INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Optimization of queueing systems with service interruptions

Araar, Abdelaziz, Ph.D.

Case Western Reserve University, 1991
We hereby approve the thesis of
Abdelaziz Araar

candidate for the Ph.D. degree.*

Signed: (Chairman)

Date September 13, 1990

*We also certify that written approval has
been obtained for any proprietary material
contained therein.
I grant to Case Western Reserve University the right to use this work, irrespective of any copyright, for the University's own purposes without cost to the University or to its students, agents and employees. I further agree that the University may reproduce and provide single copies of the work, in any format other than in or from microforms, to the public for the cost of reproduction.

[Signature]

[Signature]
OPTIMIZATION OF QUEUEING SYSTEMS WITH SERVICE INTERRUPTIONS

Abstract

by

ABDELAZIZ ARAAR

The problem of determining optimal stationary maintenance intervals that minimize the long-run average operating costs for queueing systems with non-negligible maintenance and repair times is very complex. This complexity is due to the intractability of some queueing performance measures such as waiting times, etc. A special case of the problem in which servers are under the heavy traffic condition ( $\rho \geq 0.9$ ) is investigated.

Different queueing systems are analyzed with the $(1, t_0)$ maintenance policy, i.e., maintenance is conducted at age $t_0$ or upon failure, whichever occurs first. These systems include $M/G/1$ queues, and $M/M/1$ queues with waiting space and quality degradation costs. The research also covers unreliable open queueing networks with maintenance crew-size cost. Nonlinear integer programming models are formulated and efficient algorithms are developed by investigating properties of these models. We show how the total expected costs under such policy may be computed and how the optimal policy may be determined efficiently.
Deep in my heart, I recognize the help of God in making this work possible.
Dedication

To my father Ali who sacrificed most to make this happen.

To my wife Fatma, and my two kids Omar and Rahma.

To all my Family.
ACKNOWLEDGEMENT

I would like to express my sincere thanks and my gratitude to my advisor Dr. Henry Co for his invaluable assistance in the work research and for his overall supervision of my thesis. I would like to acknowledge Dr. Daniel Solow, Dr. John Wilson and Dr. Miles Kennedy for serving on my dissertation committee and for their generous help. I would like to thank the Ministry of Higher Education for granting me a scholarship. Finally, I thank faculty, staff and my fellow graduate students who made an enjoyable atmosphere in the Operations Research Department.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>v</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES AND TABLES</td>
<td>ix</td>
</tr>
<tr>
<td><strong>CHAPTER ONE</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Problem Definition</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Model Characteristics</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Queueing Systems with Service Interruptions</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Preventive Maintenance Models</td>
<td>9</td>
</tr>
<tr>
<td>1.3.3 Unreliable Queueing Networks</td>
<td>12</td>
</tr>
<tr>
<td>1.3.4 Conclusion and Extension</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Thesis Organization</td>
<td>15</td>
</tr>
<tr>
<td><strong>CHAPTER TWO</strong></td>
<td></td>
</tr>
<tr>
<td>M/G/1 Queues with Age-Maintenance</td>
<td>17</td>
</tr>
<tr>
<td>Notation</td>
<td>18</td>
</tr>
<tr>
<td>2.1 Background</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Problem Definition</td>
<td>20</td>
</tr>
<tr>
<td>2.3 The Breakdown/Maintenance Process</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1 The On-Period</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 The Off-Period</td>
<td>25</td>
</tr>
<tr>
<td>2.3.3 The On/Off Cycle</td>
<td>26</td>
</tr>
<tr>
<td>2.4 The Expected Waiting Time of Jobs in the System</td>
<td>28</td>
</tr>
<tr>
<td>2.4.1 The Diffusion Approximation</td>
<td>28</td>
</tr>
<tr>
<td>2.4.2 The Expected Waiting Time of an M/G/1 Queueing Systems</td>
<td>30</td>
</tr>
<tr>
<td>2.4.3 The Expected Waiting Time of Jobs in the System under Maintenance and Repair</td>
<td>32</td>
</tr>
</tbody>
</table>
4.5.2 Priority Queues Method ........................................... 107
4.5.3 Diffusion Approximation Method ............................... 108
4.5.4 Equivalent Restricted Network of Queues ...................... 109
4.5.5 Simulation .............................................................. 112
4.5.6 Summary ............................................................... 113
4.6 Maintenance Optimization Model .................................. 113
4.6.1 Case 1: Cost of Waiting is Positive ......................... 113
4.6.2 Case 2: No Cost of Waiting ..................................... 117
  4.6.2.1 Linear Approximation to the Standard Deviation ........ 117
  4.6.2.2 Optimal Age-Maintenance Policy ......................... 120
  4.6.2.3 Derivation of Algorithm 4.1 ............................ 126
  4.6.2.4 Optimal Maintenance Crew Size ....................... 127
  4.6.2.4 Derivation of Algorithm 4.2 ............................ 131
4.6.3 Computational Tests for Cases 2 .............................. 132
  4.6.3.1 Illustration .................................................... 137
  4.6.3.2 GAMS ......................................................... 145
  4.6.3.3 Comparizon of Algorithm 4.1 with GAMS ............... 147
  4.6.3.4 Optizing Crew Size ..................................... 150
  4.6.3.5 Sensitivity of N* .......................................... 154
4.7 Conclusion ............................................................. 155

CHAPTER FIVE

Further Research and Conclusion ........................................... 156

5.1 Summary and Conclusion ........................................... 157
5.2 Further Research ..................................................... 159
  5.2.1 Medium Traffic Condition ................................... 159
  5.2.2 Queueing Systems without the Like-New Assumption .... 159
  5.2.3 Quality Control ............................................... 160
  5.2.4 Unreliable Queueing Networks ................................ 161
  5.2.5 Maintenance Crew size with different skills ............. 161

REFERENCES ............................................................... 163
# LIST OF FIGURES AND TABLES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Basic Components of Waiting Line Process</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>General Types of Interruptions</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>General Representation of the Model</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Classification of Different Types of Service-Interruption Interaction</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Queueing Systems with Age-Maintenance Policy</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Queueing Systems with the Age-Maintenance Policy</td>
<td>21</td>
</tr>
<tr>
<td>2.2</td>
<td>Queueing System with Alternate ON and OFF Periods</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Truncated Arrival Process Pdf</td>
<td>24</td>
</tr>
<tr>
<td>2.4</td>
<td>Representation of the Model</td>
<td>26</td>
</tr>
<tr>
<td>2.5</td>
<td>Sensitivity of Error with Traffic Intensity</td>
<td>39</td>
</tr>
<tr>
<td>2.6</td>
<td>Comparison of the Expected Waiting Times</td>
<td>41</td>
</tr>
<tr>
<td>2.7</td>
<td>Sensitivity of $E[W]$ vs. Traffic Intensity</td>
<td>49</td>
</tr>
<tr>
<td>2.8</td>
<td>Comparison of the Two Optimal Ages</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>M/M/1 Queue with Quality Control and Limited Waiting Capacity</td>
<td>65</td>
</tr>
<tr>
<td>3.2</td>
<td>Expected Waiting Time as a Function of Age-Maintenance</td>
<td>73</td>
</tr>
<tr>
<td>3.3</td>
<td>Variation of $\delta (L)$ with $L$</td>
<td>76</td>
</tr>
<tr>
<td>3.4</td>
<td>Possible Cases of the Feasible Region with $t^*_1$</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Total Cost of [SUB1] vs. buffer</td>
<td>84</td>
</tr>
<tr>
<td>3.6</td>
<td>Prob of non-def. vs. maintenance</td>
<td>86</td>
</tr>
<tr>
<td>3.7</td>
<td>Total cost of [SUB2] vs. # of defectives</td>
<td>91</td>
</tr>
<tr>
<td>3.8</td>
<td>Sequenced couples of $S$</td>
<td>93</td>
</tr>
</tbody>
</table>
Table 4.6  Optimal Age-Maintenance Policies with Different Values of N for $C_m = $10.......................... 138
Table 4.7  Optimal Age-Maintenance Policies with Different Values of N for $C_m = $50......................... 139
Table 4.8  Optimal Age-Maintenance Policies with Different Value of N for $C_m = $100........................... 140
Table 4.9  Optimal Age-Maintenance Policies with Different Values of N for $C_m = $200.......................... 141
Table 4.10 Optimal Age-Maintenance Policies with Different Values of N for $C_m = $300......................... 142
Table 4.11 Optimal Age-Maintenance Policies with Different Values of N for $C_m = $400......................... 143
Table 4.12 Optimal Age-Maintenance Policies with Different Values of N for $C_m = $500......................... 144
Table 4.13 Algorithm 4.1 vs. GAMS.......................... 148
CHAPTER ONE:

Introduction
1.1 BACKGROUND

Waiting lines are one of the most common occurrences of everyday life. Waiting line may exist in the form of an observable line of customers, or jobs waiting for processing. The study of waiting lines, known as queueing theory, is far from new. Queueing theory can be traced back to the work of A. K. Erlang in 1913, when the Danish mathematician studied the fluctuating demands on the telephone exchange. The basic components of a queueing process are: arrivals, waiting line, service and departure (see Figure 1.1).

![Diagram](image)

Figure 1.1: Basic components of waiting line process

A queueing system can be identified by the Kendall notation $a/b/c/N$, where $a$ denotes the arrival distribution, $b$ the service time distribution, $c$ the number of servers in the system, and $N$ the size of the waiting capacity. Furthermore, the following symbols are used to denote the various probability distributions: $M$ — exponential distribution, $D$ — regular distribution, $E_k$ — $k$-Erlang distribution, and $G$ — general distribution with no specification regarding its form.
Literature on queueing theory typically studies the system at steady state. The duration of the service of individual customers are assumed to be independent and identically distributed non-negative random variables. The service process is also assumed to be independent of the arrival process. The traffic intensity $\rho$ indicates the likelihood and extent of the queue. It is the product of the mean arrival rate $\lambda$ and the mean service time $E[S]$, where $\rho = \lambda E[S]$.

When the traffic intensity is less than unity ($\rho < 1$), the probability that a finite number of customers waiting in the system has a stationary distribution, and the mean waiting time is positive and finite. The focal point of queueing theory is, of course, the queue itself. The extent of which queues form depends primarily upon the nature of the arrival and service processes. However, another factor affecting the waiting line is the interruption of the service process. In the real world, customers often experience service interruptions. Such interruptions may result from (i) unscheduled server breakdowns; (ii) scheduled off-periods, such as changing shifts; or (iii) arrival of customers of a higher priority class. Figure 1.2 outlines the major effects which cause service interruptions. Unfortunately, most models in the literature do not address the issue of reducing the occurrences of such interruptions.
1.2 PROBLEM DEFINITION

This dissertation examines queueing systems with service interruptions. These interruptions are assumed to be caused by random breakdowns. The service facility may break down for several reasons. For example, an input rush may create a great pressure on the server(s), causing the shutting down of the system. In manufacturing, machines are subject to deterioration with age and usage. A radical change in service, or arrival of jobs with higher priorities may also cause service interruptions.

Unlike queues with vacations, the timing of service interruptions is random and independent of the number of customers waiting in the system.
The impact of breakdowns on the performance of a given queueing system depends on the specific interaction between the repair process and the service process. Such a system, which is presented in Figure 1.3, can be modeled as a queueing system with "on-time periods" and "off-time periods".

![Diagram](image)

**Figure 1.3:** General representation of the model

Queueing models subject to service interruptions are useful in many applications. For example, in computer systems performance modeling, interruption may occur due to hardware failure, fault in the operating system, or arrival of a higher-priority user. Whenever an interruption occurs, the current activity is preempted. The duration of the interruption is the time needed for repair. At the end of the interruption, service on the preempted activity is resumed or repeated. The following is a classification of the different types of service interruption:
The analysis of job completion time on such systems is of much interest. The analysis provides job-oriented performance and reliability measures such as the mean time to complete a job and the probability that job completion time will not exceed a specified time limit (a deadline).

Suppose customers (or jobs) arrive at a service facility at random. If the server is working without failure for age $t_0$, we stop the operation of the working unit for preventive maintenance, and the cycle is repeated. On the other hand, if the server fails before $t_0$, it will be repaired. Figure 1.5 illustrates the
implementation of the age-maintenance policy to "slow" the deterioration process. After each maintenance or repair performed, the queueing system is assumed to be in "like-new" condition.

![Diagram]

Figure 1.5: Queueing system with age-maintenance policy

Whenever breakdown occurs, the waiting line builds up. Consequently, there are losses of time and money. Unscheduled service interruption is often reflected in higher production costs and lower product quality.

The first part of this research presents two nonlinear programming models, NLP1 and INLP1, for the preventive maintenance problem. NLP1 is a model of the age-maintenance policy for an unreliable single-server queueing system, while INLP1 is a model of the age-maintenance policy for an unreliable single-server queueing system where the quality of the output declines over time, and there is a cost associated with buffer capacity. The second part of the research presents an integer nonlinear programming model INLP2.
INLP2 is a model of the age-maintenance policy for an unreliable open queueing network, where the maintenance crew size is also a decision variable. The purpose of this study is to develop efficient algorithms for solving the nonlinear programming formulations.

1.3 MODEL CHARACTERISTICS

In this section, we briefly survey literature in three areas relevant to our research:

1.3.1 Queueing systems with service interruptions

Queueing models with interruptions and their connection with priority-queue models date back to a study by White and Christie (1958). The authors considered the case where the service time, the on-time and the off-time are exponentially distributed. Keilson (1962) extended the results to models with general off-time distribution but exponential on-time. Gaver (1968) introduced diffusion approximation to analyze the mean waiting time of an M/G/1 systems. Fischer (1977) modified Gaver's results to approximate the mean waiting time in the system for queueing system with service interruptions. He analyzed the system when both on-time and off-time have a general time distribution. Federgruen and Green (1986) found the exact expression for the expected waiting time, given that the time between interruptions is exponentially distributed. Nicola (1986) obtained the steady-state average number of customers in the
system for an M/G/1 queue with mixed types of interruptions. The inter-arrival times of these interruptions are exponentially distributed. Federgruen et al. (1988) extended their results to models where the on-time period has a phase-type distribution. They derived the lower and upper bounds of the mean waiting time in the system. Sengupta (1990) examined the steady-state queueing system where the customer service time distributions in the on-time and the off-time periods are different. He determined the exact closed-form expression for the expected waiting time, given that the on-time distribution is exponential.

1.3.2 Preventive maintenance models

The reliable performance of a system is of great importance in many industrial, military, and everyday life situations. There are several methods for improving systems reliability. These methods include using redundant servers, reducing the complexity of the system, increasing the reliability of constituent components, and using preventive maintenance.

There are extensive research in the field of preventive maintenance scheduling. Morse (1958) studied the preventive maintenance of a single operating system where the equipment is repaired at failure or at age $t_0$, whichever occurs first. Barlow and Hunter (1960) determined the optimal policy that maximizes the availability of the machine for an infinite planning horizon. They also pointed out that the problem is equivalent to the age-replacement problem, if the mean repair time is replaced by the
replacement cost of a failed machine, and the performance of preventive maintenance replaced by the exchange cost of non-failed unit. Barlow and Proschan (1965) derived an expression for the optimum age-replacement that minimizes the expected cost per unit time over an infinite planning horizon. Pierskalla and Voelker (1976) presented a survey of maintenance models of deteriorating systems. They reviewed most of the results found up to the year 1976. Nakagawa (1977) maximized the expected earnings rate that the system produces under certain conditions on the cost parameters and the failure rate function using preventive maintenance. Sherif (1981) reviewed the literature on maintenance models of system subjects to failure. His work emphasized on Pierskalla's paper. Nguyen and Murthy (1981) studied the problem of optimal preventive policies for repairable system. After each repair the system improves but the failure rate function increases with the number of repair carried out. They determined the optimal age-replacement policy assuming the maintenance, repair and replacement times are negligible. Puri and Singh (1986) proved a general mathematical lemma that helps in finding the optimal age-replacement time for systems subject to shocks. The expected cost $C(t)$ is expressed in the form $[R + \psi(t)] / \phi(t)$. If the conditions imposed on $R$, $\psi(t)$ and $\phi(t)$ are met then their lemma can be applied to find the optimal replacement time. However, if the conditions are not met — for example $\psi'(t)/\phi'(t)$ is not monotone — the lemma cannot be applied directly to obtain the optimal policy. Aven and Bergman (1986)
applied the $\lambda$-minimization technique to models having the following common form:

$$C(T) = \frac{E[\int_0^T a(t)h(t)dt + c(0)]}{E[\int_0^T h(t)dt + p(0)]}.$$  

Where $a(t)$ is non-decreasing stochastic process, $h(t)$ is non-negative stochastic process, $c(0)$ and $p(0)$ are non-negative random variables. $C(T)$ is expressed as:

$$E[\int_0^T [(a(t)-\lambda)h(t)]dt + c(0) - \lambda p(0)].$$

Under certain conditions on $a(t)$, $h(t)$, $c(0)$ and $p(0)$, they concluded that there exist $\lambda^*$ and $T^*_\lambda$ minimizing $C(T)$. Nakagawa (1986) considered periodic and sequential preventive maintenance policies for systems subject to failure. The system has a different failure time distribution between maintenance and it is replaced at the $n^{th}$ maintenance. He assumed that the failure rate function increases with maintenance and the times for maintenance, repair and replacement are negligible. Tapeiro (1986) formulated a problem of continuous quality production and maintenance of a machine. The rate of degradation is reduced by implementing a maintenance program $u$. The machine degradation state is a function of machine usage. Control theory methods are used to determine the optimal maintenance level $u^*$. However, the major assumption made in his model is that at the time of failure, the broken machine must either be junked or repaired. In other words, the level of repair (repair/replace) is
fixed and not a decision variable.

Cho and Parlar (1988) presented a survey of maintenance models for single-unit as well as for multi-unit systems. They provided a quick guide for classifying the literature in maintainability. They also introduced possible extensions and suggested problem for research. Most recently, Valdez-Flores and Feldman (1989) presented the most complete survey to-date of preventive maintenance models for stochastically deteriorating single-unit systems. They summarized results of different maintenance policies for one machine problems. They have focused on the work done since 1976. Ritchken and Wilson (1990) introduced a group maintenance policy which is conducted at time $T$ or upon $m$ failures whichever comes first. They developed a general algorithm for computing optimal policies for general failure time distribution and cost structures. Wilson el al. (1990) determined the optimal $m$-failure policies for parallel machines when the repair time is a positive random variable.

Possibly the only work to-date that addresses the optimization of queueing systems with service interruptions are Federgruen and So (1989) and (1990). Federgruen and So (1989) determined the optimal time to repair a broken server. When a breakdown occurs, the server is repaired immediately, or the repair is postponed until the number of customers waiting in the system is equal to a certain threshold (Wait/Repair). Federgruen and So (1990) used two types of repairs in the system when the server must be
repaired immediately. The decision to be taken is what type of repair (Repair/Repair) should be considered. However, the implementation of preventive maintenance to reduce the rate of failure in queueing systems is missing.

1.3.3 Unreliable queueing networks

Takahashi et al. (1980) presented an approximation method for analyzing open restricted queueing networks. Their analysis consists of node-by-node decomposition of the network and the introduction of pseudo-arrival rate and an effective service rate. Altılık et al. (1983) studied a flow-shop-type production line where the stations are subject to breakdown. Vinod and Altılık (1986) presented an exact equivalent network with two-stage coxian service time distribution for unreliable closed queueing network. They approximated the equivalent network by assuming exponentiality of the service completion time. Ramanjaneyulu et al. (1989) presented a method for modeling server unreliability in closed queueing networks. They converted the problem into a closed queue with all reliable servers. The above work on unreliable queueing networks focus only on estimating parameters such as mean waiting times and service completion times. None of these models addresses the optimization of a queueing network through preventive maintenance.
1.3.4 Conclusion and further research

Most models on maintenance related to our problem assumed that the times for maintenance, repair, and replacement are negligible. This assumption is not appropriate for many queueing systems. In particular, if preventive maintenance (PM) is too frequent, the queueing system may become unstable (i.e., the formulation part of the expected waiting time is infinite or negative). By assuming that the off-period is zero, the mean waiting time in the system is always positive and constant, regardless of PM frequencies. Obviously, this is not true for real queueing systems.

Many possible avenues for expansion of the area are still open. Expansion of the area is listed below:

a) The implementation of preventive maintenance policies for queueing systems with service breakdowns does not exist.

b) In the literature, the maintenance, repair and replacement times are assumed to be negligible. This assumption may not apply in many queueing systems.

c) When the times for maintenance and repair are not negligible, the output process may not be a renewal process.

d) When the output process of a queueing system is not acceptable, maintenance action can be used to improve the quality as well.

e) Optimization of unreliable open queueing networks with general failure time distributions do not exist in the literature. Discrete-event simulation can be used as an obvious alternative
to evaluate the system performance. However, the high cost involved in simulation studies, have brought about increasing interest in the use of approximation methods.

1.4 THESIS ORGANIZATION

This dissertation consists of five chapters. The second chapter examines a single-server queueing system. The third chapter investigates a single-server queueing system with buffer cost and quality cost. The fourth chapter examines an unreliable queueing network. Finally, conclusions and further research are discussed in Chapter 5. Basically, Chapter 2 provides tools for Chapters 3 and 4. We develop efficient methods to determine the optimal age-maintenance $t_0^*$ for M/G/1 queues and the optimal policy for M/M/1 queues with buffer and quality costs. For unreliable queueing networks with maintenance crew cost, the optimal policy is determined by solving a large integer nonlinear programming problem.

In Chapter two, we will explore the use of diffusion approximation methods to determine the expected waiting time of jobs in the system under the age-maintenance policy. We will also show that when the traffic intensity is high, the maintenance decision that maximizes the fraction of time the server is on, also can minimize the expected waiting time in the system. We will provide some conditions on the cost parameters, maintenance and repair times, and failure rate function so that the optimal age-maintenance can be found efficiently.
In Chapter three, we will study an M/M/1 queue with buffer and quality costs. We will show that when the cost of quality is zero, the total expected cost is unimodal with respect to the buffer and when the cost of buffer is zero, the expected total cost is unimodal with respect to the allowed number of defective items. The maintenance decision which decreases the waiting capacity, increases the number of defective item per cycle and vice versa. The trade-off between the buffer cost and the quality cost must be considered. We will develop an efficient procedure for determining the optimal policy.

In Chapter four, we will examine the maintenance problem of an open queueing network with maintenance crew size cost. At each feasible crew size, an algorithm will be developed to find the optimal maintenance decision. The algorithm will be tested against a nonlinear programming software (GAMS). The optimal Lagrange multiplier of crew size $N$ can be used as the upper bound for the optimal Lagrange multiplier of crew size $N+1$. We will show that the total expected cost is unimodal with respect to the maintenance crew size. Based on unimodality property, a computationally expedient procedure will be determined to find the optimal crew size. In general, the optimal maintenance crew size is unique. Basically, the proposed algorithm can provide the optimal solution for any size of the network.
CHAPTER TWO

M/G/1 Queues with Age-Maintenance
**Notation**

- $\beta$: the arrival rate of a breakdown after the last repair or PM
- $C_b$: the cost of each breakdown per unit time
- $C_m$: the cost of each maintenance per unit time
- $C_w$: the cost of waiting per unit time
- $f(t)$: the probability density function that a breakdown occurs at time $t$ since the last repair or maintenance
- $F(t_0)$: the probability that the server fails before $t_0$
- $\bar{F}(t_0)$: the probability that the server is switched off before breakdown. \[ \bar{F}(t_0) = 1 - F(t_0) \]
- $r(t)$: the failure rate $= f(t)/\bar{F}(t)$
- $E[.]$: the expectation
- $V[.]$: the variance
- $\tau_b$: the mean time of the working server without PM $(1/\beta)$
- $\tau_m$: the average time to perform a maintenance
- $\tau_r$: the average repair time
- $\lambda$: the job's arrival rate
- $E[S]$: the mean service time
- $E[S^2]$: the second moment of the service time
- $W(t_0)$: the expected waiting time in the system under $t_0$
- $\varphi(t_0)$: the average cycle time
- $A(t_0)$: the mean fraction of time the server is working under $t_0$
2.1 BACKGROUND

This chapter examines the problem of determining the optimal maintenance policy for operating devices that are subject to breakdown. The decision variable is calculated only once and it is used over the entire planning horizon. We assume that the maintenance activity is permitted to occur at continuous time (i.e., the interval between maintenance is not necessarily integer). The goal is to determine the optimal maintenance policy which minimizes the expected cost per unit time, over an infinite planning horizon. The cost function includes the cost of jobs waiting in the system, the cost of maintenance and the cost of repairs.

In order to find some analytical results, assumptions regarding the basic elements of the queueing system are made. These assumptions are:

1. Customers/jobs arrive to the system according to a Poisson process.

2. The traffic intensity in the system is high.

3. The waiting capacity is infinite and at no cost, this assumption is relaxed in Chapter 3.

4. After undergoing maintenance or repair, the queueing system is like new. This means that after each maintenance or repair we have a new cycle, and the cycles are independent. For queueing
systems without the "like-new" condition, a cycle is completed only when the server is completely replaced.

5. The failure rate \( r(t) \) is a continuous and strictly increasing function of time. Clearly, preventive maintenance for constant failure rates (e.g., exponential failure time distribution) is not worthwhile for single-servers.

6. When a customer/job is preempted by an interruption, the server will resume (not repeat) its service after the interruption.

2.2 PROBLEM DEFINITION

Suppose customers (or jobs) arrive at a service facility according to a Poisson process with arrival rate \( \lambda \). The service time of arriving customers has a general distribution with mean \( E[S] \). While the server is up, the server is subject to breakdowns with general failure time distribution \( F(.) \). If the server is working without failure for age \( t_0 \), we stop the operation of the working unit for preventive maintenance. The maintenance is performed with a random time with mean \( \tau_m \). The cycle is then repeated. If the server fails before \( t_0 \), it will be repaired with an average repair time \( \tau_r \). The server is repaired upon failure or maintained at age \( t_0 \) whichever occurs first. The value of \( \tau_r \) is often longer than \( \tau_m \) (because a breakdown is more serious than PM, or a breakdown comes at an unplanned time and this involves additional delays). We can divide the sequence of
events into two kinds of cycles, both starting when the server is just put into running condition (see Figure 2.1).

![Diagram showing PM maintenance cycle with \( \bar{F}(t_0) \) and breakdown repair cycle with \( F(t_0) \).]

**Figure 2.1: Queueing system with age-maintenance policy**

Let \( f(t) \) be the density function of the probability that a breakdown occurs at time \( t \) since the last maintenance or repair. Let \( F(t_0) \) be the cumulative probability that the server fails before \( t_0 \). When the server is like-new after performing repair or maintenance, we have

\[
F(t_0) = \int_0^{t_0} f(t)dt
\]  

(1)

The problem is to determine the value of \( t_0 \) that minimizes the total expected cost per unit time, for an infinite planning horizon. The cost function includes the cost of jobs waiting in the system, the cost of maintenance, and the cost of repairs.
2.3 THE BREAKDOWN/Maintenance PROCESS

Figure 2.2 shows a queueing system alternating between the on-time period with random time $X_{on}$, and the off-time period with random time $X_{off}$. The server is operating in the on-time period, and it is either under maintenance or under repair in the off-time period.

![Diagram](image)

Figure 2.2: Queueing system with alternate ON and OFF periods

2.3.1 The on-period

To reduce the rate of breakdowns, we consider the age-maintenance policy which is defined in Section 2.2. This policy is known in the literature as the age-replacement policy see for instance Ritchken and Wilson (1990) and Nakagawa (1977). Suppose we combine preventive maintenance (PM) and breakdown into one process. This combined process is expressed as $(1, t_o)$ policy and it has the following truncated cumulative distribution:

$$ g(x) = \begin{cases} 
F(x) & \text{if } x < t_o \\
1 & \text{if } x \geq t_o 
\end{cases} $$

(2)
The density function is:

\[ q(x) = \begin{cases} 
  f(x) & \text{if } x < t_0 \\
  1 - F(t_0) & \text{otherwise} 
\end{cases} \quad (3) \]

**Lemma 2.1:**

Given \( F(t) \) and \( t_0 \), the mean and the variance of the PM/breakdown process are:

\[
E[X_{on}] = \int_0^{t_0} F(t) dt, \quad \text{and} \quad (4)
\]

\[
V[X_{on}] = \int_0^{t_0} 2tF(t) dt - \left( \int_0^{t_0} F(t) dt \right)^2 \quad (5)
\]

**Proof:**

The \( n \)th moment of a truncated distribution (shown in Figure 2.3) is:

\[
E[X^n] = \int_0^{t_0} t^n f(t) dt + t_0^n \bar{F}(t_0)
\]

Therefore the mean of the process can be computed as follows:

\[
E[X_{on}] = \int_0^{t_0} tf(t) dt + t_0 \bar{F}(t_0)
\]

\[
= t_0 F(t_0) - \int_0^{t_0} F(t) dt + t_0 \bar{F}(t_0)
\]

\[
= \int_0^{t_0} \bar{F}(t) dt
\]

also,

\[
E[X_{on}^2] = \int_0^{t_0} t^2 f(t) dt + t_0^2 \bar{F}(t_0)
\]
\[ E[X_{on}^2] = \int_0^t 2tF(t)dt \]

The variance is: \[ V[X_{on}] = E[X_{on}^2] - (E[X_{on}])^2. \]

Therefore \[ V[X_{on}] = \int_0^t 2tF(t)dt - \left( \int_0^t F(t)dt \right)^2 \]

Figure 2.3: Truncated Arrival Process Pdf
2.3.2 The off-period

Let \( X_m \) and \( X_r \) be random variables representing the maintenance and repair times, respectively. These random variables are assumed to be statistically independent.

**Lemma 2.2:**

The mean and the variance of the off-time period are:

\[
E[X_{off}] = \tau_m + (\tau_r - \tau_m)F(t_0) \tag{6}
\]

\[
V(X_{off}) = F(t_0)E[X_r] + \bar{F}(t_0)E[X_m] + F(t_0)\bar{F}(t_0)(\tau_r - \tau_m)^2 \tag{7}
\]

**Proof:**

Let \( \tau_m \) and \( \tau_r \) be the mean time of maintenance and repair, respectively. Since

\[
X_{off} = \begin{cases} 
X_m & \text{with probability } \bar{F}(t_0) \\
X_r & \text{with probability } F(t_0)
\end{cases}
\]

\[
\tau_m = E[X_m] \quad \text{and} \quad \tau_r = E[X_r]
\]

Then

\[
E[X_{off}] = E[X_m]\bar{F}(t_0) + E[X_r]F(t_0) = \tau_m + (\tau_r - \tau_m)F(t_0)
\]

We have

\[
E[X_{off}^2] = \begin{cases} 
E[X_m^2] & \text{with } \bar{F}(t_0) \\
E[X_r^2] & \text{with } F(t_0)
\end{cases}
\]
Therefore,
\[ E[X_{off}^2] = E[X_m^2] \bar{F}(t_0) + E[X_r^2] F(t_0), \text{ and} \]
\[ V[X_{off}] = E[X_{off}^2] - E[X_{off}]^2. \]

Consequently,
\[ V[X_{off}] = F(t_0)V[X_r] + \bar{F}(t_0)V[X_m] + F(t_0)\bar{F}(t_0)(\tau_r - \tau_m)^2. \]

2.3.3 The on/off cycle

Let \( q(x) \) and \( h(x) \) be the density functions for the length of time the server is working, i.e., \( \min(x, t_0) \) and the length of time the server is under maintenance or repair, respectively. The arrival and the service patterns behave as an M/G/1 queue during the on-time period, and as a system with arrivals and no departure in the off-time period (See Figure 2.4).

![Figure 2.4: Representation of the model](image-url)
Lemma 2.3:

Let $A(t_0)$ and $\varphi(t_0)$ be the fraction of time the server is on and the mean length of a cycle, respectively. Then:

$$\varphi(t_0) = \int_0^{t_0} \overline{F}(t) dt + \tau + (\tau - \tau) F(t_0)$$  \hspace{1cm} (8)$$

$$A(t_0) = \frac{\int_0^{t_0} \overline{F}(t) dt}{\varphi(t_0)}$$  \hspace{1cm} (9)$$

Proof:

Suppose $X_{on}$ and $X_{off}$ are independent random variables with density functions $q(.)$ and $h(.)$, respectively. Let $X_c$ be the random variable representing the length of a cycle, i.e., $X_c = X_{on} + X_{off}$.

Then $X_c$ has the following density function:

$$u(t) = \int_0^t q(x) h(t-x) dx.$$  

Since $X_{on}$ and $X_{off}$ are independent, then

$$E[X_c] = E[X_{on} + X_{off}] = E[X_{on}] + E[X_{off}]$$

$$= \int_0^{t_0} \overline{F}(t) dt + \tau + (\tau - \tau) F(t_0).$$

The fraction of time the server is on is:

$$A(t_0) = \frac{E[X_{on}]}{E[X_{on}] + E[X_{off}]}$$
\[ A(t_0) = \int_0^{t_0} \bar{F}(t) dt \]

Then \[ A(t_0) = \frac{\int_0^{t_0} \bar{F}(t) dt}{\varphi(t_0)} . \]

2.4 THE EXPECTED WAITING TIME OF JOBS IN THE SYSTEM

It is very difficult to obtain analytical results for queueing systems with non-exponential on-times. Let \( E[W(t_0)] \) be the expected waiting time of jobs in the system under the policy \( t_o \). To determine \( E[W(t_0)] \), we assume that the traffic intensity is high. This assumption allows us to apply diffusion approximation.

2.4.1 Diffusion approximation

Our motivation for using diffusion process approximation (see, Gaver (1968), Newell (1971), Kobayashi (1974), Glenbe (1975), Fisher (1977), Pujolle and Ai (1986)) is to develop a realistic analytical model when the on- and off-times have general distributions. The diffusion approximation technique assumes that the number of events in a given time interval is approximately normally distributed and that the queue is almost always non-empty. First we define some parameters:

\[ a(t) = \text{the number of arrivals in (0,t)} \]
\[ d(t) = \text{the number of departure in (0,t)} \]
\[ N(t) = \text{the number of customer present in the system at t} \]

i.e., \[ N(t) = a(t) - d(t) \]

and \[ N(0) = 0 \]
Let $\sigma^2_a(t)$ and $\sigma^2_d(t)$ be the variances of the arrival $a(t)$ and the departure $d(t)$ processes, respectively. These variances represent random fluctuations of the processes about their means. The fluctuations about the mean value of the processes are depicted by the Normal distributions [Kleinrock (1975)]. The processes $a(t)$ and $d(t)$ are approximated by continuous random processes with independent increments. These processes are normally distributed at time $t$ with means $\mathbb{E}[a(t)]$ and $\mathbb{E}[d(t)]$ and variances $\sigma^2_a(t)$ and $\sigma^2_d(t)$, respectively.

The linear combination of two independent normally distributed processes is a normally distributed process. Thus, $N(t) = a(t) - d(t)$ is a normally distributed process with mean $\mathbb{E}[N(t)]$ and variance $\sigma^2_N$ ($\mathbb{E}[N(t)] = \mathbb{E}[a(t)] - \mathbb{E}[d(t)]$ and $\sigma^2_N(t) = \sigma^2_a(t) + \sigma^2_d(t)$). When $N(t)$ is large, the departure process is approximately independent of the arrival process. This is "almost" true, when the traffic intensity is close to unity. Since $N(t)$ cannot be negative, then a reflecting boundary is created at its origin. We are interested in changes of $N(t)$ at a small time interval. The time interval must be small enough so that $N(t)$ changes by a small fraction of its value. On the other hand, $N(t)$ must be large enough to permit enough discrete jumps to take place. With this condition, $N(t)$ can be approximated by a continuous random process $X(t)$. $X(t)$ is assumed to be a continuous Markov process.

The conditional continuous-state transition probability is given by:

$$P(y, x, T, t) = P[X(T) = y \mid X(t) = x] \quad T > t \quad (10)$$
Equation (10) is the probability that \( X(t) \) takes on a value less than or equal to \( y \) by time \( T \), given that it took on the value \( x \) at time \( t \). We assume \( \partial P / \partial x \) and \( \partial^2 P / \partial x^2 \) exist and are continuous. The conditional mean \( M(x, T) \) and the conditional variance \( V(x, T) \) are given as follows:

\[
M(x, T) = E[ X(T) \mid X(t) = x ] \tag{11}
\]

\[
V(x, T) = E[ (X(T) - M(x, T))^2 \mid X(t) = x ] \tag{12}
\]

The rate of change of (11) and (12) with respect to \( T \), at point \( T = t \) are:

\[
m(x, t) = \frac{\partial M(x, T)}{\partial T} \bigg|_{T=t} \tag{13}
\]

\[
\omega(x, t) = \frac{\partial V(x, T)}{\partial T} \bigg|_{T=t} \tag{14}
\]

Expressions (13) and (14) are called the infinitesimal mean and the variance, respectively.

2.4.2 The expected waiting time of an M/G/1 queueing system

Suppose the waiting capacity is infinite. Let \( W \) be the waiting time (in the queue) of the customer arriving at time \( t \). Assume customers are served on a first-come-first-served basis. Then \( W \) is a random process, considered as a function of time and modified by a reflecting barrier at \( W = 0 \). Assume the traffic intensity \( \lambda E[S] \) is close to unity. Thus the process \( W \) is
replaced with an approximating continuous process, diffusion process. In order to approximate the distribution of $W$, Gaver (1968) derived an expression for the infinitesimal mean and variance, $\mu$ and $\psi$ respectively. The values of $\mu$ and $\psi$ are:

$$\mu = \lambda E[S] - 1$$

$$\psi = \lambda E[S^2]$$

(16)

(17)

where $E[S^2]$ is the second moment of the service time.

Gaver showed that the Laplace-Steiltjes transform of the mean waiting time for a simple $M/G/1$ system is:

$$F^*(\alpha) = \int_0^\infty e^{-\alpha x} \, dP[W \leq x] = \frac{2\mu}{2\mu - \alpha \psi}$$

(18)

Thus the cumulative distribution of the waiting time can be stated as follows:

$$P[W \leq x] = 1 - e^{2(\mu x)/\psi}$$

(19)

The expected steady state waiting time in the system is:

$$E[W] = \int_0^\infty x \, dP[W \leq x] = - \frac{\psi}{2\mu}$$
Therefore the mean waiting time of an $M/G/1$ queue is:

$$E[W] = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} \quad (\text{See Gaver 1968})$$

(20)

2.4.3 The expected jobs waiting time in the system under maintenance and repair

The expected waiting time in a system with breakdowns and maintenance, under preemptive-resume, is stated in the following theorem.

Theorem 2.1:

The expected waiting time in the system is:

$$E[W(t_0)] = \frac{\lambda E[S^2] + \frac{V[X_{\text{on}}](1-A(t_0))^2 + V[X_{\text{off}}]A(t_0)^2}{\varphi(t_0)}}{2(A(t_0) - \lambda E[S])}$$

(21)

Proof:

Let $Y(t)$ be the fraction of time the server is on during $(0,t)$, for large $t$. Let $\lim_{t \to \infty} E[Y(t)] = E[Y]$ and $\lim_{t \to \infty} V[Y(t)] = V[Y]$. From Cox (1962), we have:

$$E[Y] = \frac{E[X_{\text{on}}]}{E[X_{\text{on}}} + E[X_{\text{off}}]$$

(22)
\[ V[Y] = \frac{V[X_{on}]E[X_{off}]^2 + V[X_{off}]E[X_{on}]^2}{(E[X_{on}] + E[X_{off}])^3} \]  
(23)

Fischer (1977) modified \( m \) and \( \varphi \) in (16) and (17) for queueing systems with general on-time and off-time periods and under the heavy traffic condition as follows:

\[ m = \lambda E[S] - E[Y] \]  
(24)

and \[ \varphi = \lambda E[S^2] + V[Y] \]  
(25)

The expected waiting time in the system \( E[W] \) can be expressed as follows:

\[ E[W] = \frac{\lambda E[S^2] + V[Y]}{2(E[Y] - \lambda E[S])} \]

Using Lemma (2.1) - (2.3), we compute the following expressions under the age-maintenance \( t_0 \):

\[ E[Y] = A(t_0) = \frac{\int_0^{t_0} \bar{F}(t)dt}{\varphi(t_0)} \]

\[ V[Y] = V[Y(t_0)] = \frac{V[X_{on}](1-A(t_0))^2 + V[X_{off}]A(t_0)^2}{\varphi(t_0)} \]

\( A(t_0) \) and \( \varphi(t_0) \) are defined in Lemma (2.3). The average waiting time of jobs in the system under age-maintenance \( t_0 \) is approximated
as:

\[ E[W(t_0)] = \frac{\lambda E[S^2] + V[Y(t_0)]}{2(A(t_0) - \lambda E[S])} \]

Consequently,

\[ E[W(t_0)] = \frac{\lambda E[S^2] + \frac{V[X_{\text{off}}]A(t_0)^2 + V[X_{\text{on}}](1-A(t_0))^2}{\phi(t_0)}}{2(A(t_0) - \lambda E[S])} \]

Note that if the off-period is zero, \( E[W(t_0)] = \frac{\lambda E[S^2]}{2(1-\lambda E[S])} \).

This is simply the waiting time of an M/G/1 queue.

Q.E.D.

### 2.5 SOME COMPUTATIONAL RESULTS

This section illustrates the computation of the expected waiting time. We consider two cases. The first case assumes that there is no maintenance \( (t_0 = \infty) \). The second case uses maintenance \( (t_0 < \infty) \).

#### 2.5.1 No preventive maintenance \( (t_0 = \infty) \)

In order to compare expression (21) with known results, we consider the following:

1. No preventive maintenance \( (t_0 = \infty) \)
2. The on- and off-times have exponential distributions
Federgruen et al. (1985) found the exact expression for the expected waiting time when the on- and off-times have exponential distributions. Table 2.1 to Table 2.3 compare the values \( W_a \) computed from expression (21), with Federgruen's results \( W_e \). 

\[ \lambda E[S] \] is the traffic intensity and \( 1 - A(\omega) \) is the fraction of time the server is off. Note that when \( \lambda E[S] \geq 0.9 \), \( 1 - A(\omega) \) must be less or equal to 0.1, the level of the error is no more than 9% which implies that the diffusion approximation is accurate. Note that the approximation is always underestimated. The error is illustrated in Figure 2.5. However, when \( \lambda E[S] \leq 0.8 \) and \( 1 - A(\omega) \geq 0.2 \), the error is significant. Consequently, the diffusion approximation method cannot be applied for this case. Thus we will concentrate on queueing systems with average on-times must be greater than average off-times.
<table>
<thead>
<tr>
<th>λS [s]</th>
<th>W_a</th>
<th>W_c</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.800</td>
<td>0.42500</td>
<td>0.50070</td>
<td>15.10</td>
</tr>
<tr>
<td>0.805</td>
<td>0.44138</td>
<td>0.51797</td>
<td>14.78</td>
</tr>
<tr>
<td>0.810</td>
<td>0.45893</td>
<td>0.53648</td>
<td>14.45</td>
</tr>
<tr>
<td>0.815</td>
<td>0.47778</td>
<td>0.55635</td>
<td>14.12</td>
</tr>
<tr>
<td>0.820</td>
<td>0.49808</td>
<td>0.57775</td>
<td>13.78</td>
</tr>
<tr>
<td>0.825</td>
<td>0.52000</td>
<td>0.60087</td>
<td>13.45</td>
</tr>
<tr>
<td>0.830</td>
<td>0.54375</td>
<td>0.62591</td>
<td>13.12</td>
</tr>
<tr>
<td>0.835</td>
<td>0.56957</td>
<td>0.65313</td>
<td>12.79</td>
</tr>
<tr>
<td>0.840</td>
<td>0.59773</td>
<td>0.68282</td>
<td>12.46</td>
</tr>
<tr>
<td>0.845</td>
<td>0.62857</td>
<td>0.71534</td>
<td>12.12</td>
</tr>
<tr>
<td>0.850</td>
<td>0.66250</td>
<td>0.75112</td>
<td>11.79</td>
</tr>
<tr>
<td>0.855</td>
<td>0.70000</td>
<td>0.79066</td>
<td>11.46</td>
</tr>
<tr>
<td>0.860</td>
<td>0.74167</td>
<td>0.83459</td>
<td>11.13</td>
</tr>
<tr>
<td>0.865</td>
<td>0.78824</td>
<td>0.88369</td>
<td>10.80</td>
</tr>
<tr>
<td>0.870</td>
<td>0.84063</td>
<td>0.93893</td>
<td>10.46</td>
</tr>
<tr>
<td>0.875</td>
<td>0.90000</td>
<td>1.00154</td>
<td>10.13</td>
</tr>
<tr>
<td>0.880</td>
<td>0.96786</td>
<td>1.07308</td>
<td>9.80</td>
</tr>
<tr>
<td>0.885</td>
<td>1.04615</td>
<td>1.15564</td>
<td>9.47</td>
</tr>
<tr>
<td>0.890</td>
<td>1.13750</td>
<td>1.25195</td>
<td>9.14</td>
</tr>
<tr>
<td>0.895</td>
<td>1.24545</td>
<td>1.36578</td>
<td>8.81</td>
</tr>
<tr>
<td>0.900</td>
<td>1.37500</td>
<td>1.50237</td>
<td>8.47</td>
</tr>
<tr>
<td>0.905</td>
<td>1.53333</td>
<td>1.66931</td>
<td>8.14</td>
</tr>
<tr>
<td>0.910</td>
<td>1.73125</td>
<td>1.87799</td>
<td>7.81</td>
</tr>
<tr>
<td>0.915</td>
<td>1.98572</td>
<td>2.14630</td>
<td>7.48</td>
</tr>
<tr>
<td>0.920</td>
<td>2.32500</td>
<td>2.50404</td>
<td>7.15</td>
</tr>
<tr>
<td>0.925</td>
<td>2.80000</td>
<td>3.00487</td>
<td>6.81</td>
</tr>
<tr>
<td>0.930</td>
<td>3.51250</td>
<td>3.75612</td>
<td>6.48</td>
</tr>
<tr>
<td>0.935</td>
<td>4.70000</td>
<td>5.00821</td>
<td>6.15</td>
</tr>
<tr>
<td>0.940</td>
<td>7.07501</td>
<td>7.51238</td>
<td>5.82</td>
</tr>
<tr>
<td>0.945</td>
<td>14.20004</td>
<td>15.02491</td>
<td>5.49</td>
</tr>
</tbody>
</table>

Table 2.1: Comparison of the waiting times when $1 - A(\infty) = 0.05$
<table>
<thead>
<tr>
<th>$\lambda E[S]$</th>
<th>$W_a$</th>
<th>$W_e$</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.750</td>
<td>0.7815</td>
<td>0.9256</td>
<td>15.57</td>
</tr>
<tr>
<td>0.755</td>
<td>0.8100</td>
<td>0.9574</td>
<td>15.39</td>
</tr>
<tr>
<td>0.760</td>
<td>0.8405</td>
<td>0.9914</td>
<td>15.36</td>
</tr>
<tr>
<td>0.765</td>
<td>0.8733</td>
<td>1.0279</td>
<td>15.04</td>
</tr>
<tr>
<td>0.770</td>
<td>0.9086</td>
<td>1.0672</td>
<td>14.86</td>
</tr>
<tr>
<td>0.775</td>
<td>0.9466</td>
<td>1.1096</td>
<td>14.68</td>
</tr>
<tr>
<td>0.780</td>
<td>0.9878</td>
<td>1.1555</td>
<td>14.51</td>
</tr>
<tr>
<td>0.785</td>
<td>1.0326</td>
<td>1.2054</td>
<td>14.33</td>
</tr>
<tr>
<td>0.790</td>
<td>1.0814</td>
<td>1.2598</td>
<td>14.09</td>
</tr>
<tr>
<td>0.795</td>
<td>1.1348</td>
<td>1.3193</td>
<td>13.38</td>
</tr>
<tr>
<td>0.800</td>
<td>1.1936</td>
<td>1.3847</td>
<td>13.27</td>
</tr>
<tr>
<td>0.805</td>
<td>1.2584</td>
<td>1.4569</td>
<td>13.44</td>
</tr>
<tr>
<td>0.810</td>
<td>1.3304</td>
<td>1.5371</td>
<td>13.27</td>
</tr>
<tr>
<td>0.815</td>
<td>1.4107</td>
<td>1.6265</td>
<td>13.09</td>
</tr>
<tr>
<td>0.820</td>
<td>1.5009</td>
<td>1.7271</td>
<td>12.91</td>
</tr>
<tr>
<td>0.825</td>
<td>1.6031</td>
<td>1.8409</td>
<td>12.73</td>
</tr>
<tr>
<td>0.830</td>
<td>1.7197</td>
<td>1.9707</td>
<td>12.56</td>
</tr>
<tr>
<td>0.835</td>
<td>1.8540</td>
<td>2.1204</td>
<td>12.38</td>
</tr>
<tr>
<td>0.840</td>
<td>2.0103</td>
<td>2.2944</td>
<td>12.20</td>
</tr>
<tr>
<td>0.845</td>
<td>2.1962</td>
<td>2.4997</td>
<td>12.02</td>
</tr>
<tr>
<td>0.850</td>
<td>2.4151</td>
<td>2.7453</td>
<td>11.85</td>
</tr>
<tr>
<td>0.855</td>
<td>2.6836</td>
<td>3.0444</td>
<td>11.67</td>
</tr>
<tr>
<td>0.860</td>
<td>3.0178</td>
<td>3.4167</td>
<td>11.49</td>
</tr>
<tr>
<td>0.865</td>
<td>3.4450</td>
<td>3.8926</td>
<td>11.32</td>
</tr>
<tr>
<td>0.870</td>
<td>4.0106</td>
<td>4.5226</td>
<td>11.14</td>
</tr>
<tr>
<td>0.875</td>
<td>4.7947</td>
<td>5.3958</td>
<td>10.96</td>
</tr>
<tr>
<td>0.880</td>
<td>5.9533</td>
<td>6.6867</td>
<td>10.79</td>
</tr>
<tr>
<td>0.885</td>
<td>7.8411</td>
<td>8.7895</td>
<td>10.06</td>
</tr>
<tr>
<td>0.890</td>
<td>11.4606</td>
<td>12.8213</td>
<td>9.20</td>
</tr>
<tr>
<td>0.895</td>
<td>21.6139</td>
<td>23.6856</td>
<td>9.01</td>
</tr>
</tbody>
</table>

Table 2.2: Comparison of the waiting times when $1-A(\infty) = 0.1$
<table>
<thead>
<tr>
<th>$\lambda_E[S]$</th>
<th>$W_a$</th>
<th>$W_e$</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.700</td>
<td>1.63000</td>
<td>2.11750</td>
<td>23.02</td>
</tr>
<tr>
<td>0.705</td>
<td>1.71842</td>
<td>2.22908</td>
<td>22.90</td>
</tr>
<tr>
<td>0.710</td>
<td>1.81667</td>
<td>2.35306</td>
<td>22.79</td>
</tr>
<tr>
<td>0.715</td>
<td>1.92647</td>
<td>2.49162</td>
<td>22.68</td>
</tr>
<tr>
<td>0.720</td>
<td>2.05000</td>
<td>2.64750</td>
<td>22.56</td>
</tr>
<tr>
<td>0.725</td>
<td>2.19000</td>
<td>2.82417</td>
<td>22.45</td>
</tr>
<tr>
<td>0.730</td>
<td>2.35000</td>
<td>3.02607</td>
<td>22.34</td>
</tr>
<tr>
<td>0.735</td>
<td>2.53462</td>
<td>3.25904</td>
<td>22.22</td>
</tr>
<tr>
<td>0.740</td>
<td>2.75000</td>
<td>3.53083</td>
<td>22.11</td>
</tr>
<tr>
<td>0.745</td>
<td>3.00455</td>
<td>3.85205</td>
<td>22.00</td>
</tr>
<tr>
<td>0.750</td>
<td>3.31000</td>
<td>4.23750</td>
<td>21.88</td>
</tr>
<tr>
<td>0.755</td>
<td>3.68333</td>
<td>4.70861</td>
<td>21.77</td>
</tr>
<tr>
<td>0.760</td>
<td>4.15000</td>
<td>5.29750</td>
<td>21.66</td>
</tr>
<tr>
<td>0.765</td>
<td>4.75000</td>
<td>6.05464</td>
<td>21.54</td>
</tr>
<tr>
<td>0.770</td>
<td>5.55000</td>
<td>7.06417</td>
<td>21.43</td>
</tr>
<tr>
<td>0.775</td>
<td>6.67000</td>
<td>8.47750</td>
<td>21.32</td>
</tr>
<tr>
<td>0.780</td>
<td>8.35000</td>
<td>10.59751</td>
<td>21.20</td>
</tr>
<tr>
<td>0.785</td>
<td>11.15001</td>
<td>14.13084</td>
<td>21.09</td>
</tr>
<tr>
<td>0.790</td>
<td>16.75002</td>
<td>21.19752</td>
<td>20.98</td>
</tr>
<tr>
<td>0.795</td>
<td>33.55008</td>
<td>42.39760</td>
<td>20.86</td>
</tr>
</tbody>
</table>

Table 2.3: Comparison of the waiting times when $1-A(\emptyset) = 0.2$
Figure 2.5 Sensitivity of Error with Traffic Intensity
2.5.2 With preventive maintenance \((t_0 < \infty)\)

We define the "total traffic" to be the traffic intensity of customers, plus the frequency of maintenance plus the frequency of breakdowns. While no direct comparison can be made, we can see from the following table that for different values of \(\lambda E[S]\) and \(t_0\), the expected waiting times \(E[W(t_0)]\) are in the same order of magnitude, when values of the total traffic intensity are close to each other.

<table>
<thead>
<tr>
<th>(\lambda E[S])</th>
<th>(t_0)</th>
<th>Total traffic</th>
<th>(E[W(t_0)])</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>1.058</td>
<td>0.999984</td>
<td>2552.29</td>
</tr>
<tr>
<td>0.75</td>
<td>1.368</td>
<td>0.999985</td>
<td>2744.18</td>
</tr>
<tr>
<td>0.80</td>
<td>1.843</td>
<td>0.999981</td>
<td>2088.70</td>
</tr>
<tr>
<td>0.85</td>
<td>2.671</td>
<td>0.999985</td>
<td>2515.57</td>
</tr>
<tr>
<td>0.90</td>
<td>4.559</td>
<td>0.999983</td>
<td>2081.43</td>
</tr>
<tr>
<td>0.95</td>
<td>21.540</td>
<td>0.999985</td>
<td>2485.06</td>
</tr>
</tbody>
</table>

* where the failure time distribution is Erlang-2 and \(\tau_m < \tau_r / 2\)

Table 2.4: Variation of the waiting time with respect to the traffic intensity

Figure 2.6 contrasts the expected waiting time when the off-time period is assumed to be zero, versus when it is a positive random variable. The failure time distribution is assumed to be Erlang-2 with the following parameters: \(\lambda E[S] = 0.8\), \(\tau_m = 0.3\), and \(\tau_r = 1\).
Figure 2.6: Comparison of the expected waiting times
2.6 THE AGE-MAINTENANCE OPTIMIZATION MODEL

The downtime while a facility is waiting for repair assistance is unproductive time. This downtime can be reduced by increasing the maintenance rate. The cost of jobs waiting, the cost of breakdown and the cost of maintenance are parts of our objective function.

2.6.1 Model formulation

Let $C_w$, $C_m$ and $C_b$ be the cost per unit time of waiting in the system, the fixed maintenance cost per occurrence, and the fixed breakdown cost per occurrence, respectively. Each cycle is a renewal process with mean $\varphi(t_0)$. This cycle is completed whenever each renewal takes place for a renewal reward process. Let $\psi(t_0)$ be the cost of PM/breakdown per unit time for an infinite time span. $\psi(t_0)$ is equal to the expected PM/breakdown cost incurred in a cycle divided by the expected length of a cycle. Thus,

$$\psi(t_0) = \frac{C_w + (C_m - C_b)F(t_0)}{\int_0^{to} \frac{F(t)dt}{\tau_0} + \frac{\tau - \tau_0}{\tau}F(t_0)}$$  \hspace{1cm} (26)
The following one-dimensional (R → R) nonlinear programming problem \([\text{NLP1}]\) minimizes the total expected cost, subject to a nonlinear constraint.

\[
\text{[NLP1]: \quad \text{MINIMIZE} \quad J(t_0) = C \cdot \lambda \cdot E[W(t_0)] + \psi(t_0) \quad (27)}
\]

\[
\text{s.t.}
\]

\[
\lambda E[S] - A(t_0) \leq 0 \quad (28)
\]

\[
t_0 \geq 0. \quad (29)
\]

Constraint (28) ensures that the expression of \(E[W(t_0)]\) to be always positive. When \(A(t_0)\) is less than \(\lambda E[S]\), \(E[W(t_0)]\) becomes negative. Constraint (28) is used as the boundary for maintenance.

2.6.2 Feasible age-maintenance policy

If \(t_0\) is feasible, then the queueing system is stable under \(t_0\).

Let \(r(t)\) be the failure rate function. By definition:

\[
r(t) = f(t) / \bar{F}(t) \quad (30)
\]

Let \(\tau_b\) be the mean failure time without PM,

\[
R_1 = \frac{\tau_r}{\tau_b (\tau_r - \tau_a)}, \quad \text{and} \quad t_0^* \quad \text{be the value of} \ t_0 \ \text{that maximizes} \ A(t_0).
\]
Proposition 2.1:

If \( r(\omega) \leq R_1 \) then \( t_A^* = \infty \) and \( A(t_0) \) is concave in \( t_0 \geq 0 \).
Otherwise \( t_A^* \) is unique and finite, and \( A(t_0) \) is concave in \( t_0 \) for \( t_0 \leq t_A^* \).

\( \text{Proof:} \)

Differentiating \( A(t_0) \) in Equation (9) with respect to \( t_0 \), and equating to zero:

\[
-f(t_0) \left( (\tau_r - \tau_m) \int_0^{t_0} \bar{F}(t) \, dt \right) + \tau_m \bar{F}(t_0) + (\tau_r - \tau_m) F(t_0) \bar{F}(t_0) = 0 \quad (31)
\]

Dividing (31) by \( \bar{F}(t_0) \) and \( (\tau_r - \tau_m) \), we have:

\[
r(t_0) \int_0^{t_0} \bar{F}(t) \, dt - F(t_0) = \tau_m / (\tau_r - \tau_m),
\]

Let \( q(t_0) = r(t_0) \int_0^{t_0} \bar{F}(t) \, dt - F(t_0) \),

then \( q'(t_0) = r'(t_0) \int_0^{t_0} \bar{F}(t) \, dt \)

Since \( r(t) \) is monotonically increasing, we have:

\[
q'(t_0) > 0 \quad \text{for all } t_0 \geq 0
\]

\[
q(\infty) = r(t_0) \quad \text{for all } t_0 \geq 0
\]

Therefore \( q'(t_0) > 0 \) and hence \( q(t_0) \) is strictly monotonically increasing.

Also, \( q(0) = 0 \) and \( q(\infty) = \tau_b r(\infty) - 1 \). Consider the following cases:

(i) \( r(\infty) \leq R_1 \)

This implies that \( q(\infty) = \tau_m / (\tau_r - \tau_m) \). Since \( q(t_0) \) is monotonically increasing then \( t_A^* = \infty \). Thus \( A'(t_0) \geq 0 \) for all \( t_0 \).

Let \( \delta = \tau_m / (\tau_r - \tau_m) - q(t_0) \) then \( \delta \geq 0 \) for all \( t_0 \geq 0 \).
\[ \frac{\delta^2 A(t_0)}{\delta t_0^2} \leq 0 \] for all \( t_0 \) if and only if

\[ -\left[r'(t_0)(\tau_r - \tau_m)\int_{t_0}^{t_0} \tilde{F}(t)dt + r(t_0)(\tau_r - \tau_m)\mathcal{A} \right] \varphi(t_0) - 2(\tau_r - \tau_m) \mathcal{A} \left[ \tilde{F}(t_0) + (\tau_r - \tau_m)f(t_0) \right] \leq 0 \]

Since \( \tau_r \leq \tau_m \) and \( \mathcal{A} \geq 0 \), therefore \( A(t_0) \) is concave all for \( t_0 \geq 0 \).

(ii) \( r(\omega) > R_1 \)

This implies that \( q(\omega) > \tau_m/(\tau_r - \tau_m) > 0 \). From the monotonocity and the continuity of \( q(t_0) \), there exists a unique and finite \( t^*_A \) satisfying (31) which maximizes \( A(t_0) \). For all \( t_0 \leq t^*_A \), we have \( \mathcal{A} \geq 0 \), implies \( A(t_0) \) is concave. On the other hand, for all \( t_0 > t^*_A \), \( A(t_0) \) is simply decreasing asymptotically in \( t_0 \) to \( A(\omega) = \tau_b/(\tau_b + \tau_r) \). Q.E.D.

Constraints (28) and (29) can be written as

\[ t_L \leq t_0 \leq t_R \]  \hspace{1cm} (32)

\( t_L \) and \( t_R \) are the boundaries of feasibility. They can be determined using numerical analysis methods on the following equation:

\[ A(t_0) = \lambda E[S] \]  \hspace{1cm} (33)
We assume that the system is stable without maintenance so that
\[ t_R = \infty, \quad \text{i.e., The feasible interval is } [t_L, \infty). \]

2.6.3 Optimal age-maintenance policy

In our model, some standard performance measures of queues appear in the objective function and the constraints. The tractability of our optimization model depends on whether these measures are convex or concave functions of \( t_0 \).

Proposition 2.2:

The total traffic intensity in the system is:
\[
\rho = 1 + \lambda E[S] - A(t_0)
\]  \hspace{1cm} (34)

Proof:

The expected number of repaired units and the number of preventive maintenance per unit time, respectively are:
\[
\lambda_b = F(t_0)/\varphi(t_0), \quad \text{and} \\
\lambda_m = \bar{F}(t_0)/\varphi(t_0).
\]

Then the total traffic intensity is:
\[
\rho = \lambda E[S] + \lambda_m \tau_m + \lambda_b \tau_r
\]
\[
= \lambda E[S] + \left( \tau_m \bar{F}(t_0) + \tau_r F(t_0) \right) / \varphi(t_0).
\]

We have
\[
\left( \int_0^{t_0} \bar{F}(t)dt \right) \frac{\tau_m \bar{F}(t_0) + \tau_r F(t_0)}{\varphi(t_0)} = 1 \Rightarrow
\]
\[
\left( \tau_m \bar{F}(t_0) + \tau_r F(t_0) \right) / \varphi(t_0) = 1 - A(t_0). \quad \text{Therefore, } \rho = 1 + \lambda E[S] - A(t_0)
\]

Q.E.D.
Since the expected waiting time must increase with the traffic intensity, we conjecture the following:

**Conjecture 1:**

\[ E[W(t_0)] \] is monotonically increasing in \( \rho \) (\( \rho < 1 \)). When \( \rho \) is close to one, \( E[W(t_0)] \) approaches infinity.

Substituting \( \rho \) in Equation (34) to Equation (21), the waiting time is:

\[
E[W(t_0)] = \frac{\lambda E[S^2]}{2(1 - \rho)} + \frac{V[X_{off}]}{\varphi(t_0)} A(t_0)^2 + \frac{V[X_{on}]}{\varphi(t_0)} \left(1 - A(t_0)\right)^2
\]

(35)

Clearly, as \( \rho \) approaches 1, \( E[W] \) approaches infinity. The first term of (35) is strictly increasing in \( \rho \) (\( \rho < 1 \)). Since the numerator of the second term of (35) cannot be written explicitly as a function of \( \rho \), we verified Conjecture 1 numerically. Tables 2.5 shows that as traffic intensity increases, the waiting time increases, and vice versa. This is also illustrated in Figure 2.7.
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$E[W]$</th>
<th>$\rho$</th>
<th>$E[W]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.98516</td>
<td>2.43161</td>
<td>0.99477</td>
<td>8.95561</td>
</tr>
<tr>
<td>0.98538</td>
<td>2.65424</td>
<td>0.99498</td>
<td>9.35800</td>
</tr>
<tr>
<td>0.98579</td>
<td>2.80604</td>
<td>0.99517</td>
<td>9.76574</td>
</tr>
<tr>
<td>0.98632</td>
<td>2.97839</td>
<td>0.99535</td>
<td>10.17771</td>
</tr>
<tr>
<td>0.98690</td>
<td>3.16688</td>
<td>0.99552</td>
<td>10.59274</td>
</tr>
<tr>
<td>0.98749</td>
<td>3.36914</td>
<td>0.99567</td>
<td>11.00958</td>
</tr>
<tr>
<td>0.98809</td>
<td>3.58390</td>
<td>0.99582</td>
<td>11.42698</td>
</tr>
<tr>
<td>0.98868</td>
<td>3.81049</td>
<td>0.99595</td>
<td>11.84365</td>
</tr>
<tr>
<td>0.98925</td>
<td>4.04858</td>
<td>0.99608</td>
<td>12.25833</td>
</tr>
<tr>
<td>0.98979</td>
<td>4.29808</td>
<td>0.99619</td>
<td>12.66976</td>
</tr>
<tr>
<td>0.99030</td>
<td>4.55897</td>
<td>0.99630</td>
<td>13.07672</td>
</tr>
<tr>
<td>0.99079</td>
<td>4.83132</td>
<td>0.99640</td>
<td>13.47802</td>
</tr>
<tr>
<td>0.99126</td>
<td>5.11517</td>
<td>0.99649</td>
<td>13.87256</td>
</tr>
<tr>
<td>0.99169</td>
<td>5.41056</td>
<td>0.99658</td>
<td>14.25930</td>
</tr>
<tr>
<td>0.99210</td>
<td>5.71749</td>
<td>0.99666</td>
<td>14.63728</td>
</tr>
<tr>
<td>0.99248</td>
<td>6.03587</td>
<td>0.99673</td>
<td>15.00564</td>
</tr>
<tr>
<td>0.99284</td>
<td>6.36554</td>
<td>0.99680</td>
<td>15.36362</td>
</tr>
<tr>
<td>0.99318</td>
<td>6.70626</td>
<td>0.99686</td>
<td>15.71055</td>
</tr>
<tr>
<td>0.99349</td>
<td>7.05767</td>
<td>0.99692</td>
<td>16.04588</td>
</tr>
<tr>
<td>0.99379</td>
<td>7.41933</td>
<td>0.99697</td>
<td>16.36916</td>
</tr>
<tr>
<td>0.99406</td>
<td>7.79067</td>
<td>0.99702</td>
<td>16.68003</td>
</tr>
<tr>
<td>0.99431</td>
<td>8.17103</td>
<td>0.99736</td>
<td>19.09331</td>
</tr>
<tr>
<td>0.99455</td>
<td>8.55963</td>
<td>0.99757</td>
<td>21.06978</td>
</tr>
</tbody>
</table>

Table 2.5: Variation of $E[W]$ and $\rho$ with $t_0$ when $\tau_{m \tau} = 0.1$
Figure 2.7 Sensitivity of $E[W]$ vs. Traffic Intensity
We will show that when the server is under the heavy traffic condition, the value of $t_0$ which maximizes the availability of the server, also minimizes the expected waiting time. The approximation of the waiting time listed in Theorem 2.1, is accurate only when the traffic intensity is greater than or equal to 0.9.

If the times for maintenance and repair are not negligible, then finding a closed-form solution for $t_0$ which minimizes the total expected cost is very difficult. We will provide two theorems for determining the optimal age-maintenance $t_0^*$ under certain specified conditions.

Theorem 2.2 provides the necessary conditions on the failure rate, the cost parameters, and the maintenance and repair times such that the age $t_A^*$ which maximizes the availability of the server also minimizes the expected total cost. The theorem also provides conditions where $t_A^*$ can be used as an upper bound for the optimal policy. Theorem 2.3 gives some conditions of the failure rate such that $t_0^*$ which minimizes the total cost, is infinite, i.e., no maintenance is needed. When the conditions of Theorems 2.2 and 2.3 are not met, standard numerical techniques (such as Newton-Raphson procedure) may be used for determining $t_0^*$. The above results will be used as a tool for the following chapters.
Proposition 2.3:

From Conjecture 1, and assuming that $\lambda E[S]$ is high, then for any failure time distribution, $t^*$ minimizes $E[W(t_o)]$.

Proof:

Let $t^*$ be the age-maintenance that minimizes the expected waiting time. If $E[W(t_o)]$ is monotonically increasing with respect to $\rho$, then $t^*_w$ also minimizes $\rho$, i.e., $\rho(t^*_w) = \rho_{\min}$. Since $\rho(t_o) = 1 + \lambda E[S] - A(t_o)$, this implies that $A(t^*_w) = 1 + \lambda E[S] - \rho_{\min}$. Therefore $A(t_o)$ is maximized at $t^*_w$. When $r(\omega) \leq R_1$, $A(t_o)$ is maximized only at $t_o = \infty$. Thus $t^*_A = \infty$ and $t^*_w = \infty$, therefore $t^*_A = t^*_w$. Preventive maintenance will not improve the expected waiting time in the system. On the other hand, when $r(\omega) > R_1$, the finite and unique value of $t_o$ which maximizes $A(t_o)$ is $t^*_A$. We know that $t^*_w$ also maximizes $A(t_o)$. Based on the uniqueness property, $t^*_A$ must be equal to $t^*_w$.

Q.E.D.

Figure 2.8 compares $t^*_A$ and $t^*_w$ for $\tau_m / \tau_r = 0.1$ and $\tau_m / \tau_r = 0.3$ under various values of the total traffic intensity. The failure time distribution is assumed to be Erlang-2. Note that when $\rho < 0.93$, $t^*_A$ and $t^*_w$ are equal.
Figure 2.8 Comparison of the two optimal ages
Once we have determined the value of \( t^*_A \) which satisfies (31), we can find conditions for the value of \( t^*_0 \) which minimizes the total expected cost.

**Proposition 2.4:**

Let \( R_2 = \frac{C_b}{b_\tau_m - C_\tau_m + \tau_b(C_b - C_m)} \).

If \( r(\omega) > R_2 \) and \( C_\tau_m \geq C_\tau_r \), then there exists a finite and unique \( t^*_\psi \) which minimizes \( \psi(t_0) \). If \( r(\omega) \leq R_2 \) and \( C_\tau_m \geq C_\tau_r \), then \( t^*_\psi = \infty \).

**Proof:**

Differentiating \( \psi(t_0) \) in (26) with respect to \( t_0 \) and equating it to zero, we have:

\[
\frac{f(t_0)[(C_\tau_m - C_\tau_r) + (C_b - C_m)\int_0^{t_0} \tilde{F}(t)dt] - C_m \tilde{F}(t_0) - (C_b - C_m)\tilde{F}(t_0)F(t_0)}{[\tau_m + (r_r - \tau_m)F(t_0) + \int_0^{t_0} \tilde{F}(t)dt]^2} = 0
\]

Dividing both sides of the above equation by \( \tilde{F}(t_0) \) and \( C_b - C_m \), we have

\[
r(t_0)\left( \frac{C_\tau_m - C_\tau_r}{C_b - C_m} + \int_0^{t_0} \tilde{F}(t)dt \right) - F(t_0) = \frac{C_m}{C_b - C_m}
\]

(36)
Let \( q(t_0) = r(t_0) \left( \frac{C_{b,m} \tau - C_{m} \tau}{C_{b} - C_{m}} \right) + \int_{0}^{t_0} \bar{F}(t) dt \) - F(t_0)

\[ q'(t_0) = r'(t_0) \left( \frac{C_{b,m} \tau - C_{m} \tau}{C_{b} - C_{m}} \right) + \int_{0}^{t_0} \bar{F}(t) dt. \]

We have \( C_{b,m} \geq C_{m} \tau \), then \( q'(t_0) \) is always positive and therefore \( q(t_0) \) is strictly monotonically increasing.

We have \( q(0) = 0 \) and \( q(\omega) = r(\omega) \left( \frac{C_{b,m} \tau - C_{m} \tau}{C_{b} - C_{m}} \right) + \tau \) - 1.

Consider the following cases:

(i) \( r(\omega) > R_2 \)

This implies that \( q(\omega) > C_{m} / (C_{b} - C_{m}) > q(0) \). From the monotonocity and the continuity of \( q(t_0) \) there exists a finite and unique \( t_\psi \) satisfying (35).

(ii) \( r(\omega) \leq R_2 \)

This implies that \( q(\omega) \leq C_{m} / (C_{b} - C_{m}) \), and \( q(t_0) \) is monotonically increasing. Thus \( \psi'(t_0) \leq 0 \) for all \( t_0 \). Therefore \( t_\psi^* = \omega \).

Q.E.D.
Theorem 2.2:

Suppose the failure rate \( r(t) \) is continuous and monotonically increasing, and \( r(\infty) > R_1 \). The following results hold:

(i) If \( C_{b \tau} = C_{m \tau} \) then there exists a finite and unique \( t_0^* \)

such that for all \( C_w > 0 \) \( t_0^* = t_A^* \).

(ii) If \( C_{b \tau} > C_{m \tau} \) then there exists a finite and unique \( t_0^* \)

such that \( \forall C_w > 0 \) \( t_0^* \leq t_A^* \).

Proof:

From (27), the total cost is equal to \( C_{w \lambda E}[W(t_0)] + \psi(t_0) \).

(i) From proposition 2.3, we have \( t_A^* \) that minimizes \( \text{E}[W] \). We will prove that \( t_A^* \) also minimizes \( \psi(t_0) \). We must show that \( \psi'(t_A^*) = 0 \).

Using (36), we must prove that there exists a unique and finite \( t_A^* \)

which satisfies the following:

\[
r(t_A^*)\left( \frac{C_{b \tau} - C_{m \tau}}{C_b - C_m} + \int_0^{t_A^*} F(t)dt \right) - F(t_A^*) = \frac{C_m}{C_b - C_m}. \tag{37}
\]

Differentiating \( A(t_0) \) in (9) with respect to \( t_0 \) and equating it to zero,

\[
- f(t_A^*)[(\tau_r - \tau_m)\int_0^{t_0} F(t)dt] + \tau_m F(t_0) + (\tau_r - \tau_m)F(t_0)F(t_0)
\]

\[
\frac{[\tau_m + (\tau_r - \tau_m)F(t_0) + \int_0^{t_0} F(t)dt]^2}{[\tau_m + (\tau_r - \tau_m)F(t_0) + \int_0^{t_0} F(t)dt]^{\frac{3}{2}}}
\]

= 0
Dividing both sides by \( \bar{F}(t_0) \) and \( \tau_m - \tau_r \), we have

\[
\frac{r(t_0)}{\tau_m - \tau_r} \int_0^{t_0} \bar{F}(t)dt - F(t_0) = \tau_m/(\tau_r - \tau_m) \tag{38}
\]

Note that \( r(0) > R_1 \). From Proposition 2.2, there exists a finite and unique \( t^*_A \) which satisfies \( (30) \). We have \( \frac{\tau_m}{\tau_r} = \frac{\tau_m}{\tau_r} \) which implies \( \frac{\tau_m}{\tau_r} = \frac{\tau_m}{\tau_r} \) and \( R_1 = R_2 \). Substituting in \( (38) \) we have:

\[
r(t^*_A) \left( \frac{\tau_m - \tau_r}{\tau_m - \tau_r} + \int_0^{t^*_A} \bar{F}(t)dt \right) - F(t^*_A) = \tau_m/(\tau_r - \tau_m) = \frac{\tau_m}{\tau_r}
\]

(ii) We must prove that \( J_1(t^*_A) < J_1(t_0) \) for all \( t_0 > t^*_A \).

Since \( t^*_A \) minimizes \( E[W] \) then \( E[W(t^*_A)] < E[W(t_0)] \) for all \( t_0 > t^*_A \). It remains to prove \( \psi(t^*_A) < \psi(t_0) \) for \( t_0 > t^*_A \).

We must show that \( \psi'(t_0) > 0 \) for \( t_0 > t^*_A \). From the proof of (i) we will prove the following:

\[
r(t_0) \left( \frac{\tau_m - \tau_r}{\tau_m - \tau_r} + \int_0^{t_0} \bar{F}(t)dt \right) - F(t_0) > \frac{\tau_m}{\tau_r} \tag{39}
\]

for all \( t_0 > t^*_A \).
$A(t_0)$ is decreasing for $t_0 > t^*$. This implies that $A'(t_0) < 0$ for $t_0 > t^*$. From the proof of (i), we have:

$$-r(t_0)(\tau - \tau_m) \int_0^{t_0} F(t) dt + F(t_0)(\tau - \tau_m) + \tau_m < 0 \text{ for } t_0 > t^*.$$

Hence $r(t_0) \int_0^{t_0} F(t) dt - F(t_0) > \frac{\tau}{\tau_m} - \frac{\tau}{\tau_m} \tau_m$ for $t_0 > t_A^*$ \quad (40)

We have $C_{b_m} > C_{m_r}$ which implies $\frac{\tau}{\tau_m} > \frac{C_m}{(C_b - C_m)}$.

$R_1 > R_2$ and let $d = \frac{C_{b_m} - C_{m_r}}{C_b - C_m} > 0$.

We have $r(\omega) > R_1 > R_2$ and $r(t_0)$ is strictly increasing with $t_0$, then substituting $d$ in (40), we have:

$$r(t_0)d + r(t_0) \int_0^{t_0} F(t) dt - F(t_0) > \frac{\tau}{\tau_m} - \frac{\tau}{\tau_m} \tau_m \text{ for } t_0 > t_A^*$$

Therefore $r(t_0) \left( d + \int_0^{t_0} F(t) dt - F(t_0) \right) > \frac{C_m}{(C_b - C_m)}$ for $t_0 > t_A^*$

Q.E.D.

The following theorem specifies when preventive maintenance is not worthwhile.

**Theorem 2.3:**

Suppose the failure rate $r(t)$ is continuous and monotonically increasing, $r(\omega) \leq R_2$. The following results hold:

If $C_{b_m} \geq C_{m_r}$ then $t_0^* = \omega$, i.e., no preventive maintenance is needed.
*Proof*:

We have $C_b \tau_m \geq C_m \tau_r$, this implies $R_1 \geq R_2$. Since $r(\infty) \leq R_2$, then $r(\infty) \leq R_1$. Using propositions (2.2) and (2.4), we have $t^*_A = \infty$ and $t^*_\psi = \infty$. Consequently, the total cost is minimized at $t^*_0 = \infty$.

Q.E.D.

Theorems 2.2 and 2.3 provide some results concerning the optimal $t^*_0$.

For other cases, a numerical method can be used to determine $t^*_0 \geq t_L$ where $t_L$ is the lower bound of feasibility.

### 2.9 Numerical Example

We illustrate Theorems (2.2) and (2.3) of Section 2.8 using the following parameters:

1. The failure time has Erlang$_2$ distribution, i.e.,
   
   \[ F(t_0) = 1 - (1 + \beta t_0) e^{-\beta t_0} \]
   
   and $\beta = 0.1$.

2. The maintenance and repair times have exponential distributions.

3. The interarrival and service times for jobs have exponential distributions.

4. Let $r(\infty) = 0.1$, $C_b = 1000$, $C_\psi = 1$, and $\lambda E[S] = 0.95$. 
Differentiating $A(t_0)$ in (9) with respect to $t_0$ and equating it to zero, we have:

$$\frac{2/\beta + t_0}{1 + \beta t_0} \left(1 + \beta t_0(1 - \beta)\right) e^{-\beta t_0} + \frac{2\beta t_0}{1 + \beta t_0} = \frac{\tau_r}{\tau_r - \tau_m}$$

(41)

Let $t_A^*$ be the solution of equation (41). From Theorem 2.2, if $\tau > 2\tau_m$ and $C_B \tau_m - C_B \tau_r = 0$, then $t_0^*$ is unique and finite such that $t_0^* = t_A^*$. The following table compares the values of $t_A^*$ and $t_0^*$:

<table>
<thead>
<tr>
<th>$\tau_r$</th>
<th>$\tau_m$</th>
<th>$C_m$</th>
<th>$C_B \tau_m - C_B \tau_r$</th>
<th>$t_A^*$</th>
<th>$t_0^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>100</td>
<td>0</td>
<td>6.8</td>
<td>6.40</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>200</td>
<td>0</td>
<td>13.2</td>
<td>12.97</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>300</td>
<td>0</td>
<td>23.3</td>
<td>23.24</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>400</td>
<td>0</td>
<td>49.5</td>
<td>49.25</td>
</tr>
</tbody>
</table>

Table 2.6: Comparison of the optimal policies

Clearly, the values of $t_A^*$ and $t_0^*$ are very close. When $t_A^*$ is large, the two values are almost identical. Note that $t_A^*$ and $t_0^*$ are not exactly equal due to numerical errors.

From Theorem 2.2, if $\tau > 2\tau_m$ and $C_B \tau_m > C_B \tau_r$, then $t_0^* \leq t_A^*$. Also, $t_A^*$ satisfies equation (41). Table 2.8 shows that this is true. Therefore $t_A^*$ can be used as an upper bound for $t_0^*$. 
<table>
<thead>
<tr>
<th>$\tau_r$</th>
<th>$\tau_m$</th>
<th>$C_m$</th>
<th>$C_{b\tau_m} - C_{m\tau_r}$</th>
<th>$t_A^*$</th>
<th>$t_0^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>150</td>
<td>50</td>
<td>6.8</td>
<td>5.47</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>450</td>
<td>50</td>
<td>49.5</td>
<td>40.27</td>
</tr>
</tbody>
</table>

Table 2.7: Comparison of the optimal policies

Theorem 2.3 says that if the following holds: $R_2 \geq (\beta = 0.1)$, and $C_{b\tau} \geq C_{m\tau}$, then $t_0^*$ is infinite (no PM). Table 2.8 shows that this true.

<table>
<thead>
<tr>
<th>$\tau_r$</th>
<th>$\tau_m$</th>
<th>$C_b$</th>
<th>$C_m$</th>
<th>$R_2$</th>
<th>$t_A^*$</th>
<th>$t_0^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>1000</td>
<td>500</td>
<td>0.10</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>1</td>
<td>0.7</td>
<td>1000</td>
<td>600</td>
<td>0.12</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>1</td>
<td>0.8</td>
<td>1000</td>
<td>700</td>
<td>0.16</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>1</td>
<td>0.9</td>
<td>1000</td>
<td>800</td>
<td>0.24</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
</tbody>
</table>

Table 2.8: Comparison of the optimal policies
2.9 CONCLUSION

This chapter presented some analytical results for the optimal age-maintenance policy in a queueing system with breakdowns. We studied a simple model with one server, where the failure time can have general distribution. Basically, the advantage of using queueing models is that the cost of the time a customer waits in the system can be taken into account. For general up-time distribution, it does not appear possible to compute exactly the expected waiting time in the system. When \( \lambda E[S] \geq 0.9 \) and \( 1 - \lambda(t_0) \leq 0.1 \), the diffusion approximation method can provide accurate computation of the waiting time. Our aim is to study queueing systems with average on-times that are greater than the average off-times. The optimal expected total cost can be easily determined when the conditions in Theorems 2.1 and 2.2 on the cost parameters, the failure rate function, and the maintenance and repair times are met. For possible extensions of this chapter, one may consider a queueing system which is not under the heavy traffic assumption.
CHAPTER THREE

Age-Maintenance Policy of an M/M/1 Queue with

Buffer and Quality Costs
Notation

$\beta$ the arrival rate of a breakdown after the last repair or PM

$C_b$ the cost of each breakdown

$C_\xi$ the cost of each waiting capacity per unit time

$C_m$ the cost of each maintenance

$C_q$ the cost of each defective item per unit time

$f(t)$ the probability density function that a breakdown occurs at time $t$ since the last repair or maintenance

$F(t_0)$ the probability that the server fails before $t_0$

$\bar{F}(t_0)$ the probability that the server is switched off before breakdown.

$L$ the available waiting space (threshold)

$Q$ the allowable defective items (threshold)

$r(t)$ the failure rate $= f(t)/\bar{F}(t)$

$\tau_b$ the mean time of the working server without PM $(1/\beta)$

$\tau_m$ the average time to perform a maintenance

$\tau_r$ the average repair time

$\lambda$ the job's arrival rate

$E[W(t_0)]$ the expected waiting time in the system under $t_0$

$\varphi(t_0)$ the average cycle time

$A(t_0)$ the mean fraction of time the server is working under $t_0$

$t_A^*$ the age-maintenance $t_0$ such that the availability is maximized

$t_1^*$ the age-maintenance $t_0$ such that the cost in Chapter 2 is minimized

$t_L$ the lower bound of age-maintenance so that the queueing system is stable
3.1 BACKGROUND

Just-In-Time (JIT) production forces a different mind set for equipment maintenance - the understanding that facilities need proper care and attention in order to produce the right quantity of a product as needed, with the highest possible quality. This mind set, which is known as "total productive maintenance", includes: (1) maintenance as part of the business strategy, (2) elimination of breakdowns, (3) inclusion of maintenance as part of the reliability and quality design.

Some models on production systems with maintenance and/or quality control can be found in Thompson (1968), Vickson (1982), and Tapeiro (1986). These models are finite-planning-horizon models where the maintenance action is of the "bang-bang" type, i.e., the maintenance rate is either at zero or at the maximum limit. None of these models addresses the ramification of maintenance on the waiting line. As maintenance reaches a certain limit, the available waiting capacity may be exceeded. Therefore, there is a trade-off to be considered.

The problem of concern in this chapter is to determine the optimal age-maintenance policy which minimizes the cost of producing defective items, the cost of breakdowns and the cost of additional waiting capacity. The model is illustrated in Figure 3.1.
**Figure 3.1:** M/M/1 queueing system with quality control and limited waiting capability
3.2 AVAILABLE WAITING CAPACITY

The waiting line (or buffer) is the place for which jobs wait until they are loaded for processing. When the waiting line of a queueing system with finite capacity reaches a certain length, no further customer can be allowed to join the queue until a space becomes available by the departure of a customer. For our model, we assume that when a customer arrives and the waiting line is full, the customer will not be lost. Instead, the waiting capacity is increased at a certain cost. With this assumption, the waiting time is computed as if the waiting capacity is infinite.

From Little's formula the expected number of jobs in the system is equal to $\lambda E[W]$, where $\lambda$ and $E[W]$ are the arrival rate and the expected waiting time in the system, respectively. From Chapter 2, the probability that $W > x \ (x > 0)$ is:

$$P[W > x] = e^{-x/E[W]} \quad (1)$$

Let $L$ be the available waiting capacity and $y$ be the amount of waiting time in the system such that the expected number of customer waiting in the system is equal to $L$. From little result, we define $y = L/\lambda$. Thus the probability that the waiting time is greater than $y$ is:

$$P[W > y] = P[W > L/\lambda] = e^{-L/(\lambda E[W])} \quad (2)$$

Suppose we require the following constraint:

$$P[W > L/\lambda] < \theta \quad (3)$$

In our model, $L$ is a decision variable and $0 \leq \theta \leq 1$. 
3.3 QUALITY REQUIREMENT

The objective of quality control includes reducing the number of jobs for rework, maintaining the desired degree of conformance to design specification, etc. In this chapter, we are interested in how maintenance improves quality. Our objective is to determine the trade-off between the cost of poor quality and the cost of additional buffer, and the cost of maintenance and breakdown.

It has been observed that maintenance not only decreases the probability of failure, but also decreases the probability of producing defective parts (see for example Tapeiro 1986). Suppose the output of the system is not always acceptable. The output process can be decomposed into an "acceptance" process and a "rejection" process.

Let $q(t)$ be the probability that an item produced at time $t$ is not defective such that $0 \leq q(t) \leq 1$. Let $u(t)$ be the "degradation" state of the server at time $t$. At the "new" condition, $u = 0$. Higher values of $u$ represent poorer quality and vice versa.

Our purpose is to integrate quality control using maintenance action into an $M/M/1$ queueing system. We make the following assumptions:

1. The server produces no defective items at time zero, i.e., $q(0) = 1$.
2. The maintenance/repair action restores the server to the new condition (see Chapter 2).
3. The service process is exponentially distributed.
\( q(t+dt) \) is the probability that an item produced at time \( t \) is not defective minus the probability that an item produced in \( (t, t+dt) \) is not defective due to the growth rate of defective items, i.e.,

\[
q(t+dt) = q(t) - u(t)q(t)dt, \quad \text{or} \quad [q(t+dt) - q(t)] / dt = -u(t)q(t).
\]

Therefore

\[
q'(t) = -u(t)q(t) \quad \text{and} \quad q(0) = 1.
\]

The average probability of producing non-defective item in \([0, t_0]\) is:

\[
\bar{q} = \frac{\int_0^{t_0} q(t)dt}{t_0}.
\]

Suppose the server is operating without failure up to \( t_0 \), then the inter-departure time of items in \([0, t_0]\) is exponentially distributed. In Chapter 2, we have restricted the on-time period to be sufficiently longer than the off-time period due to the heavy traffic condition and under the stability condition of the queueing system, the output rate is equal to the arrival rate. Consequently, the expected number of produced items up to \( t_0 \), can be approximated to \( \lambda t_0 \).
Let \( X(t_0) \) be the number of departures up to time \( t_0 \) and \( Y(t_0) \) be the cumulative number of defective items before interrupting the queueing systems for maintenance. Let \( Q \) be the threshold value for \( Y(t_0) \). The stochastic process \( \{ X(t_0), t_0 > 0 \} \) is a Poisson process with mean \( \lambda t_0 \). The number of defective items at the \( n \)-th departure \( \xi_n \) is:

\[
\xi_n = \begin{cases} 
0 & \text{with probability } \tilde{q} \\
1 & \text{with probability } (1-\tilde{q})
\end{cases}
\]

and \( E[\xi_n] = 1 - \tilde{q} \) and \( V[\xi_n] = \tilde{q}(1 - \tilde{q}) \).

Thus \( \{ \xi_n, n > 0 \} \) is a family of independent and identically distributed random variables of \( \{ X(t_0), t_0 > 0 \} \). Furthermore, we have:

\[
Y(t_0) = \sum_{n=1}^{X(t_0)} \xi_n.
\]

The stochastic process \( \{ Y(t_0), t_0 > 0 \} \) is said to be a marked Poisson process. It can be considered also as a compound Poisson process.

Suppose we require the probability that the number of defective items up to time \( t_0 \) is less than or equal to \( Q \) be greater or equal to \( 1 - \theta \), i.e.,

\[
P[Y(t_0) \leq Q] \geq 1 - \theta.
\]
Assume that the expected number of items produced up to $t_0$ is sufficiently large so that $P[Y(t_0) \leq Q]$ can be approximated by a standard normal distribution. It follows that:

$$P[Y(t_0) \leq Q] \approx \phi \left( \frac{Q - \lambda t_0 (1-q)}{\sqrt{\lambda t_0 q (1-q)}} \right).$$

Therefore,

$$\frac{Q - \lambda t_0 (1-q)}{\sqrt{\lambda t_0 q (1-q)}} \approx \Phi^{-1}(1-\theta).$$

$Q$ is also a decision variable in our model.

### 3.4 The Maintenance Optimization Model

The problem is to find the optimal policy that minimizes the total expected cost over an infinite planning horizon. Whereas numerical methods can be used to obtain the optimal solution in the general case, an analytical approach to a special case of the model will be developed here.
3.4.1 Model formulation

Let \( C_\ell \) be the cost per unit time of each waiting space and \( C_q \) be the cost of producing a defective item per unit time. We assume that \( C_\ell \) and \( C_q \) are linear with respect to \( L \) and \( Q \), respectively. The following integer nonlinear programming problem \([\text{INLP1}]\) minimizes the total expected cost:

\[
\text{[INLP1]: MINIMIZE } J(L, Q, t_0) = J_1(t_0) + C_q Q + C_\ell L \tag{9}
\]

s.t.

\[
\lambda E[S] - A(t_0) \leq 0 \tag{10}
\]

\[e^{-L/\lambda E[W(t_0)]} \leq \theta \tag{11}\]

\[
(1-\bar{q})\lambda t_0 + \Phi^{-1}(1-\theta)\sqrt{(1-\bar{q})q\lambda t_0} \leq Q \tag{12}
\]

\[t_0 \geq 0 \tag{13}\]

\[L = 1, 2, 3, \ldots \]

\[Q = 1, 2, 3, \ldots \]

Where \( J_1(t_0) = C_w \lambda E[W(t_0)] + \frac{C_m + (C_m - C_m)F(t_0)}{\varphi(t_0)} \)

\( \theta \) is the "tolerance factor" specified by the decision maker.

Constraint (10) forces \( E[W(t_0)] \) to be always positive. Constraint (11) implies \( P[W > L/\lambda] \leq \theta \) and constraint (12) implies \( P[Y(t_0) > Q] \leq \theta \). Constraints (10) and (11) can be combined as:

\[
0 \leq E[W(t_0)] \leq -\frac{L}{\lambda \log(\theta)} \tag{14}
\]
In order to determine the optimal policy, we consider two separate subproblems. For the first and the second subproblems, we determine the joint optimal \((L^*, t_L^*)\) and \((Q^*, t_Q^*)\), respectively. Once we determine the couple \((L^*, Q^*)\), an efficient algorithm will be developed to determine the optimal policy \((L^*, Q^*, t_0^*)\).

### 3.4.2 Subproblem 1

The following subproblem is formulated for \(C_q = 0\) and without Constraint (12).

\[
\text{[SUB1]: MINIMIZE } J(L, t_0) = J_1(t_0) + C_q L
\]

s.t.

\[
0 \leq E[W(t_0)] \leq -\frac{L}{\lambda \log(\theta)}
\]

\(L = 1, 2, 3, \ldots\)

Let \(t_0^*\) be the value of \(t_0\) which minimizes the expected waiting time. Recall that \(t_0^*\) is determined in Equation (31) in Chapter 2.

From Equation (14), the lower bound of the waiting capacity \(L_0\) is equal to the smallest integer greater or equal to \(-\lambda E[W(t_A^*)] \log(\theta)\).

Figure 3.2 shows the expected waiting time as a function of age-maintenance. In Figure 3.2(a) where \(L > L_0\), the feasible region of the age-maintenance is \(t_L(L) \leq t_0 \leq t_r(L)\). In Figure 3.2(b) where \(L = L_0\) or \(E[W(t_A^*)] = -\frac{L}{\lambda \log(\theta)}\), the point of tangency \(t_0 = t_A^*\) is the only feasible age-maintenance. Therefore the optimal age-maintenance for \(L = L_0\) is \(t_A^*\). When \(L < L_0\), there is no feasible age-maintenance (see Figure 3.2 (c)).
Figure 3.2: Expected waiting time as a function of age-maintenance
Given a feasible value of L, let $t_0(L)$ be the feasible age-maintenance where the waiting capacity equals L. Suppose $t_\xi(L)$ and $t_r(L)$ are the values of $t_0$ in Figure 3.2(a) such that:

$$E[W(t_0)] = -\frac{L}{\lambda \log(\theta)}. \quad (15)$$

Thus $t_0(L)$ is feasible if and only if:

$$t_\xi(L) \leq t_0(L) \leq t_r(L). \quad (16)$$

Let the interval of feasibility $\delta(L) = t_r(L) - t_\xi(L)$. Clearly $\delta(L_0) = 0$. Given $\theta = 0.1$, Table 3.1 shows the sensitivity of $\delta(L)$ with respect to L, for $\tau_a/\tau_r = 0.1$ and 0.3. For $\tau_a/\tau_r = 0.1$, the interval of feasibility is illustrated in Figure 3.3. The following lemma shows that the region of feasibility is an increasing function of L.
The failure time distribution is Erlang-2, $\beta=0.05$, and $\lambda E[S]=0.9$

<table>
<thead>
<tr>
<th>$\tau_m/\tau_r$</th>
<th>$L$</th>
<th>$t_\xi(L)$</th>
<th>$t_r(L)$</th>
<th>$\delta(L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_0 = 31$</td>
<td>6.60</td>
<td>6.60</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>5.05</td>
<td>7.01</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3.95</td>
<td>9.64</td>
<td>5.69</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>3.43</td>
<td>11.72</td>
<td>8.28</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>35</td>
<td>3.15</td>
<td>13.85</td>
<td>10.70</td>
</tr>
<tr>
<td>36</td>
<td>2.95</td>
<td>16.11</td>
<td>13.16</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>2.73</td>
<td>18.65</td>
<td>15.92</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>2.55</td>
<td>21.55</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>2.45</td>
<td>24.95</td>
<td>22.50</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>2.38</td>
<td>29.12</td>
<td>26.74</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>2.25</td>
<td>34.25</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>2.20</td>
<td>40.80</td>
<td>38.6</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>2.10</td>
<td>48.70</td>
<td>46.6</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>2.08</td>
<td>58.00</td>
<td>55.92</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>2.03</td>
<td>69.50</td>
<td>67.47</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>2.00</td>
<td>81.90</td>
<td>79.90</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>1.95</td>
<td>99.70</td>
<td>97.75</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>1.87</td>
<td>132.00</td>
<td>130.13</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>1.83</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.80</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td></td>
</tr>
</tbody>
</table>

| $L_0 = 41$       | 20.70 | 20.70 | 0 |
| 42               | 14.35 | 33.10 | 18.75 |
| 43               | 12.35 | 43.90 | 31.55 |
| 44               | 11.10 | 54.90 | 43.80 |
| 45               | 10.35 | 66.60 | 56.25 |
| 0.3              | 46   | 9.60   | 80.15  | 70.55 |
| 47               | 9.08  | 98.10  | 89.02  |
| 48               | 8.60  | 130.70 | 122.10 |
| 49               | 8.25  | $\infty$ | $\infty$ |
| 50               | 7.95  | $\infty$ | $\infty$ |
| 51               | 7.65  | $\infty$ | $\infty$ |

Table 3.1: Sensitivity of $L_0$ vs. $\tau_m$
Figure 3.3: Variation of \( \delta(L) \) with respect to \( L \)
Lemma 3.1:

\( \delta(L) \) is a non-decreasing function of \( L \), for \( L = L_0, L_0+1, \ldots \).

**Proof:**

We must show that \( t^*_l(L+1) \leq t^*_l(L) \) and \( t^*_r(L+1) \geq t^*_r(L) \) for all \( L \geq L_0 \).

In Chapter 2, we defined \( t^*_l \) as the age-maintenance such that the fraction of time the server is on is equal to \( \lambda E[S] \), the traffic intensity. Therefore \( t^*_l \geq t^*_L \). \( E[W(t^*_l)] \) is non-increasing in \( (t^*_L, t^*_A) \) and it is non-decreasing in \( (t^*_A, \infty) \). Since \( t^*_l(L) \) and \( t^*_r(L) \) are the roots of the following: \( E[W(t^*_l)] = -L / \lambda \log(\theta) \) then for all \( L \geq L_0 \), we have:

\[
\begin{align*}
    t^*_l(L) & \leq t^*_l(L) \leq t^*_A, \\
    t^*_A & \leq t^*_r(L) \leq \infty.
\end{align*}
\]

We know that for \( L = L_0 \), we have: \( t^*_l(L) = t^*_r(L) = t^*_A \); for \( L = \infty \), we have: \( t^*_l(L) = t^*_L \) and \( t^*_r(L) = \infty \). Consequently, as \( L \) increases, \( t^*_l(L) \) approaches \( t^*_L \), and \( t^*_r(L) \) approaches infinity.

Q.E.D.

3.4.2.1 Optimal age-maintenance policy given \( L \)

Let \( t^*_1 \) be the value of \( t^*_0 \) that minimizes \( J_1(t^*_0) \) subject to Constraints (10) and (13) only. The following theorem shows that if \( t^*_1 \) is outside the feasible interval defined in Equation (16), then the upper or lower limit of that interval is the optimal age-maintenance for a given \( L \). This is illustrated in Figure 3.4.
Figure 3.4: Possible cases of the feasible region with $t^*$.
Lemma 3.2:

Let $t^*_L$ be the value of $t_0$ which minimizes the total expected cost in [SUB1], given $L$. The following relations hold:

(a) If $t^*_L \leq t_1^* \leq t_r(L)$ then $t^*_L = t_1^*$.

(b) If $t_1^* \leq t^*_L$ then $t^*_L = t^*_L$.

(c) If $t_1^* \geq t_r(L)$ then $t_L^* = t_r(L)$.

Proof:

(a) $t^*_L \leq t_1^* \leq t_r(L)$

Since $t_1^*$ minimizes $J_i(t_0)$, $C_i(L)$ is increasing with $L$ and $t_1^*$ is feasible, therefore, $t_1^*$ is optimal.

(b) $t_1^* \leq t^*_L$

Given $L$, the age-maintenance $t_0(L)$ is feasible if $t^*_L \leq t_0(L)$. We have $J_i(t_1^*) < J_i(t_0)$ for all $t_0 > t_1^*$. Therefore the smallest feasible $t_0(L)$ greater or equal to $t_1^*$ must be optimal. Thus $t^*_L = t^*_L$.

(c) $t_1^* \geq t_r(L)$

Given $L$, the age-maintenance $t_0(L)$ is feasible if $t_r(L) \geq t_0(L)$. We have $J_i(t_1^*) > J_i(t_0)$ for all $t_0 < t_1^*$. Therefore the biggest feasible $t_0(L)$ less or equal to $t_1^*$ must be optimal. Thus $t_L^* = t_r(L)$.

Q.E.D.
3.4.2.2 Optimal waiting capacity \( (L^*) \)

To reduce the region of search for finding the optimal waiting capacity \( L^* \) of [SUB1], we discovered a useful property. This property is stated in the following theorem.

Theorem 3.1:

Given \( L \), the optimal age-maintenance is \( t_L^* \). The cost function \( J_1(t_0) + C_L t \) evaluated at \( t_L^* \) is unimodal with respect to \( L \).

Proof:

From [SUB1], the optimal expected cost for a given value of \( L \) is:

\[
J(L, t_L^*) = J_1(t_L^*) + C_L t_L.
\]

Let \( L_0 \) be the smallest feasible capacity. We have:

\[
C_L t_0^* < C_L (L_0 + 1) < \ldots < C_L (L_0 + k) < C_L (L_0 + k + 1) < \ldots .
\]

Let \( (L_0 + k) \) be the smallest value of \( L \) such that \( t_L^* (L) \leq t_1 \leq t_r (L) \).

\( J_1(t_L^*) \) is decreasing in \( L \) up to \( L_0 + k \), and it is constant when \( L > L_0 + k \). Therefore we have:

\[
(1) \quad J_1(t_{L_0}^*) > J_1(t_{L_0+1}^*) > \ldots > J_1(t_{L_0+k}^*) = J_1(t_{L_0+k+1}^*) = \ldots
\]

\[
(2) \quad \left( J_1(t_{L_0}^*) - J_1(t_{L_0+1}^*) \right) \geq \left( J_1(t_{L_0+1}^*) - J_1(t_{L_0+2}^*) \right) \geq \ldots
\]

\[
\geq \left( J_1(t_{L_0+k}^*) - J_1(t_{L_0+k+1}^*) \right) = \left( J_1(t_{L_0+k+1}^*) - J_1(t_{L_0+k+2}^*) \right) = \ldots
\]
Note that buffer $L_1$ is better than buffer $L_2$ if and only if

$$J_1(t_{L2}^*) + C_{L2} > J_1(t_{L1}^*) + C_{L1}$$

The above inequality can be written as:

$$J_1(t_{L2}^*) - J_1(t_{L1}^*) + C_{L2} - C_{L1} > 0 \quad (18)$$

First it will be shown that if buffer $L_1$ is better than $L_1+1$, then it must be better than all buffers $L_2$ with $L_2 \geq L_1 + 1$. The proof is by induction. Assume that (18) is true for $L_2$, we will show that (18) is true for $L_2+1$. Using the induction hypothesis we will prove that:

$$J_1(t_{L2+1}^*) - J_1(t_{L1+1}^*) + C_{L2} - C_{L1} > 0 \quad (19)$$

Since it is true for $L_1+1$, then $C_{L2} > J_1(t_{L1}^*) - J_1(t_{L1+1}^*)$

Also we have:

$$J_1(t_{L1}^*) - J_1(t_{L1+1}^*) \geq J_1(t_{L2}^*) - J_1(t_{L2+1}^*), \text{ for } L_1 < L_2$$

Therefore $C_{L2} > J_1(t_{L2}^*) - J_1(t_{L2+1}^*)$. Substituting

$$J_1(t_{L2}^*) < (J_1(t_{L2+1}^*) + C_{L2}) \text{ in (18), } \text{ we have }$$

$$0 < J_1(t_{L2}^*) - J_1(t_{L1}^*) + C_{L2} + C_{L1} - C_{L1} - C_{L2}$$

$$J_1(t_{L2+1}^*) - J_1(t_{L1}^*) + C_{L2} - C_{L1}$$

Now suppose that buffer $L_1$ is better than buffer $L_1-1$, we will show that buffer $L_1$ is better than buffer $L_2$ for all $L_2 \leq L_1 - 1$. The proof is also by induction. We will prove that:
\[ J_1(t_{L2-1}^\ast) - J_1(t_{L1}^\ast) + C_\ell (L_2 - L_1) - C_\ell > 0 \quad (20) \]

Since (20) is true for \( L_{1-1} \), then \( C_\ell < J_1(t_{L1-1}^\ast) - J_1(t_{L1}^\ast) \).

Also we have:

\[ J_1(t_{L1-1}^\ast) - J_1(t_{L1}^\ast) \leq J_1(t_{L2-1}^\ast) - J_1(t_{L2}^\ast), \quad L_1 > L_2. \]

Therefore, \( J_1(t_{L2-1}^\ast) - J_1(t_{L2}^\ast) > C_\ell \).

By induction, \( J_1(t_{L2}^\ast) - J_1(t_{L1}^\ast) + C_\ell (L_2 - L_1) > 0 \) \quad (21)

Substituting \( J_1(t_{L2}^\ast) < (J_1(t_{L2-1}^\ast) - C_\ell) \) in (21), we have:

\[ 0 < J_1(t_{L2}^\ast) - J_1(t_{L1}^\ast) + C_\ell (L_2 - L_1) < \]

\[ J_1(t_{L2-1}^\ast) - J_1(t_{L1}^\ast) + C_\ell (L_2 - L_1) - C_\ell \]

Q.E.D.

Let \( \bar{L} \) be the smallest waiting capacity such that \( t_{1}^\ast \) is within the feasible interval, then \( \bar{L} \) is the upper bound. If the total cost with respect to \( L_0 \) is less than the total cost with respect to \( L_0 + 1 \), then \( L_0 \) must be optimal. The joint optimal \((L_{l}^\ast, t_{l}^\ast)\) is determined using the unimodality property. Table 3.2 and Figure 3.5 illustrate the above theorem.
The failure time distribution is Erlang-2, $\beta=0.05$, $t_1^* = 7.35$, $\tau_s/\tau_r = 0.1$, $\theta = 0.1$, and $\lambda E[S]=0.9$.

<table>
<thead>
<tr>
<th>L</th>
<th>$t_1^*$</th>
<th>$J_1(t_1^*)$</th>
<th>$J_1(t_1^*) + L$</th>
<th>$J_1(t_1^*) + 3L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>6.60</td>
<td>42.25</td>
<td>73.25</td>
<td>135.25</td>
</tr>
<tr>
<td>32</td>
<td>7.01</td>
<td>41.10</td>
<td>73.10</td>
<td>137.10</td>
</tr>
<tr>
<td>33</td>
<td>7.35</td>
<td>40.44</td>
<td>73.44</td>
<td>139.44</td>
</tr>
<tr>
<td>34</td>
<td>7.35</td>
<td>40.44</td>
<td>74.44</td>
<td>142.44</td>
</tr>
<tr>
<td>35</td>
<td>7.35</td>
<td>40.44</td>
<td>75.44</td>
<td>145.44</td>
</tr>
<tr>
<td>36</td>
<td>7.35</td>
<td>40.44</td>
<td>76.44</td>
<td>149.44</td>
</tr>
<tr>
<td>37</td>
<td>7.35</td>
<td>40.44</td>
<td>77.44</td>
<td>152.44</td>
</tr>
<tr>
<td>38</td>
<td>7.35</td>
<td>40.44</td>
<td>78.44</td>
<td>155.44</td>
</tr>
<tr>
<td>39</td>
<td>7.35</td>
<td>40.44</td>
<td>79.77</td>
<td>158.44</td>
</tr>
<tr>
<td>40</td>
<td>7.35</td>
<td>40.44</td>
<td>80.44</td>
<td>161.44</td>
</tr>
</tbody>
</table>

* denotes optimal

Table 3.2: Optimal waiting capacities
Figure 3.5: Total cost of [SUB1] vs. buffer
3.4.3 Subproblem 2

We examine the possibility of a closed-form solution to a special case where the deterioration state is assumed to be constant in \([0, t_0]\) such that \(u(t) = u \ (u > 0)\). Using the differential equation methods, we have \(q(t) = e^{-ut}\). Furthermore the average probability that a produced item is not defective, is given by:

\[
\bar{q} = \frac{1-e^{-ut_0}}{ut_0}.
\]

Note that \(\bar{q}\) is non-increasing function with respect to \(t_0\). Figure 3.6 shows graphically the behavior of \(\bar{q}\) with respect to \(t_0\) for different values of \(u\). [SUB2] is formulated for the case where \(C_t = 0\).

Let \(d(t_0) = \frac{\lambda(ut_0 + e^{-ut_0} - 1)}{u} + Z \sqrt{\frac{\lambda(ut_0 + e^{-ut_0} - 1)(1-e^{-ut_0})}{u^2t_0}}\)

**[SUB2]:** \(\text{MINIMIZE} \quad J(Q, t_0) = J_1(t_0) + C_Q Q\)

\[\text{s.t.}\]
\[d(t_0) - Q \leq 0\]
\[\lambda E[S] - A(t_0) \leq 0\]
\[Q = 0, 1, 2, \ldots\]

Where \(Z = \phi^{-1}(1-\theta)\).
Figure 3.6: Probability of non-defectives vs. to
Let $Q_0$ smallest integer greater or equal to $d(t_L)$. Thus $Q$ is feasible if and only if $Q \geq Q_0$. For each feasible value of $Q$, consider the age-maintenance $t_0(Q)$ such that $d(t_0) = Q$. Since $d(t_0)$ is non-decreasing with respect to $t_0$, then $t_0(Q)$ is unique. Now we can determine easily the optimal age-maintenance $t^*_Q$ for a given $Q$. When $t_0(Q) \leq t^*_1$ then $t^*_Q = t_0(Q)$, otherwise $t^*_Q = t^*_1$ and $Q$ is an upper bound denoted by $\bar{Q}$. To reduce the search for $Q^*$, we consider the following theorem:

**Theorem 3.2:**

Given $Q$, the optimal age-maintenance is $t^*_Q$. The cost function $J_1(t_0) + C_Q$ evaluated at $t^*_Q$ is unimodal with respect to $Q$.

**Proof:**

From [SUB2], the optimal expected cost for a given value of $Q$ is:

$$J(Q, t^*_Q) = J_1(t^*_Q) + C_Q.$$

Let $Q_0$ be the smallest feasible threshold. We have:

$$C_{Q_0} < C_Q(Q_0 + 1) < \ldots < C_Q(Q_0 + k) < C_Q(Q_0 + k + 1) < \ldots.$$

Let $(Q_0 + k)$ be the smallest value of $Q$ such that $t^*_1 \leq t_0(Q)$. $J_1(t^*_Q)$ is decreasing in $Q$ up to $Q_0 + k$, and it is constant when $Q > Q_0 + k$. Therefore we have:
(1) \( J_1(t^*_0) > J_1(t^*_1) > \ldots > J_1(t^*_{Q_0+k}) = J_1(t^*_{Q_0+k+1}) = \ldots \)

(2) \( \left( J_1(t^*_0) - J_1(t^*_1) \right) \geq \left( J_1(t^*_1) - J_1(t^*_2) \right) \geq \ldots \)

\[ \geq \left( J_1(t^*_{Q_0+k}) - J_1(t^*_{Q_0+k+1}) \right) = \left( J_1(t^*_{Q_0+k+1}) - J_1(t^*_{Q_0+k+2}) \right) = \ldots \]

Note that buffer \( Q_1 \) is better than buffer \( Q_2 \) if and only if

\[ J_1(t^*_0) + C_{Q_2} < J_1(t^*_0) + C_{Q_1} \]

The above inequality can be written as:

\[ J_1(t^*_0) - J_1(t^*_0) + C_{Q_2 - Q_1} > 0 \]  \hspace{1cm} (23)

First it will be shown that if buffer \( Q_1 \) is better than \( Q_{1+1} \)
then it must be better than all buffers \( Q_2 \) with \( Q_2 \geq Q_{1+1} \). The
proof is by induction. Assume that (23) is true for \( Q_2 \), we will show
that (23) is true for \( Q_{2+1} \). Using the induction hypothesis we will
prove that:

\[ J_1(t^*_0) - J_1(t^*_0) + C_{Q_2 - Q_1} + C_q > 0 \]  \hspace{1cm} (24)

Since it is true for \( Q_{1+1} \), then \( C_q > J_1(t^*_0) - J_1(t^*_0) \)

Also we have:

\[ J_1(t^*_0) - J_1(t^*_0) = J_1(t^*_0) - J_1(t^*_0), \text{ for } Q_1 < Q_2 \]

Therefore \( C_q > J_1(t^*_0) - J_1(t^*_0) \). Substituting

\[ J_1(t^*_0) < (J_1(t^*_0) + C_q) \] in (18), we have

\[ 0 < J_1(t^*_0) - J_1(t^*_0) + C_{Q_2 - Q_1} < \]

\[ J_1(t^*_0) - J_1(t^*_0) + C_{Q_2 - Q_1} + C_q \]
Now suppose that buffer $Q_1$ is better than buffer $Q_{1-1}$, we will show that buffer $Q_1$ is better than buffer $Q_2$ for all $Q_2 \leq Q_{1-1}$. The proof is also by induction. We will prove that:

$$J_1(t_{02-1}^*) - J_1(t_{01}^*) + C_q(Q_2 - Q_1) - C_q > 0$$

(25)

Since (25) is true for $Q_{1-1}$, then $C_q < J_1(t_{01-1}^*) - J_1(t_{01}^*)$.

Also we have:

$$J_1(t_{01-1}^*) - J_1(t_{01}^*) \leq J_1(t_{02-1}^*) - J_1(t_{02}^*), \quad Q_1 > Q_2.$$  

Therefore, $J_1(t_{02-1}^*) - J_1(t_{02}^*) > C_q$.

By induction, $J_1(t_{02}^*) - J_1(t_{01}^*) + C_q(Q_2 - Q_1) > 0$  

(26)

Substituting $J_1(t_{02}^*) < (J_1(t_{02-1}^*) - C_q)$ in (26), we have:

$$0 < J_1(t_{02}^*) - J_1(t_{01}^*) + C_q(Q_2 - Q_1) <$$

$$J_1(t_{02-1}^*) - J_1(t_{01}^*) + C_q(Q_2 - Q_1) - C_q.$$  

Q.E.D.

The joint optimal $(Q^*, t_{0}^*)$ can be determined easily using the unimodality property. Table 3.3 and Figure 3.7 illustrate the above theorem.
The failure time distribution is Erlang-2. $\beta = 0.05$, $t_1^* = 7.35$.

$\tau_m / \tau_r = 0.1$, $\theta = 0.1$, $u = 0.1$, and $\lambda E[S] = 0.9$.

<table>
<thead>
<tr>
<th>Q</th>
<th>$t_q^*$</th>
<th>$J_1(t_q^*)$</th>
<th>$J_1(t_q^*) + Q$</th>
<th>$J_1(t_q^*) + 3Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$8 = Q_o$</td>
<td>1.50</td>
<td>122.64</td>
<td>130.64</td>
<td>146.64</td>
</tr>
<tr>
<td>10</td>
<td>1.78</td>
<td>103.37</td>
<td>113.37</td>
<td>133.37*</td>
</tr>
<tr>
<td>15</td>
<td>2.28</td>
<td>85.92</td>
<td>100.92</td>
<td>130.92*</td>
</tr>
<tr>
<td>20</td>
<td>2.75</td>
<td>71.23</td>
<td>91.23</td>
<td>131.23</td>
</tr>
<tr>
<td>25</td>
<td>3.09</td>
<td>65.21</td>
<td>90.21</td>
<td>140.21</td>
</tr>
<tr>
<td>30</td>
<td>3.45</td>
<td>60.25</td>
<td>90.25</td>
<td>150.25</td>
</tr>
<tr>
<td>35</td>
<td>3.78</td>
<td>56.63</td>
<td>91.63</td>
<td>161.63</td>
</tr>
<tr>
<td>40</td>
<td>4.10</td>
<td>53.74</td>
<td>93.74</td>
<td>173.74</td>
</tr>
<tr>
<td>45</td>
<td>4.40</td>
<td>51.47</td>
<td>96.47</td>
<td>186.47</td>
</tr>
<tr>
<td>50</td>
<td>4.70</td>
<td>49.54</td>
<td>99.54</td>
<td>199.54</td>
</tr>
</tbody>
</table>

* denote optimal

Table 3.3 Optimal number of defective items
Figure 3.7: Total cost of [SUB2] vs. # of def.
Suppose the costs $C_t$ and $C_q$ are positive, the following cases are investigated:

If $t^*_o \in [t_L(L^*), t_r(L^*)]$ then the optimal waiting capacity and the optimal allowed number of defective items are reached as follows:

(i) $t^*_L = t^*_o$ then $(L^*, Q^*, t^*_L)$ is optimal, or

(ii) $t^*_L \neq t^*_o$ then

$Q_o = \text{the smallest integer } \geq d(t^*_L)$

$L_o = \text{the smallest integer } \geq -\lambda E[W(t^*_o)] \text{ Log(})$

if $J_1(t^*_L) + C_q L^* + C_q Q_o \geq J_1(t^*_o) + C_q L_o + C_q Q^*$

then $(L^*_0, Q^*_0, t^*_o)$ is optimal

else $(L^*, Q^*, t^*_L)$ is optimal

If $t^*_o \notin [t_L(L^*), t_r(L^*)]$ then the couple $(L^*, Q^*)$ is infeasible.

Furthermore a trade-off must be considered. The following algorithm is used to determine the optimal policy.

3.4.4 Derivation of algorithm

Let $L_{\text{min}} = L^*$, $L_{\text{max}}$ be the smallest integer greater than or equal to $-\lambda E[W(t^*_o)] \text{ Log(})$, and $Q_{\text{min}} = Q^*$. 
Let $Q_i$ be the smallest integer greater than or equal to $d(t_i^*)$, such that:

$$
\begin{align*}
\text{if } t_{L+1}^* & = \\
& \begin{cases} 
  t_{\ell}(L+1) & \text{then } t_i^* = t_{\ell}(L+1) \\
  t_{r}(L+1) & \text{then } t_{i}(L+1) \leq t_i^* \leq t_{r}(L+1)
\end{cases}
\end{align*}
$$

Where $t_{L+1}^*$ is the optimal maintenance for $L = L+1$.

For each value of $L+i$, we have many feasible number of defective items. Based on the unimadility property, we can determine easily the best feasible policy $(Q_i, t_i^*)$ in the interval $[t_{\ell}(L+1), t_{r}(L+1)]$. This can reduce dramatically the number of candidates.

We construct a sequence $S$ such that the optimal policy $(L^*, Q^*)$ belongs to $S$. This is shown in the following:

![Figure 3.8: Sequenced couples of S](image)

For each couple $(L+i, Q_i)$, the optimal age-maintenance is:

$$
t_{\ell}(L+i) \leq t_i^* \leq t_{r}(L+i)
$$

The optimal solution $(L^*, Q^*, t_0^*)$ represents the couple which has the minimum cost in $S$. 
3.4.5 Illustrative example

To illustrate the algorithm in section 3.4.4, we consider a system with Erlang-2 failure time distribution. The following parameters are used: $\beta = 0.05$, $\theta = 0.1$, $u = 0.1$, $C_b = $1000, $C_\xi = $1, $C_q = $1, $C_w = $1, $C_m = $50, and $\lambda E[S] = 0.9$.

From Tables 3.2 and 3.3, the minimum buffer and allowed number of defective items are: $L_{min} = 32$ and $Q_{min} = 27$, respectively. The following table contains the candidate groups and their associated costs:

<table>
<thead>
<tr>
<th>$L + i$</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_i$</td>
<td>59</td>
<td>35</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>$t_{i}^*$</td>
<td>5.50</td>
<td>3.78</td>
<td>3.40</td>
<td>3.24</td>
</tr>
<tr>
<td>COST</td>
<td>133.75</td>
<td>124.63</td>
<td>124.03</td>
<td>125.25</td>
</tr>
</tbody>
</table>

Table 3.4: Candidate groups for optimality

As shown in Table 3.4, the unique solution $(L^\ast, Q^\ast, t^\ast_0)$ is $(34, 28, 3.40)$. 
3.5 CONCLUSION

This chapter examined an M/M/1 queueing system that is subject to failure, degradation in quality of the output, and the system has buffer cost. We formulated a nonlinear integer programming model to simultaneously determine the optimal waiting space and the optimal number of defective items per cycle. Based on Theorems 3.1 to 3.3, we developed an algorithm to determine the optimal policy \((L^*, Q^*, t_0^*)\). In general, when the maintenance and repair times are not negligible the optimal waiting capacity \(L^*\) is unique. The age-maintenance policy which decreases the expected number of defective items per cycle, increases the expected number of items waiting in the system and vice versa. Thus a trade-off between buffer and quality costs is considered. Once the minimum numbers of waiting space and defective items are found, the proposed algorithm determined easily the optimal policy.

For possible extension, one may consider the case where the defective parts may return to the queue for rework. In this case, we have a queueing system with Bernouilli feedback. The problem is an open area for research because the feedback process in steady state may not be Poisson distributed and therefore the diffusion approximation method presented in Chapter 2 cannot be applied.
CHAPTER FOUR:

An Age-Maintenance Policy for Open Queueing Networks
Notation

$A_i(t_i)$ the fraction of time node $i$ is operating under policy $t_i$

$\beta_i$ the arrival rate of a breakdown after the last repair or PM at node $i$

$C_b$ the cost of each breakdown

$C_N$ the cost of each person per unit time

$C_m$ the cost of each maintenance

$C_w$ the cost of waiting per unit time

$f_i(t)$ the probability density function that a breakdown occurs at time $t$ since the last repair or maintenance at node $i$

$F_i(t_i)$ the probability that node $i$ fails before $t_i$

$\bar{F}_i(t_i)$ the probability that node $i$ is switched off before failure

$r_i(t)$ Failure rate of node $i$, i.e., $r_i(t) = f_i(t)/F_i(t)$

$E[.]$ the expectation

$V[.]$ the variance

$\tau_{bi}$ the mean time that node $i$ is operating without failure

$\tau_{mi}$ the average time to perform a maintenance at node $i$

$\tau_{ri}$ the average repair time at node $i$

$\lambda_i$ the net job's arrival rate of node $i$

$E[W_i(t_i)]$ the expected waiting time in the system at node $i$, under $t_i$

$\varphi_i(t_i)$ the average cycle time at node $i$ under policy $t_i$

$\mathcal{F}_i(t_i)$ the average off-time period at node $i$ under policy $t_i$

$$\mathcal{F}_i(t_i) = 1 - A_i(t_i)$$

$p_{ki}$ the probability that jobs are routed from node $k$ to $i$
4.1 BACKGROUND

Queueing networks have been used to model the performance of a variety of complex systems such as production job shops, flexible manufacturing system (FMS), computers and communication networks, and complex information processing systems. For example, Solberg (1977) showed that the performance of an FMS can be effectively predicted by modeling the system as a network of queues. Exact and approximate results exist only for limited classes of product-form type networks. Moreover, the performance of network of servers tend to be distorted by irregularities caused by machine breakdowns which affect the productive capacity of the system. The presence of unreliable machines prompts the need for performance models that capture the reliability of the system. The lack of success in attempting to solve general queueing network models have led to the extensive use of simulation models. However, the high cost involved in simulation studies, have brought about increasing interest in the use of approximation approaches.

Altiok et al. (1983) analyzed a flow-shop-type production line where the stations are subject to breakdowns. They modeled the system as a series of queues. The failure time distribution is assumed to be exponential. Vinod and Altiok (1986) presented an exact equivalent network with two-stage coxian service time distribution for unreliable closed queueing network, where the job interarrival, failure, repair and service times have exponential
distributions. The service completion times of the equivalent reliable network are approximated to be exponentially distributed. Ramanjaneyulu et al. (1989) presented a new approximation for unreliable closed queueing networks. The failure time distribution is assumed to be exponential. They transformed the unreliable closed queueing network into a closed queue with all reliable servers by creating virtual nodes. All these approximation models assume exponential failure time distribution. Also, the optimization side for unreliable queueing networks has not been explored.

4.2 PROBLEM DEFINITION

We consider an open queueing network consisting of $m$ single-server stations with infinite waiting capacity. Customers/jobs arrive from an external source to node $j$ according to a Poisson process with rate $\lambda_{0j}$. The jobs are served by an exponential server at node $j$ with rate $E[S_j]$. After completing service at node $j$, a customer either goes to node $k$ instantaneously with probability $p_{jk}$, or leaves the system with probability $\left(1 - \sum_{k=1}^{m} p_{jk}\right)$. The routing probabilities are assumed to be independent of the state of the system. To model the breakdown/PM problem, we consider the single-server at each station to be prone to failure. All stations are assumed to be stochastically independent of each other, i.e., when station $i$ fails, it will not affect other stations. The
age-maintenance policy is the vector \( t_0 = (t_1, \ldots, t_n) \), where \( t_i \) is the decision variable of the age-maintenance policy \((1, t_i)\) for node \( i \).

The network consists of a large number of different machines with individual demands for maintenance. These demands must be met by a maintenance crew of limited size. A capacity problem arises when more machines break down simultaneously than the available number of maintenance men. The waiting time for repair assistance is unproductive time which can be reduced by increasing the number of maintenance men. The optimal number of maintenance men can be found by minimizing the sum of the machine downtime cost and the wages of the maintenance men. We assume that a maintenance man cannot performed maintenance/repair at two or more machines simultaneously.

Let \( N \) denote the maintenance crew size. Since the times for maintenance and repair are not negligible, the optimal value of \( N \) must be determined. The purpose of this chapter is to determine the joint optimal \((N^*, t_0^*)\) which minimizes the total expect cost per unit time over an infinite planning horizon.

4.3 MODEL FORMULATION

In this chapter, we assume that the connection between node \( i \) and \( j \) \((p_{ij})\) is always reliable, i.e. the path between \( i \) and \( j \) never breaks down. Suppose the mean output from station \( k \) is \( \lambda_k \). When the queueing system at each station is stable then the
mean input is:

\[ \lambda_i = \lambda_{0i} + \sum_{k=1}^{m} p_{ki} \lambda_k \quad i = 1, \ldots, m \]  

(1)

The stability constraint at each station is:

\[ A_i(t) \leq \lambda_i E[S_i] \quad i = 1, \ldots, m \]  

(2)

where \( A_i(t) \) is the fraction of time station \( i \) is on under policy \( t \).

The capacity problem discussed in the previous section can occur at any time in the system. Thus, we introduce the resource constraint as follows:

\[ P[\xi(t) \leq N] \geq 1 - \alpha \]  

(3)

where \( \xi(t) \) is the number of servers under maintenance or repair at time \( t \), and \( \alpha \) is the tolerance factor of feasibility \((0 \leq \alpha \leq 1)\). We assume that all maintenance men have the same skill. In other words, they perform the same tasks with the same costs. Also we assume that there are no spare servers in inventory.

Since the maintenance crew is less than the number of stations, we consider the following probability:
\[
\lim_{t \to \infty} P[\xi(t) = k] = p_k.
\]

where 
\[
p_k = \sum_{j_1 = 1}^{n_k} \left( \prod_{l=1}^{k_1} \left( \frac{f_{j_1} f_{j_2} \cdots f_{j_{k_1}}}{(1-f_{j_1})} \right) \right)
\]

In Equation (4), \( n_k = \frac{m!}{k!(m-k)!} \), and \( f_i(t) \) is the fraction of time server \( i \) is off under \( t_1 \), such that \( f_i(t_1) = 1 - A_1(t_1) \). The mean \( \mu \) and the standard deviation \( \sigma \) of \( \xi(t) \) can be determined as follows:

\[
\mu = \sum_{k=1}^{m} kp_k = \sum_{i=1}^{m} f_i(t_1)
\]

\[
= \sum_{i=1}^{m} \left( 1 - A_1(t_1) \right),
\]

\[
\sigma^2 = \sum_{k=1}^{m} k^2 p_k - \mu^2 = \sum_{i=1}^{m} f_i(t_1)(1-f_i(t_1)),
\]

or \( \sigma = \sqrt{\sum_{i=1}^{m} A_1(t_1)(1-A_1(t_1))} \). (6)

Since the traffic intensity at each station is high, and the fraction of time the server is on at each node is greater or equal to the traffic intensity, then the difference between the values of \( f_i(t_1) \)'s is not significant. The system can be considered as \( m \) independent repetitions of the simple success-failure experiment.
At each repetition $i$, the probability of success is $\xi_i(t_i)$. Thus $\xi(t)$ is approximated as a binomial random variable that can have values $0, 1, 2, \ldots, m$, with mean $\mu$ and standard deviation $\sigma$.

Assume that the number of stations in the network is sufficiently large, so that the distribution of the random variable $\zeta(t)$, i.e., $\zeta(t) = [\xi(t) - \mu] / \sigma$ can be approximated as normal. The cumulative distribution of $\zeta(t)$ is given by:

$$P[\zeta(t) \leq (N - \mu) / \sigma] = \Phi((N - \mu) / \sigma)$$

(7)

where $\Phi(.)$ is the standard normal distribution. Therefore, constraint (4) becomes:

$$m - \sum_{i=1}^{m} A_i(t_i) + Z \left( \sum_{i=1}^{m} A_i(t_i)[1-A_i(t_i)] \right) \leq N$$

(8)

where $Z = \Phi^{-1}(1-\alpha)$.

Let $C_m$, $C_r$, $C_c$, and $C_w$ be the cost of each maintenance, the cost of each repair, the cost of each maintenance crew per unit time, and the cost of waiting per unit time, respectively. The following multidimensional $(N \times R^m \to R)$ nonlinear problem is formulated to determine the joint optimal $(N^*, t_0^*)$: 
\[ \text{[INLP2]: MINIMIZE} \]

\[
J(t_0, N) = \sum_{i=1}^{m} C_w \lambda_i E[W_i(t_i)] + \sum_{i=1}^{m} \left( \frac{C_m F_i(t_i) + C_b F_i(t_i)}{\varphi_i(t_i)} \right) + C_N \eta \quad (9)
\]

s.t.
\[
\lambda_i E[S_i] - A_i(t_i) \leq 0 \quad i = 1, \ldots, m \quad (10)
\]

\[
m - \sum_{i=1}^{m} A_i(t_i) + \frac{\sum_{i=1}^{m} \left( A_i(t_i) [1 - A_i(t_i)] \right)}{\sum_{i=1}^{m} \left( A_i(t_i) \right)} \leq N \quad (11)
\]

\[
N = 1, 2, 3, \ldots \quad (12)
\]

Where \( \varphi_i(t_i) = \tau_{s_i} + (\tau_{r_i} - \tau_{s_i}) F_i(t_i) + \int_{0}^{t_i} F_i(t) dt \), and \( F_i(t) \), \( \tau_{s_i} \) and \( \tau_{r_i} \) are the cumulative failure time distribution of server \( i \), the mean maintenance and repair times of server \( i \), respectively.

Equation (9) represents the expected total cost which includes the cost of waiting per unit time, the cost of breakdown per unit time, and the cost of maintenance per unit time for each station, and the cost of maintenance men per unit time. Constraint (10) guarantees the stability condition at each station. This means that the mean output at each station must be equal to the mean input. Constraint (11) ensures that the probability that the number of servers under maintenance or repair at any time being less or equal to the available number of maintenance men, must be greater or equal to \( 1 - \alpha \).
4.4 FEASIBLE CREW SIZE

Let $t^*_{A_1}$ be the value of $t_1$ which maximizes $A_i(t_1)$, and

$$\Gamma(t_1, \ldots, t_m) = m - \sum_{i=1}^{m} A_i(t_1) + Z \sqrt{\sum_{i=1}^{m} \left( A_i(t_1)[1-A_i(t_1)] \right)}.$$ 

Lemma 4.1:

$t^*_{A}$ is the unique value of $t_0$ that minimizes $\Gamma(t_0)$, where $t^*_{A} = (t^*_{A_1}, \ldots, t^*_{A_m})$ and $t_0 = (t_1, \ldots, t_m)$.

**Proof:**

Differentiating $\Gamma(t_0)$ and equating it to zero

$$A'_i(t_1) \left( -1 + Z \frac{1-2A_i(t_1)}{2 \sum_{i=1}^{m} A_i(t_1)[1-A_i(t_1)]} \right) = 0, \text{ for } i = 1, \ldots, m.$$ 

Since $A_i(t_1) \geq \lambda_i E[S_1]$ and $\lambda_i E[S_1]$ is high for $i = 1, \ldots, m$, this implies that

$$\left( -1 + Z \frac{1-2A_i(t_1)}{2 \sum_{i=1}^{m} A_i(t_1)[1-A_i(t_1)]} \right) = 0.$$ 

Thus $\forall \Gamma(t_0) = 0$ if and only if $A'_i(t_1) = 0$ for $i=1, \ldots, m$. We know $A'_i(t_1)$ has a unique maximum (see Chapter 2), so $t^*_{A_1}$ is the unique value of $t_1$ that maximizes $A_i(t_1)$, i.e., $A'_i(t^*_{A_1}) = 0$. 

Since $A'(t_{A1}^*) = 0$, we have

$$\Gamma''(t_{A1}^*) = A''(t_{A1}^*) \left[ -1 + Z \frac{1 - 2A_1(t_{A1}^*)}{2 \sum_{i=1}^{m} \left[ A_i(t_{A1}^*) (1 - A_i(t_{A1}^*)) \right]} \right] \geq 0$$

$\forall \Gamma(t_{A1}^*) = 0$ and $\Gamma''(t_{A1}^*) \geq 0$. Consequently, $t_{A1}^*$ is the unique value of $t_0$ that minimizes $\Gamma(t_0)$.

Q.E.D.

From Lemma 4.1, we conclude that when $N < \Gamma(t_{A1}^*, \ldots, t_{Am}^*)$, then $N$ is infeasible. Let $N_0$ be the lowest feasible number of maintenance men. Therefore, $N_0$ is the smallest integer greater or equal to the following:

$$\Gamma(t_{A1}^*) = m - \sum_{i=1}^{m} A_i(t_{A1}^*) + \phi^{-1}(1-\alpha) \sum_{i=1}^{m} \left[ A_i(t_{A1}^*) (1 - A_i(t_{A1}^*)) \right]$$

(13)

The maintenance crew size $N$ is feasible if and only if:

$$N \geq N_0$$

(14)

4.5 INTRACTABILITY OF QUEUEING PERFORMANCE MEASURES

To find the solution to INLP2, we first examine the various methods for computing the expected waiting time.
4.5.1 Equivalent reliable network of queues

The implementation of the age-maintenance policy in queueing systems with interruptions (e.g., breakdowns) causes the on-time period to have a truncated distribution (see Chapter 2). Let \( h_a(.) \) and \( h_r(.) \) be the density functions of the maintenance and repair times, respectively. Thus \( h(t) = \bar{F}(t)h_m(t) + F(t)h_r(t) \). Even if \( h_m \) and \( h_r \) are exponential, \( h(t) \) is hyper-exponential. Consequently, the off-time period is not exponentially distributed.

Altiok et al. (1983), Vinod and Altiok (1986), and Ramanjaneyulu et al. (1989) presented an equivalent reliable closed queueing network to estimate the performance measures of closed unreliable queueing networks. They assume that the job interarrival, failure, service and repair times are exponentially distributed. In this situation, the model with service breakdown can be mapped into an equivalent model of server without breakdown, using an effective service time distribution that accounts for both the actual service time and for delays due to breakdown. Unfortunately, the server interruptions that occur in our model, are much more complicated. Since their assumptions do not hold, their methods cannot be applied.

4.5.2 Priority queues method

Our model can be treated as a priority queues problem with two classes of customers and operating under the head-of-line and preemptive-resume discipline. The first class customer represents
PM/breakdown and the second class represents jobs. The study of priority queues has become an area of great interest in the field of Operations Research. Priority queues have appeared in the literature since Cobham article in 1954. Since that time there has been a large amount of work on the subject. However, most of the research assumes Poisson arrivals. Unfortunately, the arrival processes of the higher classe in our model are not Poisson.

4.5.3 Diffusion approximation method

The output process at each node is a sequence of times at which jobs leave the server. The output of one node can be the input to another node. It is proved that, for a large class of stable stationary queueing systems with renewal arrival process and without losses, a necessary condition for the departure process also to be a renewal process is that its interval distribution (output process) be the same as that of the arrival process [Berman and Wescott (1983)]. In our model the distribution of the output process cannot be the same as the distribution of the arrival process due to the presence of off-time periods. Figure 4.1 illustrates the departure process during an ON/OFF cycle.

![Image of the output process cycle](image)

**Figure 4.1:** The output process
Clearly, the output process at each node is not a renewal process. Consequently, the interarrival time at each node is not exponentially distributed.

Diffusion approximation can be used when the on- and off-times have general distribution for an M/G/1 system (see Chapter 2). In our model, the arrival process at each node is not Poisson. Therefore, the method cannot be applied even if the heavy traffic condition is imposed.

4.5.4 Equivalent restricted network of queues

Open restricted queueing networks have been studied by Hillier and Boling (1967), Labetouille and Pujolle (1980), Takahashi et al. (1980), Kerbache and Smith (1987). An open restricted queueing network is a network of queues where the waiting capacities are limited. Blocking may occur due to restrictions on the queue length. Blocking is the situation where a customer at station i cannot complete its service until the succeeding station j is released. The probability of blocking is equal to the probability that station j is filled with customers. If an unreliable open queueing network can be transformed to a restricted queueing network, then we can apply methods in the literature on restricted queueing networks to our problem. To transform the unreliable network to a restricted open queueing network, we create at each node i a virtual destination node $V_i$. The external arrivals to $V_i$ are breakdowns/maintenance. Furthermore $V_i$ can block i whenever a failure or maintenance occurs. Figure 4.2 illustrates this approach.
Figure 4.2: Virtual node process
The model can be transformed to a restricted queueing network with the following properties:

1-- The waiting capacity at $V_i$ is zero, i.e., the system at $V_i$ can have only one customer.

2-- The mean service time of jobs at $V_i$ is equal to 0 so that the expected number of jobs at $V_i$ is always 0, and let the mean off-time be $\tau_i$.

3-- Let event 1 and event 2 be a service completion and a PM/breakdown, respectively. If event 1 occurs just before event 2 then the job proceeds to $V_i$. Otherwise the processing job blocks the server until $V_i$ is released. When both events occur at the same time, the finished job will be lost.

When $V_i$ is "busy" with a breakdown or a maintenance, (i.e., $V_i$ is under maintenance or repair), it blocks i. The probability of blocking is equal to the probability that $V_i$ is busy with a breakdown or a maintenance. When $V_i$ becomes idle, server i resumes service until completion. However, when a job leaves station i, and $V_i$ becomes busy with a PM/breakdown, the job will be lost. Let $\ell_i$ be the probability that a customer leaving i towards $V_i$ will be lost, knowing that $V_i$ is full at the beginning of its service. For unreliable queueing networks, losses of customers in the system do not exist. Thus finding the coefficient of $\ell_i$ is quite important for the transformation process. We suggested the probability of
blocking at node $i$ is:  
\[ b_i = \mathcal{F}_i(t_i) - \ell_i \]  
(15)

Labetoulle and Pujolle (1980) provided an approximated closed-form for the value of $\ell_i$ when all parameters are exponentially distributed. Unfortunately, the PM/breakdown process is not Poisson and the off-time is not exponentially distributed. Therefore, it is not possible to compute the values of $\ell_i$.

4.5.5 Simulation

The literature on unreliable queueing networks do not address problems with general failure time distribution. Consequently, simulation can be used to provide some estimates to our problem. Discrete-event Monte Carlo simulation has been an obvious alternative for evaluating general queueing networks. The main drawback is the computational requirement. To optimize the age-maintenance policy, the finite difference technique can be used (i.e., changing each age $t_i$ and computing the sensitivity by doing a new simulation after each change). For $m$ stations, each iteration step requires $m$ new simulations. One simulation can take over 100 times the computer time that a queueing model would use [Suri (1984)]. Recently, gradient estimation techniques are developed to improve the efficiency of simulation and optimize the system in fewer runs. However, theoretical issues arise out of this methods. There are problems of unbiasedness and convergence.
4.5.6 Summary

Deriving analytical or closed-form expressions for the expected waiting time in the system at each station in a queueing network with non-exponential failure and repair times distributions do not appear possible.

In this model, each station in the network consists of an \( M/M/1 \) queue with interruptions. The off-time period at each station is not negligible. Thus, under any maintenance policy implemented for queueing networks, it is not possible to determine a closed-from expression of the waiting time at each node.

4.6 MAINTENANCE OPTIMIZATION MODEL

The expected waiting time in the system at each station appears in the objective function (Equation 9). Since there is no analytical or closed form expression for the expected waiting time in the system, INLP2 is intractable. However, we will attempt to solve INLP2 under two special cases.

4.6.1 Case 1: Cost of waiting is positive

Let \( \rho_i \) be the total traffic intensity (or utilization factor) at node \( i \). \( \rho_i \) is the fraction of time node \( i \) is busy (processing a job, or under maintenance/repair). The closer \( \rho_i \) approaches unity, the larger the waiting time at node \( i \). \( \rho_i \) reflects the way in which the system performance varies with the average system load. Suppose the system alternates between a busy period and an idle period.
Such a regenerative cycle is illustrated in Figure 4.3.

![Diagram of a regenerative cycle](image)

**Figure 4.3: A regenerative cycle**

Let $E[B]$ and $E[I]$ be the expected times of the busy and the idle periods, respectively. We claim that the following is true:

**Claim 1:**

The expected waiting time at node $i$ is strictly increasing with respect to $\rho_i$.

The proof is by contradiction. Suppose $\rho_{i1} < \rho_{i2}$. Then $E[W_{i1}] > E[W_{i2}]$. Figure 4.4 shows how the expected busy period varies when the traffic intensity is increased.
If the traffic intensity is increased, the arrival rate is increased. Since there is no idle period within the busy period, then the time of the busy period cannot decrease. Obviously, if $\rho_{11} < \rho_{12}$ then $E[B_1] < E[B_2]$. Consequently, the expected waiting time at each node is increasing with respect to its traffic intensity. Since $t_{A_1}^*$ is the value of $t_1$ which maximizes $A_1(t_1)$ (and minimizes $\rho_i$) then $t_{A_1}^*$ minimizes $E[W(t_1)]$. This result helps us to solve the first special case, where

$$C_{r_{bi}} = C_{r_{ai}} \text{ for } i = 1, \ldots, m.$$
Theorem 2 in Chapter 2 proved that the optimal age-maintenance policy is \( (t^*_1, t^*_2, \ldots, t^*_m) \), where \( t^*_i \) is the value of \( t_i \) satisfies the following equation:

\[
r_i(t_i) \int_0^{t_i} F_i(t) \, dt - F_i(t_i) = \frac{\tau_{m_i}}{\tau_{r_i} - \tau_{m_i}} \quad \text{for } i = 1, \ldots, m.
\]

To determine the optimal crew size for each condition, we simply find the number of maintenance men which can perform the optimal age maintenance policy found above. Since \( (C, N) \) is increasing with respect to \( N \) and \( N_0 \) is the lower bound that satisfies Equation (13). Therefore, \( N_0 \) is the optimal crew size.
4.6.2 Case 2: No cost of waiting ($C_w = 0$)

Nonlinear programming methods are NP hard. When the number of servers in the network is large, the solution for INLP2 is intractable. In order to make the problem tractable, we use a linear approximation for the standard deviation. Since the traffic intensity at each station is high then the interval $(\lambda_i E[S_i], 1)$ must be small. We know that any function in a very small interval can be linearized.

4.6.2.1 Linear approximation to the standard deviation

Our aim is to transform the standard deviation to a linear function with respect to $A_i(t_i)$ in the small intervals $[\lambda_i E[S_i], 1]$, for $i = 1, \ldots, m$.

$$
\sigma = \sqrt{\sum_{i=1}^{m} \left[ A_i(t_i)[1-A_i(t_i)] \right]} \approx \sum_{i=1}^{m} a_i A_i(t_i) + a_0
$$

If $A_i(t_i) = 1$, for $i = 1, \ldots, m$ then $\sigma = 0$. This implies that

$$
a_0 = - \sum_{i=1}^{m} a_i.
$$

If $A_i(t_i) = \lambda_i E[S_i]$, for $i = 1, \ldots, m$, then

$$
\sigma = \sqrt{\sum_{i=1}^{m} \left( \lambda_i E[S_i](1-\lambda_i E[S_i]) \right)}.
$$

This implies that

$$
\sum_{i=1}^{m} a_i (\lambda_i E[S_i] - 1) = \sqrt{\sum_{i=1}^{m} \left( \lambda_i E[S_i](1-\lambda_i E[S_i]) \right)}.
$$

(16)
There are many roots to Equation (16). However, one possible solution to (16) is the following:

\[ a_i = \sqrt{\frac{\sum_{i=1}^{m} \left( \lambda_i E[S_i]\right) \left( 1 - \lambda_i E[S_i] \right)}{m \left( \lambda_i E[S_i] - 1 \right)}} \]

and

\[ a_0 = -\sqrt{\frac{\sum_{i=1}^{m} \left( \lambda_i E[S_i]\right) \left( 1 - \lambda_i E[S_i] \right)}{m \left( \lambda_i E[S_i] - 1 \right)}} \sum_{i=1}^{m} \frac{1}{m \left( \lambda_i E[S_i] - 1 \right)} \]

To compare the values of the standard deviation with the linear approximation, we consider the following notations:

Let \( I_i = A_i - \lambda_i E[S_i] \), \( \lambda_i E[S_i] \geq 0.9 \), and \( I = \sum_{i=1}^{m} I_i/m \).

Figure 4.5 shows that the approximation is always underestimated.

For small values of \( I \), the approximation is very accurate.

Replacing \( \sigma \) in constraint (11) by its approximation, we have:

\[ \sum_{i=1}^{m} \mu_i A_i(t_i) + u_0 - N \leq 0. \] \hspace{1cm} (17)

where

\[ u_i = a_i Z - 1, \]

and

\[ u_0 = m + a_0 Z. \]

Since \( \lambda_i E[S_i] \leq 1 \) for \( i = 1, \ldots, m \), then \( u_i \leq 0 \) for all \( i \) and \( u_0 > 0 \).
Figure 4.5: Comparison of the standard deviation with linear approximation
4.6.2.2 Optimal age-maintenance for a fixed $N$

The necessary conditions for optimizing INLP2 problem are known as the classical Kuhn-Tucker conditions. Many authors have derived the necessary and sufficient conditions for various cases of nonlinear programming problems utilizing the Kuhn-Tucker conditions.

Given a fixed crew size $N$ and $C_w = 0$, Equations (9) to (11) can be transformed as follows:

\[
[NLP] \quad \text{Minimize } J(t_0) = \sum_{i=1}^{m} \left( \frac{C_{m+1} \tilde{F}_i(t_0) + CF_i(t_0)}{b_1(t_1)} \right) \quad (18)
\]

\[\text{s.t.} \quad g_1(t_0) = \lambda_i E[S_i] - A_i(t_1) \leq 0 \quad i = 1, \ldots, m \quad (19)\]

\[g_{m+1}(t_0) = \sum_{i=1}^{m} u_i A_i(t_1) + u_0 - N \leq 0 \quad (20)\]

$t_0 \in E^n$ and we assume that $F_i(.)$ is twice continuously differentiable for $i = 1, \ldots, m$. Thus $J(t_0)$ and $g_1, \ldots, g_{m+1}$ are twice continuously differentiable. The above NLP can be solved by methods which are based on the transformation of the constrained problem into a sequence of unconstrained problems. There are two classes of methods, namely, the penalty and the Lagrangian methods. The penalty method have been studied extensively and used to solve many practical problems. However, the method suffers from numerical instability. The Lagrange multiplier methods have been widely used for the analysis of economic systems. The above NLP is transformed into the following Lagrangian function (L):
Minimize $L(t_0^*, \gamma) = J(t_0) + \sum_{i=1}^{m+1} \gamma_i g_i(t_0)$ \hspace{1cm} (21)

Where $\gamma_i$, $i = 1, \ldots, m+1$, are the Lagrange multipliers, and $L(t_0^*, \gamma)$ is the multiplier function, and the computational algorithm using the function is called the multiplier method. $L(t_0^*, \gamma)$ is constructed in such a way that it is twice continuously differentiable if $J(t_0)$ and $g_i(t_0)$, $i = 1, \ldots, m+1$, are twice continuously differentiable. This property is quite important to the computational procedure for finding the unconstrained minimum of the Lagrangian. At the k-th iteration, if the following conditions are satisfied:

(i) $\nabla J(t_0^k) + \sum_{i=1}^{m+1} \gamma_i^k \nabla g_i(t_0^k) = 0$

(ii) $\gamma_i^k g_i(t_0^k) = 0$ for $i = 1, \ldots, m+1$

(iii) $\gamma_i^k \geq 0$ and $g_i(t_0^k) \leq 0$ for $i = 1, \ldots, m+1$

then a Kuhn-Tucker point is reached, i.e., $\gamma^* = \gamma^k$ and $t_0^* = t_0^k$.

For $N = N_0$, the only feasible age-maintenance which satisfies constraint (11) is $t_0^* = (t_{A1}^*, t_{A2}^*, \ldots, t_{A_m}^*)$. Therefore, when the number of maintenance men is equal to $N_0$, the availability at each station must be maximized.
For $N > N_0$, we propose an iterative procedure to determine the optimal age-maintenance at each station. Let $t_{L_1}$ be the lower bound of $t_1$, such that $A_1(t_{L_1}) = \lambda_1 E[S_1]$. If $t_1 > t_{L_1}$ then $g_1(t_0) < 0$ for $i = 1, \ldots, m$. At the $k$-th iteration we have:

(i) if $t_1^k > t_{L_1}$ for $i = 1, \ldots, m$ then $\gamma^k = (0, 0, \ldots, 0, \gamma_{m+1}^k)$, or

(ii) if $t_1^k \leq t_{L_1}$ then we force $t_1^e$ to be equal to $t_{L_1}$ and

$$\gamma_1^e = \frac{\partial J/\partial t_1 + \gamma_{m+1}^k u_1 \partial A_1/\partial t_1}{\partial A_1/\partial t_1} (t_{L_1})$$

Since $A_1(t_{L_1}) = \lambda_1 E[S_1]$, then we replace $A_1(t_1^e)$ by $\lambda_1 E[S_1]$ in Constraint (20). Thus station $i$ is removed completely from the NLP formulation (18) - (20).

Therefore the problem is reduced to one Lagrangian multiplier $\gamma_{m+1}^*$. Standard line search methods may be applied to determine $\gamma_{m+1}^*$. First we investigate the properties which are suitable for applying line search methods. Let $\gamma$ denote $\gamma_{m+1}$ and $\gamma^i$ be the value of $\gamma$ at the $i$-th iteration. Let $t_0^i$ be the value of $t_0$ such that

$$\nabla J(t_0^i) + \gamma^i \nabla g_{m+1}(t_0^i) = 0.$$ 

Theorem 4.1:

If $\gamma^i < \gamma^j$ then $g_{m+1}(t_0^i) > g_{m+1}(t_0^j)$.
Proof:

Let $t_0^i$ be the value of $t_0$ under $\gamma^i$ such that
\[ \forall J(t_0^i) + \gamma^i \nabla g_{m+1}(t_0^i) = 0. \]  \hspace{1cm} (22)

Let $t_0^j$ be the value of $t_0$ under $\gamma^j$ such that
\[ \forall J(t_0^j) + \gamma^j \nabla g_{m+1}(t_0^j) = 0, \]  \hspace{1cm} (23)

where $\forall J(t_0^i) = (J_1^i(t_1^i), \ldots, J_m^i(t_m^i))$ and
\[ \nabla g_{m+1}(t_0^i) = (\frac{\partial g_{m+1}}{\partial t_1^i}, \ldots, \frac{\partial g_{m+1}}{\partial t_m^i}) = (u_1^i A_1^i(t_1^i), \ldots, u_m^i A_m^i(t_m^i)) \]

Since the Lagrange multiplier $\gamma$ is non-negative, $J_i(t_i)$ has a unique minimum, $A_i(t_i)$ has a unique maximum, and $u_i$ is non-positive, then the solution of (22) is unique and for (23) as well. This implies

\[ J(t_0^i) + \gamma^i g_{m+1}(t_0^i) \leq J(t_0^i) + \gamma^j g_{m+1}(t_0^j) \text{ for all } t_0 \geq t_L^i, \]

and

\[ J(t_0^j) + \gamma^j g_{m+1}(t_0^j) \leq J(t_0^j) + \gamma^j g_{m+1}(t_0^j) \text{ for all } t_0 \geq t_L^j. \]

Rearranging the above:

\[ J(t_0^i) - J(t_0^j) \leq \gamma^i (g_{m+1}(t_0^i) - g_{m+1}(t_0^j)), \]  \hspace{1cm} (24)

and

\[ J(t_0^j) - J(t_0^i) \leq \gamma^j (g_{m+1}(t_0^j) - g_{m+1}(t_0^i)). \]  \hspace{1cm} (25)

Since $t_0^i \geq t_L^i$, we apply (24) with $t_0 = t_0^i$, we have:

\[ J(t_0^i) - J(t_0^i) \leq \gamma^i (g_{m+1}(t_0^i) - g_{m+1}(t_0^i)). \]

Since $t_0^j \geq t_L^j$, we apply (25) with $t_0 = t_0^j$, we have:

\[ J(t_0^j) - J(t_0^j) \leq \gamma^j (g_{m+1}(t_0^j) - g_{m+1}(t_0^j)). \]
Adding the preceding two inequalities together, we have:

$$0 \leq (\gamma^1 - \gamma^j)(g_{m+1}(t^1_0) - g_{m+1}(t^j_0)).$$

Therefore, $\gamma^1 < \gamma^j$ implies $g_{m+1}(t^1_0) \geq g_{m+1}(t^j_0)$.

Q.E.D.

The bisection method can be used for finding $\gamma^*$ for a given $N$. The method can be carried out starting with $\gamma^R$ large enough so that $g_{m+1}(t^R_0) \leq 0$ and with $\gamma^L$ small enough so that $g_{m+1}(t^L_0) \geq 0$.

**Theorem 4.2:**

$$\gamma^L < \gamma^* < \gamma^R$$

**Proof:**

If $\gamma^k = \gamma^* > 0$ then $g_{m+1}(t^k_0) = 0$. From Theorem 4.1, $g_{m+1}(t^1_0)$ is a non-increasing function with respect to $\gamma^1$. If $g_{m+1}(t^R_0) \leq 0$, and $g_{m+1}(t^L_0) = 0$, then $\gamma^* < \gamma^R$. If $g_{m+1}(t^L_0) \geq 0$ and $g_{m+1}(t^R_0) = 0$ then $\gamma^L < \gamma^*$. Consequently, $\gamma^L < \gamma^* < \gamma^R$.

Q.E.D.

At each iteration the bisection method guarantees to halve the interval $(\gamma^L, \gamma^R)$ in which $\gamma$ must lie. Since $g_{m+1}(t^k_0)$ is non-increasing with $\gamma^k$, and the interval of uncertainty $(\gamma^L, \gamma^R)$ is bounded, the bisection method converges. Thus the length of the
interval of uncertainty after k iterations is equal to 
\((\gamma_R - \gamma_L)/2^k\), so that the method converges to \(\gamma^*\) within any desired level of accuracy. If the length of the final interval of feasibility is fixed at 1, then at the worst case the number of iterations performed will be the smallest integer such that 
\(2^k \geq (\gamma_R - \gamma_L)/1\).

**Theorem 4.3:**

\(\gamma^*(N) \leq \gamma^*(N-1)\) for all \(N > N_0\), i.e., \(\gamma^*(N-1)\) is an upper bound of \(\gamma^*(N)\).

**Proof:**

Let \(\gamma_1 = \gamma^*(N) > 0\) and \(\gamma_2 = \gamma^*(N-1) > 0\). We have:

\[
g_{m+1}(t_0, N) = \sum_{i=1}^{m} u_i A_1(t_1) + u_0 - N,\]

then

\[
g_{m+1}(t_0, N) \leq g_{m+1}(t_0, N-1)\] for all \(N \geq N_0\). We have:

\[
0 = g_{m+1}(t_0^1, N) \leq g_{m+1}(t_0^1, N-1)
\]

From Theorem 4.1, \(g_{m+1}(t_0^k)\) is non-increasing with respect to \(\gamma^k\). We have:

\[
g_{m+1}(t_0^2, N-1) = 0 \quad \text{and} \quad g_{m+1}(t_0^1, N-1) \geq 0,
\]

therefore \(\gamma_1 < \gamma_2\).

Q.E.D.
4.6.2.3 Derivation of Algorithm 4.1

Given N, we develop Algorithm 4.1 which determines the optimal age-maintenance policy using the bisection method to find $\gamma^*(N)$. The steps of the algorithm are outlined below:

(i) If $N = N_0 + 1$ then determine $\gamma_L$ and $\gamma_R$.
   If $N > N_0 + 1$ then $\gamma_L = 0$ and $\gamma_R = \gamma^*(N-1)$.
   Set $k = 1$.

(ii) Set $\gamma^k = \left(\gamma_R + \gamma_L\right)/2$.
    Find $t_0^k$ such that $\nabla J(t_0^k) + \gamma^k \nabla g_{m+1}(t_0^k) = 0$.
    If $t_0^k \leq t_{L_1}$ then $t_0^* = t_{L_1}$,
    replace $A_i(t_0^*)$ by $\lambda E[S_i]$ in $g_{m+1}(t_0^k)$,
    remove station $i$ from the NLP.

(iii) If $|\gamma^k g_{m+1}(t_0^k)| \leq \varepsilon$ then go to (iv),
    else
    if $g_{m+1}(t_0^k) > 0$ then $\gamma_L = \gamma^k$,
    else $\gamma_R = \gamma^k$,
    set $k = k+1$, and go to (ii).

(iv) $\gamma^* = \gamma^k$ and $t_0^* = t_0^k$, and stop.

$\varepsilon$ is used as the value for the stopping rule.
4.6.2.4 The optimal maintenance crew size

To reduce the search for finding the optimal maintenance crew size \( N^* \), we find a very useful property as listed in the following theorem.

**Theorem 4.4:**

The expected cost function in Equation (9) with \( C_w = 0 \), evaluated at \( t_0^*(N) \) is unimodal with respect to \( N \).

**Proof:**

Given a feasible \( N \), the optimal expected cost under \( N \) is:

\[
J(t_0^*(N), N) = J(t_0^*(N)) + C_N N
\]

where

\[
J(t_0^*(N)) = \sum_{i=1}^{m} \left( \frac{C_m r_i F_i(t_1)}{\varphi_i(t_1)} \right)
\]

Let \( N_0 \) be the smallest feasible crew, we have:

\[
C_N N_0 < C_N (N_0 + 1) < \ldots < C_N (N_0 + k) < C_N (N_0 + k + 1) < \ldots
\]

Let \( (N_0 + k) \) be the value of \( N \) such that \( \gamma^* \) becomes zero. Thus, \( J(t_0^*(N)) \) is decreasing in \( N \) up to \( N_0 + k \), and it becomes constant in \( N > N_0 + k \). Therefore we have:

\[
(1) \quad J(t_0^*(N_0 + k)) = J(t_0^*(N_0 + k + 1)) = \ldots
\]
(2) \[ J(t_0^*(N_0)) - J(t_0^*(N_0 + 1)) \geq J(t_0^*(N_0 + 1)) - J(t_0^*(N_0 + 2)) \geq \ldots \]
\[ \geq J(t_0^*(N_0 + k)) - J(t_0^*(N_0 + k + 1)) = J(t_0^*(N_0 + k + 1)) - J(t_0^*(N_0 + k + 2)) \]
\[ = \ldots \]

Note that crew size $N_1$ is better than crew size $N_2$ if and only if
\[ J(t_0^*(N_2)) + CN_2 > J(t_0^*(N_1)) + CN_1. \]

The above inequality can be written as:
\[ J(t_0^*(N_2)) - J(t_0^*(N_1)) + C(N_2 - N_1) > 0. \tag{26} \]

First it will be shown that if crew size $N_1$ is better $N_1 + 1$ then it must be better than all $N_2 \geq N_1 + 1$. The proof is by induction. Assume that (26) is true for $N_2$, we will show that (26) is true for $N_2 + 1$. Using the induction hypothesis we will prove that:
\[ J(t_0^*(N_2 + 1)) - J(t_0^*(N_1)) + C(N_2 - N_1) + C > 0 \]

Since the induction is true for $N_1 + 1$, then we have the following:
\[ C_n > J(t_0^*(N_1)) - J(t_0^*(N_1 + 1)). \]

Also we have:
\[ J(t_0^*(N_1)) - J(t_0^*(N_1 + 1)) \geq J(t_0^*(N_2)) - J(t_0^*(N_2 + 1)), \text{ for } N_1 < N_2 \]

Therefore $C_n > J(t_0^*(N_2)) - J(t_0^*(N_2 + 1))$. Substituting
\[ J(t_0^*(N_2)) < (J(t_0^*(N_2 + 1)) + C_n) \]
in (26) we have:
\[ 0 < J(t_0^*(N_2)) - J(t_0^*(N_1)) + C(N_2 - N_1) < \]
\[ J(t_0^*(N_2 + 1)) - J(t_0^*(N_1)) + C(N_2 - N_1) + C_n \]
Now suppose that crew size $N_1$ is better than crew size $N_1 - 1$, the goal is to prove that crew size $N_1$ is better than crew size $N_2$ for all $N_2 \leq N_1 - 1$. The proof is by induction, we will prove that:

$$J(t_0^*(N_2 - 1)) - J(t_0^*(N_1)) + C_N(N_2 - N_1) - C_N > 0$$

Since the induction is true for $N_1 - 1$, then we have the following:

$$C_N < J(t_0^*(N_1 - 1)) - J(t_0^*(N_1)).$$

Also we have:

$$J(t_0^*(N_1 - 1)) - J(t_0^*(N_1)) \leq J(t_0^*(N_2 - 1)) - J(t_0^*(N_2)),$$

for $N_1 > N_2$.

Therefore we have $J(t_0^*(N_2 - 1)) - J(t_0^*(N_2)) > C_N$. By induction the following is true:

$$J(t_0^*(N_2)) - J(t_0^*(N_1)) + C_N(N_2 - N_1) > 0 \quad (27)$$

Substituting $J(t_0^*(N_2)) < \left( J(t_0^*(N_2 - 1)) - C_N \right)$ in (27), we have:

$$0 < J(t_0^*(N_2)) - J(t_0^*(N_1)) + C_N(N_2 - N_1) <$$

$$J(t_0^*(N_2 - 1)) - J(t_0^*(N_1)) + C_N(N_2 - N_1) - C_R$$

Q.E.D.

**Theorem 4.5:**

If $C_N > J(t_0^*(N_0)) - J(t_0^*(N_0 + 1))$ then $N^* = N_0$ and $t_0^*(N) = t_A^*$.

If $C_N = J(t_0^*(N-1)) - J(t_0^*(N))$ then $N^* = N$ and $N^* = N - 1$.

**Proof:**

From Theorem 4.4, the total cost is unimodal with respect to $N$.

(i) The first relation means that the total cost at $N_0$ is better than the cost at $N_0 + 1$. Since $N_0$ is the lowest crew size, $N^* = N_0$. 

From Theorem 4.4, we have the following relations:

\[ N^* = \begin{cases} N-1, & \text{if } J(t_0^*(N-2)) - J(t_0^*(N-1)) > C_N > J(t_0^*(N-1)) - J(t_0^*(N)), \\ N, & \text{if } J(t_0^*(N-1)) - J(t_0^*(N)) > C_N > J(t_0^*(N)) - J(t_0^*(N+1)). \end{cases} \]

If \( C_N = J(t_0^*(N-1)) - J(t_0^*(N)) \) then we have:

\[ J(t_0^*(N-2)) - J(t_0^*(N-1)) > C_N > J(t_0^*(N)) - J(t_0^*(N+1)). \]

Consequently, \( N^* = N \) and \( N^* = N-1 \). Q.E.D.

By evaluating the cost function at \( N_0 + 1 \), Theorem 4.5 provides the condition where \( N_0 \) is optimal. This is the case where the maintenance wage is high. The following theorem shows that the optimal crew size is unique when the maintenance wage has the following condition:

**Theorem 4.6:**

If \( C_N = J(t_0^*(N-1)) - J(t_0^*(N)) \) for all \( N > N_0 \), then \( N^* \) is unique.

**Proof:**

From theorem 4.5, \( N^* \) is not unique if there exists \( N \) such that

\[ J(t_0^*(N-1)) + C_N (N-1) = J(t_0^*(N)) + C_N N, \text{ i.e. } C_N = J(t_0^*(N-1)) - J(t_0^*(N)). \]

From Theorem 4.4, we have the following relations:

If \( J(t_0^*(N-2)) - J(t_0^*(N-1)) > C_N > J(t_0^*(N-1)) - J(t_0^*(N)) \) then \( N^* = N-1 \).

If \( J(t_0^*(N-1)) - J(t_0^*(N)) > C_N > J(t_0^*(N)) - J(t_0^*(N+1)) \) then \( N^* = N \).

If \( J(t_0^*(N)) - J(t_0^*(N+1)) > C_N > J(t_0^*(N+1)) - J(t_0^*(N+2)) \) then \( N^* = N+1 \).
It is clear that when \( C_n \neq J(t_0^*(N-1)) - J(t_0^*(N)) \) the above relations cannot be true at the same time. This is true for all \( N > N_0 \).

Q.E.D.

In general \( C_n \neq J(t_0^*(N-1)) - J(t_0^*(N)) \) for all \( N > N_0 \). Thus we can conclude that for most cases, \( N^* \) unique.

4.6.2.5 Derivation of Algorithm 4.2

To determine the joint optimal \( (N^*, t_0^*) \), we develop an iterative algorithm which is based on the unimodality property. Each step of the algorithm requires the solution of an NLP problem. Let \( N_{\text{max}} \) be the smallest integer greater or equal to \( \sum_{i=1}^{n} u_i \lambda_i E[S_i] + u_0 \).

The steps of the algorithm are outlined below:

(i) Compute \( J(N_0, t_{0}^*) \) and set \( N = N_0 + 1 \)

(ii) Use Algorithm 4.1 to find \( J(N, t_0^*(N)) \).

If \( J(N_0, t_0^*(N)) < J(N-1, t_0^*(N-1)) \) then if \( N < N_{\text{max}} \) then go to (iv)

else set \( N = N+1 \) and go to (ii)

If \( J(N, t_0^*(N)) > J(N-1, t_0^*(N-1)) \) then go to (iii)

If \( J(N, t_0^*(N)) = J(N-1, t_0^*(N-1)) \) then go to (iii) or (iv)

(iii) Set \( (N^*, t_0^*) = (N-1, t_0^*(N-1)) \) and stop

(iv) Set \( (N^*, t_0^*) = (N, t_0^*(N)) \) and stop
Since we can reduce the problem with \( m+1 \) Lagrange multipliers to a problem with only one Lagrange multiplier, Algorithm 4.2 always find the optimal number of maintenance men for any size of the network. Let \( N \) be the smallest maintenance crew size such that \( \gamma^* = 0 \). At the worst case, Algorithm 4.2 iterates \( \min (N_{\text{max}}, N) - N_0 + 1 \) times.

4.6.3 Computational tests for case 2

To analyze and illustrate the efficiency of Algorithm 4.1, we conducted computational tests using state-of-the-art methods. First, we consider an unreliable open queueing network with 20 stations. The failure time distribution at each station is assumed to be Erlang-2. An example of data input for illustrating the two algorithms is shown in Tables 4.1 to 4.2.

Table 4.1 contains the successor nodes for each node, it describes the entire network. The node OUT means that jobs leave the system. It contains also the routing probabilities of each station corresponding to the successor nodes. Table 4.2 contains the mean failure rate \( \beta \) which varies form 0.025 to 0.095, the mean failure time \( \tau_b \), i.e., \( \tau_b = 2/\beta \), which varies from 23.53 to 80 days, the repair time is assumed to be exponentially distributed with mean \( \tau_r \) of 0.50 to 1.50 days depending on the server, the maintenance time is assumed to be exponentially distributed with mean \( \tau_m \) of 0.05 to 0.7 days, and the traffic intensity coming from an external source \( \lambda_0 E[S] \).
<table>
<thead>
<tr>
<th>Node #</th>
<th>Successor Nodes (Probability of routing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2(.20) 3(.21) 4(.13) 5(.10) 6(.05) 7(.05) 8(.11) OUT(.15)</td>
</tr>
<tr>
<td>2</td>
<td>3(.15) 4(.21) 5(.11) 6(.17) 8(.07) 9(.09) OUT(.20)</td>
</tr>
<tr>
<td>3</td>
<td>4(.15) 5(.15) 8(.15) 9(.15) 10(.10) 11(.11) OUT(.19)</td>
</tr>
<tr>
<td>4</td>
<td>5(.12) 8(.13) 9(.21) 10(.05) 11(.14) 12(.13) 15(.10) OUT(.12)</td>
</tr>
<tr>
<td>5</td>
<td>6(.13) 7(.11) 8(.12) 9(.25) 10(.12) 13(.17) OUT(.10)</td>
</tr>
<tr>
<td>6</td>
<td>7(.17) 8(.21) 12(.15) 13(.16) 16(.11) OUT(.20)</td>
</tr>
<tr>
<td>7</td>
<td>8(.15) 9(.07) 10(.05) 11(.10) 12(.15) 14(.13) 15(.13) 16(.17) OUT(.14)</td>
</tr>
<tr>
<td>8</td>
<td>9(.17) 10(.10) 11(.21) 13(.11) 13(.12) 14(.14) 18(.15) OUT(.00)</td>
</tr>
<tr>
<td>9</td>
<td>10(.21) 11(.12) 12(.07) 13(.08) 14(.11) 15(.21) 19(.15) OUT(.05)</td>
</tr>
<tr>
<td>10</td>
<td>11(.10) 14(.20) 16(.20) 19(.20) OUT(.30)</td>
</tr>
<tr>
<td>11</td>
<td>12(.25) 13(.25) 14(.10) 15(.10) 18(.10) 19(.10) OUT (.10)</td>
</tr>
<tr>
<td>12</td>
<td>13(.20) 14(.15) 15(.15) 16(.10) 17(.14) OUT (.26)</td>
</tr>
<tr>
<td>13</td>
<td>14(.12) 15(.11) 16(.13) 17(.10) 18(.19) 19(.09) 20(.10) OUT (.16)</td>
</tr>
<tr>
<td>14</td>
<td>15(.21) 16(.17) 17(.11) 18(.08) 19(.10) 20(.13) OUT (.20)</td>
</tr>
<tr>
<td>15</td>
<td>16(.10) 17(.30) 18(.12) 19(.07) 20(.13) OUT (.28)</td>
</tr>
<tr>
<td>16</td>
<td>17(.19) 18(.11) 19(.15) 20(.15) OUT (.40)</td>
</tr>
<tr>
<td>17</td>
<td>18(.25) 19(.10) 20(.20) OUT (.45)</td>
</tr>
<tr>
<td>18</td>
<td>19(.10) 20(.10) OUT (.80)</td>
</tr>
<tr>
<td>19</td>
<td>20(.20) OUT (.80)</td>
</tr>
<tr>
<td>20</td>
<td>OUT(1.0)</td>
</tr>
</tbody>
</table>

Table 4.1: The network representation
<table>
<thead>
<tr>
<th>Station #</th>
<th>$\beta$</th>
<th>$\tau_b$</th>
<th>$\tau_r$</th>
<th>$\tau_m$</th>
<th>$\lambda_0 E[S]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.055</td>
<td>36.36</td>
<td>0.90</td>
<td>0.25</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>0.040</td>
<td>50.00</td>
<td>0.65</td>
<td>0.20</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>0.075</td>
<td>26.66</td>
<td>1.10</td>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td>4</td>
<td>0.035</td>
<td>57.14</td>
<td>0.60</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>0.080</td>
<td>25.00</td>
<td>1.20</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>6</td>
<td>0.090</td>
<td>22.22</td>
<td>1.50</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>7</td>
<td>0.045</td>
<td>44.45</td>
<td>0.70</td>
<td>0.15</td>
<td>0.55</td>
</tr>
<tr>
<td>8</td>
<td>0.050</td>
<td>40.00</td>
<td>0.75</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>9</td>
<td>0.085</td>
<td>23.53</td>
<td>1.40</td>
<td>0.35</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.070</td>
<td>28.57</td>
<td>1.00</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>11</td>
<td>0.030</td>
<td>66.66</td>
<td>0.55</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>12</td>
<td>0.095</td>
<td>23.53</td>
<td>1.40</td>
<td>0.60</td>
<td>0.25</td>
</tr>
<tr>
<td>13</td>
<td>0.050</td>
<td>40.00</td>
<td>0.75</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td>14</td>
<td>0.065</td>
<td>30.77</td>
<td>0.95</td>
<td>0.40</td>
<td>0.05</td>
</tr>
<tr>
<td>15</td>
<td>0.030</td>
<td>66.66</td>
<td>0.55</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>0.025</td>
<td>80.00</td>
<td>0.50</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>17</td>
<td>0.049</td>
<td>40.82</td>
<td>0.85</td>
<td>0.30</td>
<td>0.10</td>
</tr>
<tr>
<td>18</td>
<td>0.070</td>
<td>28.57</td>
<td>1.00</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>19</td>
<td>0.080</td>
<td>25.00</td>
<td>0.95</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>20</td>
<td>0.055</td>
<td>36.36</td>
<td>0.90</td>
<td>0.35</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 4.2: Mean failure rate ($\beta$), mean failure ($\tau_b$), repair ($\tau_r$) and maintenance ($\tau_m$) times, and traffic from external node ($\lambda_0 E[S]$).
Using Equations (1) and (2) we computed $\lambda_1 E[S_1]$ and $A_1(\omega)$ and these are tabulated in Table 4.3. From Table 4.3, we can see that the queueing network is stable without maintenance, i.e., $A_1(\omega) \geq \lambda_1 E[S_1]$, for $i = 1, \ldots, 20$.

<table>
<thead>
<tr>
<th>Station #</th>
<th>$A(\omega)$</th>
<th>$\lambda E[S]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.976</td>
<td>0.850</td>
</tr>
<tr>
<td>2</td>
<td>0.987</td>
<td>0.870</td>
</tr>
<tr>
<td>3</td>
<td>0.960</td>
<td>0.910</td>
</tr>
<tr>
<td>4</td>
<td>0.989</td>
<td>0.930</td>
</tr>
<tr>
<td>5</td>
<td>0.954</td>
<td>0.890</td>
</tr>
<tr>
<td>6</td>
<td>0.937</td>
<td>0.910</td>
</tr>
<tr>
<td>7</td>
<td>0.982</td>
<td>0.860</td>
</tr>
<tr>
<td>8</td>
<td>0.982</td>
<td>0.940</td>
</tr>
<tr>
<td>9</td>
<td>0.944</td>
<td>0.900</td>
</tr>
<tr>
<td>10</td>
<td>0.966</td>
<td>0.890</td>
</tr>
<tr>
<td>11</td>
<td>0.991</td>
<td>0.870</td>
</tr>
<tr>
<td>12</td>
<td>0.944</td>
<td>0.900</td>
</tr>
<tr>
<td>13</td>
<td>0.981</td>
<td>0.930</td>
</tr>
<tr>
<td>14</td>
<td>0.970</td>
<td>0.910</td>
</tr>
<tr>
<td>15</td>
<td>0.991</td>
<td>0.940</td>
</tr>
<tr>
<td>16</td>
<td>0.993</td>
<td>0.890</td>
</tr>
<tr>
<td>17</td>
<td>0.981</td>
<td>0.870</td>
</tr>
<tr>
<td>18</td>
<td>0.966</td>
<td>0.910</td>
</tr>
<tr>
<td>19</td>
<td>0.963</td>
<td>0.960</td>
</tr>
<tr>
<td>20</td>
<td>0.977</td>
<td>0.920</td>
</tr>
</tbody>
</table>

Table 4.3: Values of $A(\omega)$ and $\lambda E[S]$ for each station
Using Equations (31) and (33) in Chapter 2 and Equation (17') we computed $t_L$, $t_A^*$, and $u_1$, respectively. These are tabulated in Table 4.4.

<table>
<thead>
<tr>
<th>Station #</th>
<th>$t_L$</th>
<th>$t_A^*$</th>
<th>$u$ ($\alpha = 0.001$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.43</td>
<td>37.09</td>
<td>-2.36</td>
</tr>
<tr>
<td>2</td>
<td>2.16</td>
<td>61.50</td>
<td>-2.57</td>
</tr>
<tr>
<td>3</td>
<td>4.23</td>
<td>48.50</td>
<td>-3.27</td>
</tr>
<tr>
<td>4</td>
<td>3.56</td>
<td>171.50</td>
<td>-3.92</td>
</tr>
<tr>
<td>5</td>
<td>3.94</td>
<td>49.50</td>
<td>-2.86</td>
</tr>
<tr>
<td>6</td>
<td>9.60</td>
<td>166.50</td>
<td>-3.27</td>
</tr>
<tr>
<td>7</td>
<td>0.92</td>
<td>31.50</td>
<td>-2.46</td>
</tr>
<tr>
<td>8</td>
<td>3.25</td>
<td>38.50</td>
<td>-4.41</td>
</tr>
<tr>
<td>9</td>
<td>3.55</td>
<td>20.50</td>
<td>-3.04</td>
</tr>
<tr>
<td>10</td>
<td>3.81</td>
<td>142.50</td>
<td>-2.86</td>
</tr>
<tr>
<td>11</td>
<td>0.67</td>
<td>39.50</td>
<td>-2.57</td>
</tr>
<tr>
<td>12</td>
<td>6.67</td>
<td>73.50</td>
<td>-3.04</td>
</tr>
<tr>
<td>13</td>
<td>2.73</td>
<td>34.50</td>
<td>-3.92</td>
</tr>
<tr>
<td>14</td>
<td>4.27</td>
<td>97.50</td>
<td>-3.27</td>
</tr>
<tr>
<td>15</td>
<td>3.96</td>
<td>366.50</td>
<td>-4.41</td>
</tr>
<tr>
<td>16</td>
<td>0.40</td>
<td>27.50</td>
<td>-2.86</td>
</tr>
<tr>
<td>17</td>
<td>2.02</td>
<td>67.50</td>
<td>-2.57</td>
</tr>
<tr>
<td>18</td>
<td>4.89</td>
<td>142.50</td>
<td>-2.27</td>
</tr>
<tr>
<td>19</td>
<td>19.76</td>
<td>79.50</td>
<td>-6.12</td>
</tr>
<tr>
<td>20</td>
<td>5.20</td>
<td>81.50</td>
<td>-3.55</td>
</tr>
</tbody>
</table>

Table 4.4: Values of $t_L$, $t_A^*$, and $u$
### 4.6.3.1 Illustration

Tables 4.5 to 4.12 display the optimal age-maintenance policy for each station found using Algorithm 4.1.

<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
<td>6.16</td>
<td>3.78</td>
<td>1.70</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
<td>8.13</td>
<td>5.07</td>
<td>2.47</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
<td>6.85</td>
<td>4.23</td>
<td>4.23</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
<td>13.37</td>
<td>7.77</td>
<td>3.56</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
<td>6.29</td>
<td>3.49</td>
<td>3.94</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
<td>9.60</td>
<td>9.60</td>
<td>9.60</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
<td>6.06</td>
<td>3.97</td>
<td>2.23</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
<td>8.40</td>
<td>5.06</td>
<td>3.25</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
<td>5.12</td>
<td>3.55</td>
<td>3.55</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
<td>7.44</td>
<td>3.81</td>
<td>3.81</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
<td>7.82</td>
<td>5.40</td>
<td>3.47</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
<td>6.67</td>
<td>6.67</td>
<td>6.67</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
<td>7.91</td>
<td>4.80</td>
<td>2.73</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
<td>8.14</td>
<td>4.27</td>
<td>4.27</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
<td>16.88</td>
<td>9.68</td>
<td>3.96</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
<td>7.39</td>
<td>5.59</td>
<td>4.24</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
<td>8.09</td>
<td>4.79</td>
<td>2.02</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
<td>6.46</td>
<td>4.89</td>
<td>4.89</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
<td>19.76</td>
<td>19.76</td>
<td>19.76</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
<td>9.40</td>
<td>5.20</td>
<td>5.20</td>
</tr>
</tbody>
</table>

M. & B. Costs: 516.55  224.04  181.37  171.77
Lagrange Mult. 69.09   24.71   0.00
CPU Time (Sec.) 19.88   7.09    5.17

Table 4.5 Optimal age-maintenance policies with different values of $N$ for $C_m = $5
<table>
<thead>
<tr>
<th>Station #</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.09</td>
<td>6.32</td>
<td>3.86</td>
<td>2.58</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
<td>8.39</td>
<td>5.27</td>
<td>3.70</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
<td>6.75</td>
<td>4.23</td>
<td>4.23</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
<td>13.31</td>
<td>7.38</td>
<td>4.21</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
<td>6.20</td>
<td>3.94</td>
<td>3.94</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
<td>9.60</td>
<td>9.60</td>
<td>9.60</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
<td>6.41</td>
<td>4.34</td>
<td>3.32</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
<td>8.43</td>
<td>4.90</td>
<td>3.25</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
<td>5.10</td>
<td>3.55</td>
<td>3.55</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
<td>7.31</td>
<td>3.81</td>
<td>3.81</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
<td>8.53</td>
<td>6.21</td>
<td>5.12</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
<td>6.67</td>
<td>6.67</td>
<td>6.67</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
<td>7.98</td>
<td>4.71</td>
<td>2.92</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
<td>8.00</td>
<td>4.27</td>
<td>4.27</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
<td>16.71</td>
<td>9.07</td>
<td>4.95</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
<td>8.54</td>
<td>6.93</td>
<td>6.22</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
<td>8.16</td>
<td>4.70</td>
<td>2.88</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
<td>6.41</td>
<td>4.89</td>
<td>4.89</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
<td>19.76</td>
<td>19.76</td>
<td>19.76</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
<td>9.27</td>
<td>5.20</td>
<td>5.20</td>
</tr>
</tbody>
</table>

M. & B. Costs: 517.02 235.35 198.85 195.01
Lagrange Mult. 62.50 16.79 0
CPU Time (Sec.) 19.77 6.81 8.18

Table 4.9 Optimal age-maintenance policies with different values of N for \( C_m = \$10 \)
<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
<td>7.85</td>
<td>7.18</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
<td>10.87</td>
<td>10.05</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
<td>6.09</td>
<td>4.99</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
<td>13.29</td>
<td>11.44</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
<td>5.60</td>
<td>4.59</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
<td>9.60</td>
<td>9.60</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
<td>9.47</td>
<td>8.98</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
<td>9.07</td>
<td>7.99</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
<td>5.18</td>
<td>4.44</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
<td>6.50</td>
<td>5.31</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
<td>14.10</td>
<td>13.62</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
<td>6.67</td>
<td>6.67</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
<td>8.95</td>
<td>7.99</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
<td>7.12</td>
<td>5.83</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
<td>15.85</td>
<td>13.40</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
<td>16.66</td>
<td>16.44</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
<td>9.04</td>
<td>8.03</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
<td>6.27</td>
<td>5.31</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
<td>19.76</td>
<td>19.76</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
<td>8.49</td>
<td>7.04</td>
</tr>
</tbody>
</table>

M. & B. Costs: 523.94 321.33 319.53
Lagrange Mult. 14.45 0.0
CPU Time (Sec.) 18.28 14.56

Table 4.10 Optimal age-maintenance policies with different values of $N$ for $C_m = $50
<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
</tr>
</tbody>
</table>

M. & B. Costs: 532.59 411.34
Lagrange Mult. 0
CPU Time (Sec.) 23.62

Table 4.11 Optimal age-maintenance policies with different values of N for C_m = $100
<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.09</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
</tr>
</tbody>
</table>

M. & B. Costs: 549.90 511.69
Lagrange Mult. 0.
CPU Time (Sec.) 22.68

Table 4.12 Optimal age-maintenance policies with different values of $N$ for $C_m = $200
<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
</tr>
</tbody>
</table>

| M. & B. Costs: | 567.20      | 554.96      |
| Lagrange Mult. | 0.          |             |
| CPU Time (Sec.) | 20.44      |             |

Table 4.13  Optimal age-maintenance policies with different values of $N$ for $C_m = $300

<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
<td>93.34</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
<td>125.87</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
<td>67.41</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
<td>142.02</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
<td>63.02</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
<td>52.88</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
<td>114.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
<td>102.25</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
<td>64.41</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
<td>69.86</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
<td>168.25</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
<td>51.39</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
<td>102.00</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
<td>76.17</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
<td>166.07</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
<td>205.00</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
<td>102.81</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
<td>69.85</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
<td>61.72</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
<td>90.87</td>
<td></td>
</tr>
</tbody>
</table>

M. & B. Costs: 584.51 566.46
Lagrange Multi. 0.
CPU Time (Sec.) 20.93

Table 4.14 Optimal age-maintenance policies with different values of N for C_m = $400
<table>
<thead>
<tr>
<th>Station #</th>
<th>Crew size N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>37.09</td>
</tr>
<tr>
<td>2</td>
<td>61.50</td>
</tr>
<tr>
<td>3</td>
<td>48.50</td>
</tr>
<tr>
<td>4</td>
<td>171.50</td>
</tr>
<tr>
<td>5</td>
<td>49.60</td>
</tr>
<tr>
<td>6</td>
<td>166.50</td>
</tr>
<tr>
<td>7</td>
<td>31.50</td>
</tr>
<tr>
<td>8</td>
<td>38.50</td>
</tr>
<tr>
<td>9</td>
<td>20.50</td>
</tr>
<tr>
<td>10</td>
<td>142.50</td>
</tr>
<tr>
<td>11</td>
<td>39.50</td>
</tr>
<tr>
<td>12</td>
<td>73.50</td>
</tr>
<tr>
<td>13</td>
<td>34.50</td>
</tr>
<tr>
<td>14</td>
<td>97.50</td>
</tr>
<tr>
<td>15</td>
<td>366.50</td>
</tr>
<tr>
<td>16</td>
<td>27.50</td>
</tr>
<tr>
<td>17</td>
<td>67.50</td>
</tr>
<tr>
<td>18</td>
<td>142.50</td>
</tr>
<tr>
<td>19</td>
<td>79.50</td>
</tr>
<tr>
<td>20</td>
<td>81.50</td>
</tr>
</tbody>
</table>

M. & B. Costs: 601.81 566.82
Lagrange Mult. 0.
CPU Time (Sec.) 57.34

Table 4.15 Optimal age-maintenance policies with different values of $N$ for $C_m = $500
To check the quality of the results of Algorithm 4.1, first we used an NLP software called "GAMS".

4.6.3.2 GAMS

GAMS (General Algebraic Modeling System) [Brooke et. al. (1988)] is a software designed for constructing and solving large and complex mathematical programming models. GAMS was developed to provide a high level language for the compact representation of large models. GAMS applies MINOS 5 for solving nonlinear problems. GAMS/MINOS is a FORTRAN-based system designed to solve NLP problems expressed in the following form:

\[
\begin{align*}
\text{minimize } & \quad f(x) \\
\text{s.t. } & \quad g_i(x) \leq b_i \\
& \quad l \leq x \leq u
\end{align*}
\]  

The components of \( x \) are called the nonlinear variables. \( f(.) \) is the cost function and \( g(.) \) is a vector of smooth functions \( g_i \).

GAMS/MINOS employs a projected Lagrangian algorithm. This involves a sequence of major iterations. Each major iteration requires the solution of a linearly constrained subproblem. Each subproblem contains linearized versions of the nonlinear constraints as well as the original linear constraints and bounds. At the start of the \( k \)-th major iteration, let \( x_k \) be an estimate of the nonlinear variables and let \( \lambda_k \) be an estimate of the Lagrangian multipliers associated with the nonlinear constraints. The constraints are linearized by changing \( g(x) \) in (29) to its linear approximation:
\[ \tilde{g}(x) = g(x_k) + J(x)_k(x - x_k). \]

Where \( J(x_k) \) is the Jacobian matrix evaluated at \( x_k \). Note that the i-th row of Jacobian is the gradient vector for the i-th nonlinear constraint function. The subproblem to be solved during the k-th major iteration is then

\[
\begin{align*}
\text{minimize } & \quad f(x) - \lambda_k^T(g - \tilde{g}) + \frac{1}{2}\delta (g - \tilde{g})^T(g - \tilde{g}) \\
\text{s.t. } & \quad \tilde{g} \leq b \\
& \quad 1 \leq x \leq u.
\end{align*} \tag{31}
\]

The objective function (31) is called an augmented Lagrangian. The scalar \( \delta \) is a penalty parameter, and the term involving \( \delta \) is a modified quadratic penalty function. GAMS uses the reduced-gradient method to minimize (31) subject to (32) and (33).

Unfortunately, there is no guarantee the algorithm will converge from an arbitrary starting point. To influence the likelihood of convergence, we have to consider the following steps:

1. Specify initial values for \( t_i \) as carefully as possible.
2. Include sensible upper and lower bounds on \( F_i(t_i) \) and \( A_i(t_i) \).
3. Specify a penalty parameter \( \delta \) that is higher than the default value.

We used GAMS algorithm to find the optimal age-maintenance for a fixed \( N \), for (9)-(11) with \( C_N = 0 \) and \( C_w = 0 \).
4.6.3.3 Comparison of the results of Algorithm 4.1 with the GAMS output

Using a 286 IBM-AT compatible with math co-processor, Table 4.13 compares the optimal costs and CPU times for different values of $C_m$ and $N$. The stopping rule is $\varepsilon = 10^{-6}$. The cost of the maintenance crew size is not included in the objective function. First, we determine the value of $N_0$ from Equation (13), i.e., $N_0 = 3$. For $N = 3$, $t^*_A$ is optimal. For $N = 4$, Algorithm 4.1 guesses the value of $\gamma_{\max}(N)$. For $N > 4$, the algorithm uses $\gamma^*(N-1)$ as the value of $\gamma_{\max}(N)$ which decreases the CPU time almost 3 fold. When $\gamma^*(N) = 0$, this means that $N$ is an upper bound.

The results of Algorithm 4.1 is compared against the output from GAMS for different values of $C_m$ and $N$. We conclude that Algorithm 4.1 is much faster that GAMS. The maximum error is of the order of 5.5%. For $C_m \approx 100$, we have the same results. Note the difference in CPU times. For $C_m = 300$, the GAMS software failed to give a solution. After more than 30 minutes of execution, GAMS gave the following message: "Exit - Too Many Iterations". When the number of stations increases (e.g., $M = 40$), GAMS cannot provide an optimal solution. Figures 4.6 and 4.7 compare graphically differences in CPU times which are taken from Table 4.13.
<table>
<thead>
<tr>
<th>$C_m$</th>
<th>N</th>
<th>COST</th>
<th>CPU TIME</th>
<th>COST</th>
<th>CPU TIME</th>
<th>% ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>224.04</td>
<td>19.88 Sec</td>
<td>212.36</td>
<td>8.67 Min</td>
<td>5.50%</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>181.37</td>
<td>7.09 Sec</td>
<td>173.08</td>
<td>3.98 Min</td>
<td>4.78%</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>171.77</td>
<td>5.17 Sec</td>
<td>171.24</td>
<td>3.38 Min</td>
<td>0.31%</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>235.35</td>
<td>19.77 Sec</td>
<td>224.79</td>
<td>6.23 Min</td>
<td>4.69%</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>198.85</td>
<td>6.81 Sec</td>
<td>194.67</td>
<td>4.10 Min</td>
<td>2.15%</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>195.01</td>
<td>8.18 Sec</td>
<td>194.60</td>
<td>4.00 Min</td>
<td>0.21%</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>321.33</td>
<td>18.28 Sec</td>
<td>319.68</td>
<td>8.87 Min</td>
<td>0.52%</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>319.53</td>
<td>14.56 Sec</td>
<td>319.56</td>
<td>8.78 Min</td>
<td>BETTER</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>411.34</td>
<td>23.62 Sec</td>
<td>411.34</td>
<td>12.17 Min</td>
<td>0.00%</td>
</tr>
<tr>
<td>200</td>
<td>4</td>
<td>511.69</td>
<td>22.68 Sec</td>
<td>511.69</td>
<td>24.08 Min</td>
<td>0.00%</td>
</tr>
<tr>
<td>300</td>
<td>4</td>
<td>554.96</td>
<td>20.44 Sec</td>
<td>-</td>
<td>&gt; 30 Min</td>
<td>-</td>
</tr>
</tbody>
</table>

* Stopping Rule: ($\varepsilon = 10^{-6}$)

**Table 4.7: Algorithm 4.1 vs. GAMS**
Figure 4.7: Comparison of CPU times
4.6.3.4 Optimizing the crew size

We have proved that the total expected cost is unimodal with respect to $N$. This is illustrated in Figure 4.8 for $C_m = 5$ and $C_R = 10$ and in Figure 4.9 for $C_m = 5$ and $C_R = 100$. For a given feasible value of $N$, Algorithm 4.1 is used to determine the optimal age-maintenance policy and its expected cost. Using the unimodality property, we can determine the optimal maintenance men $N^*$. We observe that the cost of having fewer men is considerably greater than the cost of having more. When the number of nodes increases, Algorithm 4.1 provides the optimal age-maintenance policy with the CPU time approximately linear to the number of stations. This is shown in Figure 4.10. the couples $(20,5)$, $(40,10)$, $(60,15)$, $(80,20)$, and $(100,25)$ have almost the same optimal Lagrange multipliers.
Figure 4.8: Total cost as function of maintenance crew size
\[ C_m = $5 \text{ and } C_R = $10 \]
Figure 4.9: Total cost as function of maintenance crew size

$C_a = $5$ and $C_n = $100$
Figure 4.10: CPU time vs. \((m,N)\)
4.6.3.5 Sensitivity of $N^*$

From Tables 4.5 to 4.9, the optimal maintenance crew size $N^*$ is sensitive to the values $C_N$ and $C_m$ as follows:

$$C_m = 5, \quad N^* = \begin{cases} 
6 & \text{if } C_N < 9.6 \\
5 & \text{if } 9.6 \leq C_N < 42.67 \\
4 & \text{if } 42.67 \leq C_N < 292.50 \\
3 & \text{if } C_N \geq 22.50 
\end{cases}$$

$$C_m = 10, \quad N^* = \begin{cases} 
6 & \text{if } C_N < 3.85 \\
5 & \text{if } 3.85 \leq C_N < 36.50 \\
4 & \text{if } 36.50 \leq C_N < 281.70 \\
3 & \text{if } C_N \geq 281.70 
\end{cases}$$

$$C_m = 50, \quad N^* = \begin{cases} 
5 & \text{if } C_N < 1.8 \\
4 & \text{if } 1.8 \leq C_N < 202.60 \\
3 & \text{if } C_N \geq 202.60 
\end{cases}$$

$$C_m = 100, \quad N^* = \begin{cases} 
4 & \text{if } C_N < 121.25 \\
3 & \text{if } C_N \geq 121.25 
\end{cases}$$

$$C_m = 200, \quad N^* = \begin{cases} 
4 & \text{if } C_N < 38.20 \\
3 & \text{if } C_N \geq 38.20 
\end{cases}$$

When the value of $C_m$ is high, $N^*$ is insensitive with respect to the maintenance times. On the other hand, when $C_m$ is low and $C_N$ is low, $N^*$ increases as we increase the maintenance time at each node. When $C_N$ is high, $N^*$ decreases with respect to the maintenance times.
4.7 CONCLUSION

A central issue in this chapter is the modeling of equipment failures in a network of queues. By imposing the heavy traffic condition at each node, the problem can be modeled as a nonlinear program. Since the objective function has a unique minimum and the resource constraint has a unique minimum, the bisection method can be used successfully to determine the optimal Lagrange multiplier for each feasible value of \( N \). The optimal Lagrange multiplier for the crew size \( N \) can be used as an upper bound of \( N+1 \). As the optimal fraction of time the server is on approaches the traffic intensity, the solution of Algorithm 4.1 is close to GAMS. The advantage of Algorithm 4.1 is not only the speed but also in providing a solution for any number of stations. The optimum corresponds to the minimum of the total cost which includes the server off-time cost and the wages for the maintenance crew. The solution can be found iteratively using Algorithm 4.1 for solving an NLP formulation with different crew sizes. Using the unimodality property, the optimal number of maintenance men is the value of \( N \) such the total cost at \( N \) is less than the total cost at \( N+1 \). In general the optimal crew size is unique. Extensions of this chapter are provided in Chapter 5.
CHAPTER FIVE

Conclusions and Further Research
5.1 SUMMARY AND CONCLUSION

The problem of providing the optimal maintenance decision to queueing systems with service interruptions where the maintenance and repair times are not negligible, is very important. Unfortunately, finding the solution in the general case is very complex due to the intractability of some queueing performance measures. We investigated a special case of the problem where queueing systems are under the heavy traffic condition and servers are assumed to be like-new after undergoing maintenance or repairs.

In Chapter one, we began by providing some background about the problem and exploring possible research avenues. In Chapter two, we explored the use of diffusion approximation methods to determine the expected waiting time of jobs in the system under the age-maintenance policy. We conjectured that the expected waiting time is increasing with the total traffic intensity (i.e., job traffic intensity plus the fraction of time the server is under maintenance or repair). The maintenance decision that maximizes the fraction of time the server is on, also minimizes the expected waiting time in the system. We provided some conditions on the cost parameters, maintenance and repair times, and failure rate function so that the optimal age-maintenance can be found efficiently.

In Chapter three, we studied an $M/M/1$ queue with buffer and quality costs. We showed that when the cost of quality is zero, the total expected cost is unimodal with respect to the buffer and when
the cost of buffer is zero, the expected total cost is unimodal with respect to the allowed number of defective items. The maintenance decision which decreases the waiting capacity, increases the number of defective item per cycle and vice versa. The trade-off between the buffer cost and the quality cost is considered. We developed an efficient procedure for determining the optimal policy.

In Chapter four, we examined the maintenance problem of an open queueing network with maintenance crew size cost. At each feasible crew size, the bisection method is used successfully to update the Lagrange multiplier in order to determine the corresponding optimal age-maintenance policies. We found that the optimal Lagrange multiplier of crew size $N$ can be used as the upper bound for the optimal Lagrange multiplier of crew size $N+1$. This reduced the CPU time almost 3 folds. We showed that the total expected cost is unimodal with respect to the maintenance crew size. We developed a computationally expedient procedure to determine the optimal crew size. This procedure is based on the unimodality property. In general, the optimal maintenance crew size is unique. Basically, the proposed algorithm can provide the optimal solution for any size of the network.
5.2 FURTHER RESEARCH

The implementation of preventive maintenance policies for queueing systems with service interruptions needs much work. This thesis opens up large avenues for further research. Extensions of the thesis are listed below:

5.2.1 Medium Traffic Intensity

When the traffic intensity is not high, the diffusion approximation method presented in Chapter 2 cannot provide an accurate approximation of the expected waiting time in the system. The analysis in Chapter four is also based on the heavy traffic assumption. When the traffic intensity is not high, other approaches must be used.

5.2.2 Queueing Systems without the like-new condition

Most models in the literature on maintenance have assumed that the system is like new after undergoing preventive maintenance or repair. This assumption might not hold true for some queueing systems, where the improvement of repair and maintenance depends on the age of the server. Since the maintenance action is not the replacement of the queueing system, maintenance may not renew the system completely. Some models without the like-new condition are
presented in the literature on reliability. For example, Nguyen (1981) and Nagakawa (1986) presented a model where the failure rate increases with the number of PM/repair carried out. The server is replaced at the $n^{th}$ maintenance/repair. Kapur et al. (1989) developed models where the system is replaced after the $n^{th}$ failure, or when the estimated minimal repair cost exceeds a predetermined limit cost. In their model the minimal repair cost has a cumulative distribution which increases with each repair. Kijima (1989) developed general repair models for a repairable system by using the idea of the virtual age process of the system. He assumed that the virtual age accumulated with failures. However, the above models assumes that the times for maintenance, repair, and replacement are negligible. A possible extension of this thesis is to drop the above assumptions.

5.2.3 Quality Control Requirement

In Chapter 3, we have assumed that defective items are rejected from the system. A possible extension is to allow the defective items to join the queue for rework. The queueing system with bernouilli feedback is an open area for research because the feedback process is not Poisson process.
Also we have assumed that maintenance/repair action restores the system to its new condition. Another possible extension is to incorporate Section 5.2.2 to the quality model.

5.2.4 Unreliable queueing networks

From Chapter 4, we have found that the output process at each station is not a renewal process due to the presence of off-time periods. Since the output of one station is the input to other stations, the arrival process of each node does not occur in Poisson stream. Consequently, existing approximation methods in the literature cannot be used to determine the expected waiting time in the system at each node. Development of an approximation method for this particular problem which can be validated by simulation experiment, is an important result. One may try to use the proposed transformation in Chapter 4 by estimating the probability of rejection.

5.2.5 Maintenance crew size with different skills

The maintenance division can be divided into separate shops for maintenance, e.g., mechanical, electrical, electronic, etc. The
wages are not necessarily equal in these different maintenance shops. Maintenance men may have different skills. A single-skill person can perform only one job. A double-skill person can perform two jobs (mechanical and electrical, or mechanical and electronic, or, ...). However, a multi-skill person can perform all types of jobs. The wage of multi-skill person is higher than a double-skill person and so on. Any device can have all type of breakdowns with the same failure time distribution and same failure rate. For possible extension one may incorporate the different skills problem to Chapter 4, and determine the optimal number of each level of skills.
REFERENCES


