W_q^* for the T policy M/E$_3$(E$_4$,E$_3$)/1 and M/M(E$_3$,E$_2$)/1 queues are shown in Table 5.3 for the above five cases based on the constraint of the first moment of the queue length. The most relative error percentages are very small (0.2-4.3%). Table 5.4 displays numerical results using the constraint of the second moment of the queue length. The relative error percentages are lager.

5.5.3 Comparative analysis for the T policy **M/E$_3$(E$_4$,D)/1** *and* **M/E$_3$(E$_4$,M)/1** *queues*

Here we perform a comparative analysis between the exact W_q and the two approximate (maximum entropy) W_q^* for the T policy M/E$_3$(E$_4$,D)/1 and M/E$_3$(E$_4$, M)/1 queues. For the T policy M/E$_3$(E$_4$,D)/1 queue, we get $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $E[S^3] = 20/9\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 5/4\beta^2$, $E[R^3] = 15/8\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 1/\gamma^2$, and $E[U^3] = 6/\gamma^3$. For the T policy M/E$_3$(E$_4$,M)/1 queue, we obtain $\mu_S = 1/\mu$, $E[S^2] = 4/3\mu^2$, $E[S^3] = 20/9\mu^3$, $\mu_R = 1/\beta$, $E[R^2] = 5/4\beta^2$, $E[R^3] = 15/8\beta^3$, $\mu_U = 1/\gamma$, $E[U^2] = 2/\gamma^2$ and $E[U^3] = 6/\gamma^3$.

Numerical results of W_q and W_q^* for the T policy M/E$_3$(E$_4$,D)/1 and M/E$_3$(E$_4$,M)/1 queues are shown in Table 5.5 for the above five cases based on the first moment of queue length. Again, the most relative error percentages are very small (0.2-4.4%). Table 5.6 shows numerical results using the constraint of the second moment of the queue length. The relative error percentages are lager (1.1-26.8%).

Table 5.1. Comparison of exact W_q and approximate W_q^* for the T policy M/M(M,M)/1 and M/D(D,D)/1 queues with the first moment of the queue length.

	M/M(M,M)/1						M/D(D,D)/1					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. $(\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0)$											
0.2	2.9380	2.9248	0.4487	5.4329	5.4114	0.3957	2.7971	2.8878	3.2408	5.2971	5.3794	1.5540
0.4	3.3776	3.3483	0.8673	5.8726	5.8267	0.7821	3.0169	3.1968	5.9607	5.5169	5.6801	2.9575
0.6	4.2758	4.2229	1.2378	6.7708	6.6929	1.1502	3.4660	3.7313	7.6530	5.9660	6.2063	4.0271
0.8	7.1307	7.0220	1.5234	9.6256	9.4837	1.4742	4.8935	5.2281	6.8386	7.3935	7.6948	4.0754
μ	Case 2. $(\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0)$											
0.5	5.8660	5.7870	1.3466	8.3609	8.2570	1.2432	4.2611	4.8183	13.0769	6.7611	7.2933	7.8718
1.0	3.1256	3.1049	0.6626	5.6206	5.5874	0.5904	2.8909	3.0265	4.6872	5.3909	5.5140	2.2817
1.5	2.8515	2.8397	0.4126	5.3464	5.3263	0.3756	2.7539	2.8113	2.0868	5.2539	5.3030	0.9352
2.0	2.7695	2.7613	0.2959	5.2645	5.2501	0.2741	2.7129	2.7436	1.1319	5.2129	5.2374	0.4692
α	Case 3. $(\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0)$											
0.05	3.1256	3.1049	0.6626	5.6206	5.5874	0.5904	2.8909	3.0265	4.6872	5.3909	5.5140	2.2817
0.10	3.1462	3.1044	1.3284	5.6411	5.5744	1.1832	2.9012	3.0220	4.1645	5.4012	5.4970	1.7741
0.15	3.1672	3.104	1.9973	5.6622	5.5615	1.7782	2.9117	3.0177	3.6387	5.4117	5.4802	1.2648
0.20	3.1888	3.1037	2.669	5.6838	5.5488	2.3753	2.9225	3.0134	3.1097	5.4225	5.4634	0.7539
β	Case 4. $(\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0)$											
2.0	3.1376	3.1064	0.9950	5.6326	5.5827	0.8865	2.8969	3.0251	4.4246	5.3969	5.5064	2.0276
3.0	3.1256	3.1049	0.6626	5.6206	5.5874	0.5904	2.8909	3.0265	4.6872	5.3909	5.5140	2.2817
4.0	3.1201	3.1047	0.4967	5.6151	5.5903	0.4426	2.8882	3.0273	4.8177	5.3882	5.5180	2.4084
5.0	3.1170	3.1046	0.3972	5.6120	5.5921	0.3540	2.8866	3.0280	4.8958	5.3866	5.5205	2.4844
γ	Case 5. $(\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0)$											
2.0	3.2213	3.2001	0.6578	5.7105	5.6768	0.5890	2.9743	3.1094	4.5418	5.4743	5.5969	2.2393
3.0	3.1256	3.1049	0.6626	5.6206	5.5874	0.5904	2.8909	3.0265	4.6872	5.3909	5.5140	2.2817
4.0	3.0795	3.0590	0.6651	5.5766	5.5436	0.5912	2.8493	2.9850	4.7630	5.3493	5.4725	2.3033
5.0	3.0524	3.0321	0.6665	5.5505	5.5177	0.5916	2.8243	2.9601	4.8096	5.3243	5.4476	2.3165

Table 5.2. Comparison of exact W_q and approximate W_q^* for the T policy M/M(M,M)/1 and M/D(D,D)/1 queues with the second moment of the queue length.

	M/M(M,M)/1						M/D(D,D)/1					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	2.9380	2.4113	17.9259	5.4329	4.3457	20.0120	2.7971	2.3352	16.5134	5.2971	4.2826	19.1522
0.4	3.3776	3.1851	5.7009	5.8726	5.3318	9.2091	3.0169	2.9650	1.7229	5.5169	5.1444	6.7537
0.6	4.2758	4.2200	1.3039	6.7708	6.3699	5.9205	3.4660	3.6093	4.1346	5.9660	5.8317	2.2516
0.8	7.1307	7.2423	1.5663	9.6256	9.2248	4.1640	4.8935	5.1280	4.7924	7.3935	7.2667	1.7141
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	5.8660	6.1944	5.5987	8.3609	8.2019	1.9021	4.2611	4.8731	14.3622	6.7611	6.9833	3.2869
1.0	3.1256	2.8071	10.1918	5.6206	4.8890	13.0167	2.8909	2.6753	7.4599	5.3909	4.7784	11.3621
1.5	2.8515	2.3752	16.7033	5.3464	4.4470	16.8223	2.7539	2.3247	15.5840	5.2539	4.4079	16.1013
2.0	2.7695	2.2199	19.8454	5.2645	4.2831	18.6423	2.7129	2.1894	19.2948	5.2129	4.2612	18.2561
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	3.1256	2.8071	10.1918	5.6206	4.8890	13.0167	2.8909	2.6753	7.4599	5.3909	4.7784	11.3621
0.10	3.1462	2.8136	10.5705	5.6411	4.8845	13.4122	2.9012	2.6760	7.7622	5.4012	4.7686	11.7118
0.15	3.1672	2.8203	10.9541	5.6622	4.8802	13.8099	2.9117	2.6768	8.0686	5.4117	4.7589	12.0632
0.20	3.1888	2.8272	11.3426	5.6838	4.8761	14.2100	2.9225	2.6777	8.3789	5.4225	4.7493	12.4165
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	3.1376	2.3125	26.2994	5.6326	4.0656	27.8209	2.8969	2.6767	7.6015	5.3969	4.7745	11.5324
3.0	3.1256	2.3076	26.1719	5.6206	4.0649	27.6784	2.8909	2.6753	7.4599	5.3909	4.7784	11.3621
4.0	3.1201	2.3057	26.1037	5.6151	4.0651	27.6049	2.8882	2.6748	7.3872	5.3882	4.7806	11.2760
5.0	3.1170	2.3047	26.0615	5.6120	4.0653	27.5601	2.8866	2.6747	7.3429	5.3866	4.7820	11.2241
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	3.2213	2.8870	10.3778	5.7105	4.9703	12.9622	2.9743	2.7411	7.8406	5.4743	4.8522	11.3630
3.0	3.1256	2.8071	10.1918	5.6206	4.8890	13.0167	2.8909	2.6753	7.4599	5.3909	4.7784	11.3621
4.0	3.0795	2.7698	10.0576	5.5766	4.8497	13.0346	2.8493	2.6431	7.2362	5.3493	4.7418	11.3569
5.0	3.0524	2.7483	9.9629	5.5505	4.8266	13.0422	2.8243	2.6240	7.0903	5.3243	4.7199	11.3521

Table 5.3. Comparison of exact W_q and approximate W_q^* for the T policy $M/E_3(E_4,E_3)/1$ and $M/M(E_3,E_2)/1$ queues with the first moment of the queue length.

	$M/E_3(E_4,E_3)/1$						$M/M(E_3,E_2)/1$					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	2.8440	2.9000	1.9704	5.3423	5.3900	0.8931	2.9323	2.9191	0.4489	5.4298	5.4083	0.3957
0.4	3.1370	3.2471	3.5104	5.6353	5.7288	1.6586	3.3712	3.3419	0.8677	5.8687	5.8228	0.7822
0.6	3.7356	3.8948	4.2617	6.2339	6.3681	2.1530	4.2677	4.2149	1.2382	6.7652	6.6874	1.1503
0.8	5.6382	5.8251	3.3149	8.1365	8.2901	1.8876	7.1175	7.0091	1.5238	9.6150	9.4732	1.4743
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	4.7957	5.1409	7.1971	7.2940	7.6142	4.3895	5.8579	5.7790	1.3471	8.3554	8.2515	1.2434
1.0	2.9691	3.0525	2.8100	5.4674	5.5383	1.2975	3.1196	3.0990	0.6629	5.6171	5.5839	0.5905
1.5	2.7863	2.8207	1.2343	5.2847	5.3107	0.4932	2.8458	2.8340	0.4127	5.3433	5.3232	0.3756
2.0	2.7317	2.7495	0.6494	5.2301	5.2415	0.2198	2.7640	2.7558	0.2960	5.2615	5.2471	0.2742
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	2.9691	3.0525	2.8100	5.4674	5.5383	1.2975	3.1196	3.0990	0.6629	5.6171	5.5839	0.5905
0.10	2.9827	3.0493	2.2335	5.4810	5.5226	0.7596	3.1393	3.0976	1.3292	5.6368	5.5701	1.1833
0.15	2.9966	3.0461	1.6536	5.4949	5.5070	0.2198	3.1596	3.0964	1.9985	5.6571	5.5565	1.7784
0.20	3.0109	3.0431	1.0703	5.5092	5.4915	0.3220	3.1804	3.0954	2.6708	5.6779	5.5430	2.3757
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	2.9770	3.0520	2.5210	5.4753	5.5316	1.0285	3.1306	3.0995	0.9956	5.6281	5.5782	0.8866
3.0	2.9691	3.0525	2.8100	5.4674	5.5383	1.2975	3.1196	3.0990	0.6629	5.6171	5.5839	0.5905
4.0	2.9655	3.0531	2.9540	5.4638	5.5420	1.4318	3.1145	3.0990	0.4969	5.6120	5.5871	0.4427
5.0	2.9634	3.0535	3.0401	5.4617	5.5443	1.5123	3.1115	3.0992	0.3974	5.6090	5.5891	0.3540
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	3.0565	3.1395	2.7154	5.5529	5.6234	1.2698	3.2091	3.188	0.6584	5.7037	5.6701	0.5891
3.0	2.9691	3.0525	2.8100	5.4674	5.5383	1.2975	3.1196	3.099	0.6629	5.6171	5.5839	0.5905
4.0	2.9259	3.0096	2.8589	5.4250	5.4961	1.3116	3.0757	3.0553	0.6653	5.5743	5.5413	0.5912
5.0	2.9002	2.9840	2.8886	5.3996	5.4709	1.3201	3.0497	3.0294	0.6667	5.5487	5.5159	0.5916

Table 5.4. Comparison of exact W_q and approximate W_q^* for the T policy $M/E_3(E_4,E_3)/1$ and $M/M(E_3,E_2)/1$ queues with the second moment of the queue length.

	$M/E_3(E_4,E_3)/1$						$M/M(E_3,E_2)/1$					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	2.8440	2.3601	17.0161	5.3423	4.3031	19.4523	2.9323	2.4047	17.9936	5.4298	4.3419	20.0361
0.4	3.1370	3.0358	3.2263	5.6353	5.2046	7.6445	3.3712	3.1777	5.7384	5.8687	5.3272	9.2262
0.6	3.7356	3.8050	1.8578	6.2339	6.0034	3.6981	4.2677	4.2111	1.3262	6.7652	6.3637	5.9353
0.8	5.6382	5.8136	3.1111	8.1365	7.8945	2.9743	7.1175	7.2283	1.5561	9.6150	9.2132	4.1789
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	4.7957	5.3012	10.5402	7.2940	7.3744	1.1016	5.8579	6.1859	5.5987	8.3554	8.1957	1.9111
1.0	2.9691	2.7180	8.4575	5.4674	4.8141	11.9485	3.1196	2.8001	10.2412	5.6171	4.8848	13.0365
1.5	2.7863	2.3412	15.9772	5.2847	4.4206	16.3494	2.8458	2.3685	16.7732	5.3433	4.4432	16.8444
2.0	2.7317	2.1994	19.4860	5.2301	4.2684	18.3877	2.7640	2.2133	19.9243	5.2615	4.2794	18.6653
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	2.9691	2.7180	8.4575	5.4674	4.8141	11.9485	3.1196	2.8001	10.2412	5.6171	4.8848	13.0365
0.10	2.9827	2.7204	8.7912	5.4810	4.8059	12.3160	3.1393	2.8057	10.6280	5.6368	4.8795	13.4361
0.15	2.9966	2.7230	9.1293	5.4949	4.7979	12.6855	3.1596	2.8114	11.0196	5.6571	4.8742	13.8381
0.20	3.0109	2.7257	9.4719	5.5092	4.7899	13.0571	3.1804	2.8173	11.4159	5.6779	4.8692	14.2423
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	2.9770	2.2502	24.4134	5.4753	4.0131	26.7055	3.1306	2.3054	26.3590	5.6281	4.0610	27.8446
3.0	2.9691	2.2475	24.3020	5.4674	4.0145	26.5736	3.1196	2.3016	26.2212	5.6171	4.0614	27.6967
4.0	2.9655	2.2465	24.2439	5.4638	4.0155	26.5065	3.1145	2.3001	26.1495	5.6120	4.0619	27.6214
5.0	2.9634	2.2460	24.2082	5.4617	4.0162	26.4658	3.1115	2.2992	26.1057	5.6090	4.0623	27.5758
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	3.0565	2.7885	8.7688	5.5529	4.8904	11.9301	3.2091	2.8729	10.4768	5.7037	4.9622	13.0004
3.0	2.9691	2.7180	8.4575	5.4674	4.8141	11.9485	3.1196	2.8001	10.2412	5.6171	4.8848	13.0365
4.0	2.9259	2.6841	8.2662	5.4250	4.7766	11.9514	3.0757	2.7654	10.0896	5.5743	4.8470	13.0478
5.0	2.9002	2.6642	8.1389	5.3996	4.7543	11.9511	3.0497	2.7451	9.9868	5.5487	4.8245	13.0523

Table 5.5. Comparison of exact W_q and approximate W_q^* for the T policy M/E_3(E_4,D)/1 and M/E_3(E_4,M)/1 queues with the first moment of the queue length.

	M/E_3(E_4,D)/1						M/E_3(E_4,M)/1					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu = 1.0,\ \alpha = 0.05,\ \beta = 3.0,\ \gamma = 3.0$)											
0.2	2.8405	2.8966	1.9732	5.3405	5.3882	0.8935	2.8509	2.9070	1.9648	5.3459	5.3936	0.8923
0.4	3.1336	3.2437	3.5151	5.6336	5.7270	1.6593	3.1440	3.2540	3.5012	5.6389	5.7324	1.6571
0.6	3.7321	3.8914	4.2666	6.2321	6.3664	2.1539	3.7425	3.9017	4.2519	6.2375	6.3717	2.1512
0.8	5.6347	5.8217	3.3177	8.1347	8.2883	1.8883	5.6451	5.8319	3.3091	8.1401	8.2936	1.8862
μ	Case 2. ($\lambda = 0.3,\ \alpha = 0.05,\ \beta = 3.0,\ \gamma = 3.0$)											
0.5	4.7922	5.1374	7.2031	7.2922	7.6124	4.3908	4.8027	5.1477	7.1853	7.2976	7.6177	4.3868
1.0	2.9656	3.0491	2.8139	5.4656	5.5366	1.2981	2.9760	3.0594	2.8023	5.4710	5.5419	1.2963
1.5	2.7829	2.8173	1.2362	5.2829	5.3089	0.4935	2.7933	2.8276	1.2304	5.2882	5.3143	0.4926
2.0	2.7283	2.7460	0.6506	5.2283	5.2398	0.2199	2.7387	2.7564	0.6471	5.2336	5.2451	0.2195
α	Case 3. ($\lambda = 0.3,\ \mu = 1.0,\ \beta = 3.0,\ \gamma = 3.0$)											
0.05	2.9656	3.0491	2.8139	5.4656	5.5366	1.2981	2.9760	3.0594	2.8023	5.4710	5.5419	1.2963
0.10	2.9792	3.0458	2.2373	5.4792	5.5208	0.7602	2.9896	3.0562	2.2260	5.4846	5.5262	0.7585
0.15	2.9931	3.0427	1.6572	5.4931	5.5052	0.2203	3.0035	3.0530	1.6463	5.4985	5.5105	0.2186
0.20	3.0074	3.0397	1.0739	5.5074	5.4897	0.3215	3.0179	3.0499	1.0633	5.5128	5.4950	0.3231
β	Case 4. ($\lambda = 0.3,\ \mu = 1.0,\ \alpha = 0.05,\ \gamma = 3.0$)											
2.0	2.9735	3.0486	2.5248	5.4735	5.5298	1.0290	2.9839	3.0589	2.5133	5.4789	5.5351	1.0273
3.0	2.9656	3.0491	2.8139	5.4656	5.5366	1.2981	2.9760	3.0594	2.8023	5.4710	5.5419	1.2963
4.0	2.9620	3.0496	2.9579	5.4620	5.5402	1.4324	2.9724	3.0600	2.9462	5.4674	5.5456	1.4306
5.0	2.9599	3.0500	3.0440	5.4599	5.5425	1.5128	2.9703	3.0604	3.0323	5.4653	5.5479	1.5111
γ	Case 5. ($\lambda = 0.3,\ \mu = 1.0,\ \alpha = 0.05,\ \beta = 3.0$)											
2.0	3.0489	3.1320	2.7233	5.5489	5.6195	1.2711	3.0717	3.1546	2.6995	5.5608	5.6313	1.2673
3.0	2.9656	3.0491	2.8139	5.4656	5.5366	1.2981	2.9760	3.0594	2.8023	5.4710	5.5419	1.2963
4.0	2.9239	3.0076	2.8611	5.4239	5.4951	1.3119	2.9299	3.0135	2.8543	5.4270	5.4981	1.3109
5.0	2.8989	2.9827	2.8901	5.3989	5.4702	1.3203	2.9028	2.9866	2.8856	5.4009	5.4722	1.3197

Table 5.6. Comparison of exact W_q and approximate W_q^* for the T policy M/E_3(E_4,D)/1 and M/E_3(E_4,M)/1 queues with the second moment of the queue length.

	M/E_3(E_4,D)/1						M/E_3(E_4,M)/1					
	T =5			T =10			T =5			T =10		
	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error	W_q	W_q^*	%Error
λ	Case 1. ($\mu=1.0,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.2	2.8405	2.3560	17.0567	5.3405	4.3009	19.4659	2.8509	2.3682	16.9332	5.3459	4.3075	19.4248
0.4	3.1336	3.0319	3.2443	5.6336	5.2024	7.6526	3.1440	3.0438	3.1873	5.6389	5.2088	7.6276
0.6	3.7321	3.8012	1.8513	6.2321	6.0013	3.7037	3.7425	3.8127	1.8739	6.2375	6.0076	3.6861
0.8	5.6347	5.8102	3.1148	8.1347	7.8926	2.9768	5.6451	5.8205	3.1056	8.1401	7.8984	2.9688
μ	Case 2. ($\lambda=0.3,\ \alpha=0.05,\ \beta=3.0,\ \gamma=3.0$)											
0.5	4.7922	5.2977	10.5476	7.2922	7.3724	1.0997	4.8027	5.3082	10.5272	7.2976	7.3783	1.1059
1.0	2.9656	2.7140	8.4845	5.4656	4.8120	11.9587	2.9760	2.7260	8.4012	5.4710	4.8184	11.9276
1.5	2.7829	2.3371	16.0190	5.2829	4.4185	16.3620	2.7933	2.3494	15.8916	5.2882	4.4250	16.3238
2.0	2.7283	2.1953	19.5346	5.2283	4.2662	18.4013	2.7387	2.2077	19.3871	5.2336	4.2727	18.3601
α	Case 3. ($\lambda=0.3,\ \mu=1.0,\ \beta=3.0,\ \gamma=3.0$)											
0.05	2.9656	2.7140	8.4845	5.4656	4.8120	11.9587	2.9760	2.7260	8.4012	5.4710	4.8184	11.9276
0.10	2.9792	2.7165	8.8176	5.4792	4.8038	12.3261	2.9896	2.7284	8.7361	5.4846	4.8102	12.2954
0.15	2.9931	2.7191	9.1551	5.4931	4.7957	12.6955	3.0035	2.7310	9.0755	5.4985	4.8021	12.6652
0.20	3.0074	2.7218	9.4970	5.5074	4.7878	13.0669	3.0179	2.7336	9.4192	5.5128	4.7941	13.0371
β	Case 4. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \gamma=3.0$)											
2.0	2.9735	2.2467	24.4404	5.4735	4.0112	26.7149	2.9839	2.2571	24.3578	5.4789	4.0167	26.6864
3.0	2.9656	2.2441	24.3293	5.4656	4.0127	26.5831	2.9760	2.2545	24.2458	5.4710	4.0182	26.5543
4.0	2.9620	2.2431	24.2714	5.4620	4.0137	26.5160	2.9724	2.2535	24.1874	5.4674	4.0192	26.4871
5.0	2.9599	2.2426	24.2358	5.4599	4.0144	26.4753	2.9703	2.2530	24.1515	5.4653	4.0199	26.4464
γ	Case 5. ($\lambda=0.3,\ \mu=1.0,\ \alpha=0.05,\ \beta=3.0$)											
2.0	3.0489	2.7798	8.8285	5.5489	4.8857	11.9525	3.0717	2.8062	8.6426	5.5608	4.9000	11.8837
3.0	2.9656	2.7140	8.4845	5.4656	4.8120	11.9587	2.9760	2.7260	8.4012	5.4710	4.8184	11.9276
4.0	2.9239	2.6818	8.2815	5.4239	4.7754	11.9572	2.9299	2.6886	8.2346	5.4270	4.7790	11.9396
5.0	2.8989	2.6627	8.1487	5.3989	4.7535	11.9548	2.9028	2.6671	8.1187	5.4009	4.7558	11.9435

Chapter 6

Conclusions

In this thesis, we consider the N policy and the T policy for M/G/1 queues with server breakdowns and general distributed startup times, respectively. We develop the theoretical results for various system performance measures, such as the expected number of customers in the system, the expected length of the turned-off, complete startup, busy, and break-down periods, and the expected length of the busy cycle. In the N policy or T policy M/G/1 queue with general service times and startup times, we prove that the probability that the server is busy in the steady-state is equal to the traffic intensity ρ. We construct a cost model to determine the optimal threshold N or T so as to minimize this cost function. We also provide sensitivity analysis to discuss how the system performance measures can be affected by the changes of the input parameters (or cost parameters) in the investigated queueing service model. The sensitivity investigation is particularly valuable when evaluating future condition for the system analyst

We have utilized the maximum entropy principle to develop the maximum entropy (approximate) solutions for the N policy M/G/1 queue with general repair times and startup times. We perform a comparative analysis between the approximate results obtained using maximum entropy principle and established exact results. We have demonstrated that the relative error percentages are very small (below 6.8%). For the T policy M/G/1 queue with server breakdowns and general startup times, we have utilized the maximum entropy principle to develop the two maximum entropy (approximate) solutions for the T policy M/G/1 queue with general repair times and startup times. We perform a comparative analysis between two approximate results obtained using maximum entropy principle with different constraints and exact results. We have demonstrated that the relative error percentages are very small for maximum entropy solutions with the first moment of queue length (below 7.3%) and are very large for maximum entropy with the second moment of queue length. The numerical solutions show that the maximum entropy solutions with the first moment of queue length are much better than the maximum entropy solutions with the second moment of queue length. The numerical results indicate that the use of maximum entropy principle is accurate enough for practical purposes and suitable choosing the constraints. In a word, the maximum entropy principle provides a helpful method for analyzing complex queueing systems. Finally, incorporating the startup failure into a complex queueing system with combined multiple threshold policies is worthy of further investigation.

References

1. Alfa, A. S. and Li, W. (2000). Optimal (*N*, *T*)-policy for M/G/1 system with cost structure. *Performance Evaluation*, 42, 265-277.
2. Ali, O. M. E. and Neuts, M. F. (1984). A service system with two stages of waiting and feedback of customers. *Journal of Applied Probability***,** 21, 404-413.
3. Arizono, I., Cui, Y. and Ohta, H. (1991). An analysis of M/M/S queueing systems based on the maximum entropy principle. *Journal of the Operational Research Society*, 42, 69-73.
4. Artalejo, J. R. and López-Herrero, M. J. (2004). Entropy maximization and the busy period of some single-server vacation models. *RAIRO Operations Research*, 38, 195-213.
5. Baker, K. R. (1973). A note on operating policies for the queue M/M/1 with exponential startup. *INFOR*, 11, 71-72.
6. Balachandran, K. R. (1973). Control Policies for a Single Server System. *Management Science*, 19, 1013-1018.
7. Bell, C. E. (1971). Characterization and computation of optimal policies for operating an M/G/1 queueing system with removable server. *Operations Research*, 19, 208-218.
8. Borthakur, A. Medhi, J. and Gohain, R. (1987). Poisson input queueing systems with startup time and under control operating policy. *Computers and Operations Research*, 14, 33-40.
9. Cooper, R. B. (1970). Queues served in cyclic order: waiting times. *Bell System Technical Journal*, 49, 399-413.
10. Doganata, Y. N. (1990) NT-vacation policy for M/G/1 queue with starter in: E. Arikan (ed.), *Communication, Control, and Signal Processing*, Elsevier Science, Amsterdam 1663-1669.
11. Doshi, B. T. (1986). Queueing systems with vacations – a survey. *Queueing Systems*, 1, 29-66.
12. El-Affendi, M. A. and Kouvatsos, D. D. (1983). A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems at equilibrium. *Acta Informatica*, 19, 339-355.
13. Ferdinand, A. E. (1970). A statistical mechanical approach to system analysis. *IBM Journal of Research and Development*, 14, 539-547.

14. Fuhrmann, S. W. and Cooper, R. B. (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. *Operations Research*, 33, 1117-1129.
15. Gakis, K. G., Rhee, H. K. and Sivazlian, B. D. (1995). Distributions and first moments of the busy and idle periods in controllable M/G/1 queueing models with simple and dyadic policies. *Stochastic Analysis and Applications*, 13, 47-81.
16. Gaver, D. P. (1962). A waiting line with interrupted service, including priorities. *Journal of the Royal Statistical Society,* B24, 73-90
17. Gupta, S. M. (1995). N-policy Queueing System with Finite Population. *Transactions on Operational Research*, 7, 45-62.
18. Gross, D. and Harris, C. M. (1985). *Fundamentals of Queueing Theory*. 2nd ed, John Wiley and Sons, New York.
19. Heyman, D. P. (1968). Optimal operating policies for M/G/1 queueing system. *Operations Research*, 16, 362-382.
20. Heyman, D. P. (1977). The T-policy for the M/G/1 queue. *Management Science*, 23, 775-778.
21. Hur, S. and Paik, S. J. (1999). The effect of different arrival rates on the N-policy of M/G/1 with server setup. *Applied Mathematical Modelling*, 23, 289-299.
22. Hur, S., Kim, J. and Kang, C. (2003). An analysis of the M/G/1 system with N and T policy. *Applied Mathematical Modelling*, 27, 665-675.
23. Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4), 620–630
24. Ke, J.-C. (2003). The optimal control of an M/G/1 queueing system with server vacations, startup and breakdowns. *Computers and Industrial Engineering*, 44, 567-579.
25. Ke, J.-C. (2003). The operating characteristic analysis on a general input queue with N policy and a startup time. *Mathematical Methods of Operations Research*, 57, 235-254.
26. Ke, J.-C. (2005). Modified T vacation policy for M/G/1 queueing system with an unreliable server and startup. *Mathematical and Computer Modeling*, 41, 1267 -1277.
27. Ke, J.-C. and Pearn, W. L. (2004). Optimal management policy for heterogeneous arrival queueing systems with server breakdowns and vacations. *Quality Technology & Quantitative Management*, 1(1), 149-162.

28. Kleinrock, L. (1975). *Queueing Systems*. Vol. I: Theory, John Wiley and Sons, New York.
29. Kouvatsos, D. D. (1986). Maximum entropy and the G/G/1/N queue. *Acta Informatica*, 23, 545-565.
30. Kouvatsos, D. D. (1988). A maximum entropy analysis of the G/G/1 queue at equilibrium. *Journal of the Operational Research Society*, 39, 183-200.
31. Krishna, Reddy G. V., Nadarajan, R. and Arumuganathan, R. (1998). Analysis of a bulk queue with N-policy multiple vacations and setup times. *Computers and Operations Research*, 25, 957-967.
32. Lee, H. W. and Park, J. O. (1997). Optimal strategy in N-policy production system with early set-up. *Journal of the Operational Research Society*, 48, 306-313.
33. Levy, Y. and Yechiali, U. (1975). Utilization of idle time in an M/G/1 queueing system. *Management Science*, 22, 202-211.
34. Medhi, J. and Templeton, J. G. C. (1992). A Poisson input queue under N-policy and with a general start up time. *Computers and Operations Research*, 19, 35-41.
35. Pearn, W. L., Ke, J.-C. and Chang, Y. C. (2004). Sensitivity analysis of optimal management policy for a queueing system with a removable and non-reliable server. *Computers and Industrial Engineering*, 46, 87-99.
36. Ross, S. M. (2003). *Introduction to Probability Models*. 8th ed, Academic Press.
37. Scholl, M. and Kleinrock, L. (1983). On the M/G/l queue with rest period and certain service-independent queueing disciplines. *Operations Research*, 31, 705-719.
38. Shanthikumar, J. G. (1980). Some analyses of the control of queues using level crossing of regenerative processes. *Journal of Applied Probability*, 17, 814-821.
39. Shore, J. E. (1978). Derivation of equilibrium and time-dependent solutions to M/M/∞/N and M/M/∞ queueing systems using entropy maximization. *In Proceedings, National Computer Conference, AFIPS*, 483-487.
40. Shore, J. E (1982). Information theoretic approximations for M/G/1 and G/G/1 queueing systems. *Acta Informatica*, 17, 43-61
41. Tadj, L. (2003). On an M/G/1 quorum queueing system under T policy. *Journal of the Operational Research Society*, 54, 466-471.
42. Tadj, L. and Choudhury G. (2005). Optimal design and control of queues. *Top*, 13, 359-414.

43. Takagi H. (1991). *Queueing Analysis: A Foundation of Performance Evaluation* (Volume 1). North-Holland.

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

45. Teghem, Jr. J. (1986). Control of the service process in a queueing system. *European Journal of Operational Research*, 23, 141-158.

46. Tijms, H. C. (1986). *Stochastic Modelling and Analysis: a computational approach*. John Wiley and Sons., New York.

47. Wang, K.-H. (1995). Optimal operation of a Markovian queueing system with a removable and non-reliable server. *Microelectronics and Reliability*, 35, 1131-1136.

48. Wang, K.-H. (1997). Optimal control of an $M/E_k/1$ queueing system with removable service station subject to breakdowns. *Journal of the Operational Research Society*, 48, 936-942.

49. 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.

50. Wang, K.-H., Chang, K.-W. and Sivazlian, B. D. (1999). Optimal control of a removable and non-reliable server in an infinite and a finnite $M/H_2/1$ queueing system. *Applied Mathematical Modelling*, 23, 651-666.

51. Wang, K.-H. and Huang, H.-M. (1995). Optimal control of an $M/E_k/1$ queueing system with a removable service station. *Journal of the Operational Research Society*, 46, 1014-1022.

52. Wang, K.-H., Kao, H.-T. and Chen, G. (2004). Optimal management of a removable and non-reliable server in an infinite and a finite $M/H_k/1$ queueing system. *Quality Technology & Quantitative Management*, 1(2), 325-339.

53. Wang, K.-H. and Ke, J.-C. (2000). A recursive method to the optimal control of an M/G/1 queueing system with finite capacity and infinite capacity. *Applied Mathematical Modelling*, 24, 899-914.

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

55. K.-H. Wang, S.-L. Chuang, and W.-L. Pearn (2002). Maximum entropy analysis to the N policy M/G/1 queueing system with a removable server. *Applied Mathematical Modelling*, 26, 1151-1162.
56. Wang, K.-H. and Yen, K.-L. (2002). Optimal control of an $M/H_k/1$ queueing system with a re-movable server. *Mathematical Methods of Operations Research*, 57, 255-262.
57. Wu, J.-S. and Chan, W.C. (1989). Maximum entropy analysis of multiple-server queueing systems. *Journal of the Operational Research Society*, 40, 815–825.
58. Yadin, M. and Naor, P. (1963). Queueing systems with a removable service station. *Operational Research Quarterly*, 14, 393-405.